Scientific Basis of Water-Resource Management (1982)

Pages
142

Size
8.5 x 10

ISBN
030903244X

Geophysics Study Committee; Geophysics Research Board; Assembly of Mathematical and Physical Sciences; National Research Council

STUDIES IN GEOPHYSICS

Scientific Basis of Water-Resource Management

Panel on Scientific Basis of Water-Resource Management

Geophysics Study Committee
Geophysics Research Board
Assembly of Mathematical and Physical Sciences
National Research Council

NATIONAL ACADEMY PRESS
Washington, D.C. 1982

82-0052
C. 1

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the Councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the Committee responsible for this report were chosen for their special competences and with regard for appropriate balance.

This report has been reviewed by a group other than the authors according to procedures approved by a Report Review Committee consisting of members of the National Academy of Sciences, the National Academy of Engineering, and tne Institute of Medicine.

The National Research Council was established by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and of advising the federal government. The Council operates in accordance with general policies determined by the Academy under the authority of its congressional charter of 1863, which establishes the Academy as a private, nonprofit, self-governing membership corporation. The Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in the conduct of their services to the government, the public, and the scientific and engineering communities. It is administered jointly by both Academies and the Institute of Medicine. The National Academy of Engineering and the Institute of Medicine were established in 1964 and 1970, respectively, under the charter of the National Academy of Sciences.

The Geophysics Study Committee is pleased to acknowledge the support of the National Science Foundation, the Defense Advanced Research Projects Agency, the National Aeronautics and Space Administration, the U.S. Geological Survey, the National Oceanic and Atmospheric Administration, and the Department of Energy (Grant #DE-FG02-80-ER10757) for the conduct of this study.

Library of Congress Cataloging in Publication Data

National Research Council. Geophysics Study Committee.
Scientific basis of water-resource management.
(Studies in geophysics)
Based on papers presented at the American
Geophysical Union meeting in San Francisco in
Dec. 1979.
1. Hydrology--Congresses. 2. Water resources
development—Congresses. I. American Geophysical
Union. II. Title. III. Series.
GV652.N37 1982 333.91 82-3538
ISBN 0-309-03244-X AACR2

Available from

NATIONAL ACADEMY PRESS
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Printed in the United States of America

Geophysics Research Board

Geophysics Study Committee

CHARLES L. DRAKE, Dartmouth College, *Chairman*
LOUIS J. BATTAN, University of Arizona, *Vice Chairman*
JOHN D. BREDEHOEFT, U.S. Geological Survey
ALLAN V. COX, Stanford University
HUGH ODISHAW, University of Arizona
CHARLES B. OFFICER, Dartmouth College
RAYMOND G. ROBLE, National Center for Atmospheric Research

Liaison Representatives

LEONARD E. JOHNSON, National Science Foundation
GEORGE A. KOLSTAD, Department of Energy
NED A. OSTENSO, National Oceanic and Atmospheric Administration
WILLIAM RANEY, National Aeronautics and Space Administration
CARL F. ROMNEY, Defense Advanced Research Projects Agency
ROBERT L. WESSON, U.S. Geological Survey

Staff

THOMAS M. USSELMAN

Panel on Scientific Basis of Water-Resource Management

MYRON B. FIERING, Harvard University, *Chairman*
VICTOR R. BAKER, University of Texas
J. W. BIGGAR, University of California, Davis
JOHN D. BREDEHOEFT, U.S. Geological Survey
JOHN CAIRNS, JR., Virginia Polytechnic Institute and State University
THOMAS DUNNE, University of Washington
PETER S. EAGLESON, Massachusetts Institute of Technology
VIT KLEMES, Fisheries and Environment Canada
GEORGE KUCZERA, Harvard University
JURATE MACIUNAS LANDWEHR, U.S. Geological Survey
LUNA B. LEOPOLD, University of California, Berkeley
NICHOLAS C. MATALAS, U.S. Geological Survey
DONALD NEILSEN, University of California, Davis
RAYMOND SIEVER, Harvard University
WERNER STUMM, Swiss Federal Institute of Technology
M. GORDON WOLMAN, The Johns Hopkins University

Studies in Geophysics*

ENERGY AND CLIMATE
Roger R. Revelle, *panel chairman*, July 1977, 158 pp.

CLIMATE, CLIMATIC CHANGE, AND WATER SUPPLY
James R. Wallis, *panel chairman*, August 1977, 132 pp.

ESTUARIES, GEOPHYSICS, AND THE ENVIRONMENT
Charles B. Officer, *panel chairman*, August 1977, 127 pp.

THE UPPER ATMOSPHERE AND MAGNETOSPHERE
Francis S. Johnson, *panel chairman*, October 1977, 169 pp.

GEOPHYSICAL PREDICTIONS
Helmut E. Landsberg, *panel chairman*, May 1978, 215 pp.

IMPACT OF TECHNOLOGY ON GEOPHYSICS
Homer E. Newell, *panel chairman*, September 1979, 121 pp.

CONTINENTAL TECTONICS
B. Clark Burchfiel, Jack E. Oliver, and Leon T. Silver,
panel co-chairmen, February 1980, 197 pp.

MINERAL RESOURCES: GENETIC UNDERSTANDING FOR PRACTICAL APPLICATIONS
Paul B. Barton, Jr., *panel chairman*, 1981, 119 pp.

SCIENTIFIC BASIS OF WATER-RESOURCE MANAGEMENT
Myron B. Fiering, *panel chairman*, 1982, 127 pp.

*Published to date.

Preface

In 1974 the Geophysics Research Board completed a plan, subsequently approved by the Committee on Science and Public Policy of the National Academy of Sciences, for a series of studies to be carried out on various subjects related to geophysics. The Geophysics Study Committee was established to provide guidance in the conduct of the studies.

One purpose of the studies is to provide assessments from the scientific community to aid policymakers in decisions on societal problems that involve geophysics. An important part of such an assessment is an evaluation of the adequacy of present geophysical knowledge and the appropriateness of present research programs to provide information required for those decisions.

This study was motivated by the perceived need for substantial improvements in the hydrologic sciences to enable more firmly based decisions involving water. This is particularly important in dealing with water quality, the interaction of water quality and the ecosystem, toxic and radioactive wastes, and the management of water for agricultural and energy production. Larger-scale and longer-term understanding of the scientific basis will enable more accurate predictive statements on problems involving large investments and human welfare.

The study was developed through meetings of the panel and presentation of papers in a preliminary form at the American Geophysical Union meeting in San Francisco in December 1979. The papers provide examples of our current geophysical knowledge base in hydrology and how that knowledge base interacts with the management and planning of our water resources. In completing their papers, the authors had the benefit of discussion at this symposium as well as the comments of several scientific referees. Responsibility for the individual essays rests with the corresponding authors.

The Overview of the study summarizes the highlights of the essays and formulates conclusions and recommendations. In preparing it, the panel chairman had the benefit of meetings and discussion that took place at the symposium and the comments of the panel of authors and selected referees. Responsibility for its contents rests with the Geophysics Study Committee and the chairman of the panel.

Contents

Scientific Basis of Water-Resource Management

Overview and Recommendations

INTRODUCTION

Hydrologists and water-resource managers and planners can deal effectively with many, if not most, questions of local water supply. In fact most existing water-management institutions (e.g., state and local water boards and irrigation districts) were designed to manage water supplies. However, hydrologists are confronted increasingly with long-term and large-scale (regional and national) problems of water management, such as (1) groundwater contamination due to toxic and nuclear-waste disposal; (2) nonpoint sources of pollution on our stream systems; (3) impacts of changes in both flow and water quality on the aquatic ecosystem; (4) the frequency, duration, and impacts of droughts including long-term trends toward desertification; (5) long-term hydrologic budgets for assessing the adequacy of regional or national water resources; (6) global geochemical cycles such as the fate of nitrogen and sulfur (e.g., acid rain); and (7) protection of engineered systems (e.g., nuclear power plants) against hydrologic extrema. These macroscale and long-term problems, involving large investments and the health and well-being of much of the world's population, demand increasingly precise and accurate predictive statements.

Until recently there has been a crude congruence between the knowledge needed to make forecasts and the time horizon (the economic period over which the project costs are amortized) of most hydrologically dependent decisions. The report, *Geophysical Predictions* (NRC Geophysics Study Committee, 1978), describes many of these forecasts, for example, the prediction of flood heights produced by rainstorms. There are many other such forecasts, e.g., next season's water supply, the drawdown of water levels by pumping, and the oxygen sag-curve in a polluted stream. These forecasts are short-term forecasts so that projects depending on them can be monitored, controlled, reversed, or abandoned to avoid egregious errors. But increasingly intense use of the hydrosphere for storage, detoxification, and assimilation of wastes, along with the potential global hydrologic effects of other societal activities (e.g., energy production, intensive agricultural development, and urbanization) and climatic change, has shifted much of the emphasis in prediction to long-term hydrologic phenomena. What makes the problems confronting the hydrologic community particularly important now are their global and temporal scales and the rate at which they are imposed on us.

Some important problems confronting hydrology were outlined in the previous paragraphs. Solutions to these problems require new approaches extending beyond conventional hydrologic knowledge and practice. This report examines the problems and the differences in approaches, with emphasis on gaps in our scientific understanding. Examples of these differences follow:

1. The difference in viewpoint, as well as in practice, between the use of *empirical and causal models* has deep roots in hydrology. Empirical models offer a practical answer to today's problems. These models require fitting to onsite data and continuous updating with further collection of data. They can have low capacity for transferability in time and place and for extrapolation beyond the range of the data. Causal models, based on the principles of physics, chemistry, and biology, make possible (at least in theory) the explicit incorporation of the effects of natural and artificial changes in the land and water environments provided the same hydrologic mechanisms are involved. Whether they do is a pivotal issue because it is tempting to resort to familiar mechanistic constructs or to develop new ones on theoretical arguments, but these cannot adequately be tested over large temporal and spatial scales. In principle, causal models are transferable in space and time, but in practice this is often found to be impossible. Accordingly, statistical tools, or inferential analysis, necessarily have been used to plan water-resource systems.

An operational and intermediate view is that hydrologic determinism admits statistical as well as causal relations. Excessive emphasis on causal or theoretical models diffuses hydrologic inquiry by concentrating primarily on description and explanation, when these might not be possible. The two categories are not in conflict but rather in a symbiotic relationship. Statistical principles help bridge those gaps in causal relationships that arise from unknown and perhaps unknowable components called uncertainties or noise (e.g., in quantum mechanics, scattering, and turbulence).

2. Related to the matter of models is the difference between *prediction and understanding*--the first is directed to immediate needs by whatever methods are most practical or that meet some optimization objective and the second toward improved ability to explain the connecting factors that control events in time and space. These should be intertwined because progress in predicting the performance of complex systems almost always requires both. Some might take prediction (see Chapter 11) as the ultimate purpose of hydrology--and science in general. On the other hand, there are others who hold that prediction is a means of assessing understanding and that understanding is itself a temporal phenomenon.

3. Attempts at solutions for most of today's water problems are concentrated at a *local* (watershed or aquifer) spatial and temporal scale with little reference to the *global* hydrologic cycle of which the local systems are a part. Many hydrologists believe that interaction between continental waters, the oceans, and the atmosphere must also become a part of hydrologic inquiry especially as large-scale problems of climate and environmental change become critical.

The controversies associated with empirical versus causal models, prediction versus understanding, and local versus global scale are the main concerns of this report. They are all related to the apparent conflict between immediate or short-run concerns and those of the long run. This was reflected in some unresolved differences among panel members.

The time horizon for hydrologic prediction is of the essence, because short-term forecasts often are insensitive to a wide range of imperfections that would be intolerable in long-range schemes.

PERSPECTIVES ON WATER-MANAGEMENT PRACTICE

Society has long had to make major decisions involving large investments under conditions of uncertain costs (fiscal, societal, and environmental) and uncertain benefits. The large investment of time and resources associated with initiation of water-resource projects often makes it practically impossible to discard a project in favor of a significantly different alternative--the inertia can be enormous. What makes such management decision making so critical is the possibility of irreversible, unanticipated effects--perhaps on a global scale--such as those on climate, water chemistry, and land erosion. Any environmental decision involves risk. By minimizing this risk we can make more efficient use of our resources; there is a penalty for uncertainty in that full utilization of resources is compromised when consequences cannot be predicted closely. The long-term effects of irreversible error might be intolerable.

The long lead times and the complexity of the planning process--involving scientific, economic, environmental, political, recreational, and societal concerns (Lord *et al.*, 1979)--put a premium on the accuracy of the predictions of benefits and effects. Confidence in our predictive skills can be maintained only through audits of completed projects. The difficulty with retrospective audits is that the projects themselves may significantly change the local societal and hydrologic environment so that there is insufficient opportunity to monitor the system for its long-term performance. It frequently becomes impossible to make reasonable postproject assessments of the impact of a project unless there are causal models to estimate reliably (after the fact) what would have happened had the project not been constructed (or had it been constructed and operated differently). Such estimates are needed to evaluate the success or failure of the large investment in water-supply projects and in environmental control. Moreover, as the chapters in this report stress, substantial improvements in hydrologic science are necessary (but not always sufficient) for the construction of such causal models. For example, conjunctive use of surface water and groundwater will, under conditions of resource scarcity, become more important in water-resource planning. We must know more about water and chemical transport through the vadose zone--the unsaturated zone between the surface and the groundwater table (see Chapter 3 for further discussion).

Few would dispute the need for an improved physical, chemical, and biological understanding of the hydrologic regime, but modern computers and the opportunities that they afford for use of probabilistic modeling techniques suggest an extension of this basic knowledge. For example, in modeling aquatic ecosystems it would be inadequate to deal only with deterministic processes because of inherent random variabilities in the system. One issue facing the hydrologic community is determination of when one mode or another of investigation is more appropriate; the symbiosis of theoretical and empirical approaches mentioned above is essential.

In addition, customary statistical analyses that have been based on hydrologic extremes need further attention. Floods, historically dramatic and threatening, have been the wellspring of statistical hydrology --but droughts are an equally devastating kind of extreme albeit different scales of time and space and different consequences. The estimation of hydrologic extremes and their consequences--floods, droughts, glacial advances and retreats, channel degradation and aggradation--is a pervasive goal of hydrology, often beyond the powers of the customary statistics derived from an observational record; these phenomena require further understanding if their impacts are to be predicted reliably. Moreover, the phenomena are distinct: Droughts are not merely the statistical opposite of floods. Dunne (Chapter 1) shows that climatic fluctuations alter rainfall-runoff relationships so that even long records might not provide an adequate sample. Leopold (Chapter 9) and Baker (Chapter 10) argue persuasively for the use of the geomorphic record as a source of information on floods and droughts. This suggests the possible application of the ergodic theorem (observations in space as a substitute for those in time). Matalas *et al*. (Chapter 11) endorse the use of geomorphic data as well as techniques such as regionalization of hydrologic data. The problems of extremes deserve scientific attention on their own merit to achieve better estimation of the magnitude of hydrologic events of low probability.

MODES OF HYDROLOGIC ANALYSES

The recent history of water-resource investigations has been dominated by computer models for planning and management, without corresponding attention to the scientific advances in understanding physical hydrology. There are several observations concerning this trend:

1. Computers have facilitated the estimation of empirical numerical relationships, thus making the publication of papers on numerical models more attractive than further basic research in hydrologic processes. "Fascination with automatic computation has encouraged a new set of mathematical formalisms simply because they can now be computed; we have not often enough asked ourselves whether they ought to be computed or whether they make any difference . . . " (Fiering, 1976). This focus and reliance on computational power and empirical relationships have consumed resources that might have been applied to theoretical, descriptive, and field hydrology.
2. Several early and highly innovative water-resource research programs emphasized the interdisciplinary (physical, social, economic, and political) approach to planning. This, unfortunately, has overwhelmed the interest in theoretical hydrology and masked its importance.
3. Many available theoretical watershed models require an extraordinary number of parameters--perhaps as many as 50. The estimation of these parameters by enforcing agreement between basin inputs and outputs leads to nonunique sets of results. This inherent nonuniqueness makes applications of these models potentially inappropriate when applied to hydrologic phenomena on different spatial and temporal scales. To minimize this problem, generalizations from offsite data and hydrologic experience are required.
4. The availability of mathematical programming software for rapid solution of management and decision-oriented problems prompted promiscuous usage that soon dominated water-resource planning activities. Decisions based on these available (and often inadequate) models might not have been optimal because the models necessarily discounted some significant hydrologic realities that could not be accommodated within computationally tractable analyses.
5. New problems associated with water-quality management and indices of water-quality trends have focused attention on assessment of standards, economic equity, and regulatory and management policies.
6. The indiscriminate use of large computers in the modeling of large systems concentrated on the linkages (such as ecological impact, economic equity, recreational uses, and other societal impacts) between the model components and not on the hydrologic bases of the components themselves. There has been a trend to rely on interdisciplinary committees to develop strategies for managing these large systems.

Just as insistence on causal models can lead to some nonunique estimation of parameters (with large fluctuations of the parameter set due to minor changes in the data), so might emphasis on probabilistic and management models without adequate concern for hydrologic data or scientific understanding lead to some scientific traps. For example, the extreme value statistical theory (Gumbel, 1954) used in the derivation of some flood-frequency distributions does not fit well with many observational hydrologic data. Because empirical evidence often did not support proposed statistical models, other (and presumably better) statistical models were sought, generally with little concern for the hydrologic phenomena that they attempted to describe. Another example is that hydrologic practice stresses the use of unbiased estimates. However, Fiering and Kuczera (Chapter 7) and Matalas *et al.* (Chapter 11) argue that, for many predictive and operational purposes, biased regional estimators have desirable properties that offer more advantages than disadvantages.

FINDINGS

The three major findings of this report are as follows:

1. Neither empirical-optimization models nor increased causal understanding alone will enable hydrologists to respond with the necessary effectiveness to current societal demands--a combination of the two is required. The scientific understanding of hydrology is currently inadequate (documented in all the accompanying chapters) and needs substantial advances. Statistical models also are inadequate, but the panel did not address this issue. Both forms of analysis have strengths and weaknesses.

Even though this study emphasizes causal understanding, it is argued that it is unrealistic to cast aside statistical models in the hope that deterministic or causal principles will be found to provide timely answers to the urgent planning questions. The most challenging hydrologic problems occur at intermediate scales because of the difficulty in extrapolating from microlevel understanding or from macrolevel description. We cannot afford to be selective in the quest for usable tools; indeed, even if refined tools are found they might not be as useful to the planner as more pedestrian models.

2. The problems of hydrologic scale impact not only scientific alternatives and research directions but also the management of water-resource systems. Institutional jurisdictions extend to those mesoscale concerns (typically basins) that are of a similar scale to the principal unresolved hydrologic problems.

3. Although these institutional aspects are not specifically addressed within the context of this study, the panel recognized the importance of these problems. No scientific approaches will be totally successful without reduction of existing institutional difficulties (see NRC Water Resources Research Review Committee, 1981). Consider the following examples. (1) Some institutions collect, while others are consumers of, data. This has often led to fragmented data bases for which integrated data-retrieval systems are prepared at great expense but only used infrequently. (2) Watersheds and political boundaries rarely coincide. This often leads to jurisdictional problems that should and could be overcome. (3) There is evidence that some institutions do not now efficiently utilize the available knowledge in planning and discharging their mandated responsibilities [e.g., the requirement that virtually all flood-frequency analyses be based on the log-Pearson III density (U.S. Water Resources Council, 1977) despite its recognized inadequacies and proposed modifications]. One might thus infer that more thorough understanding and regional data bases might not be fully utilized because of traditions, inertia, and the longer time constants of institutional change. (4) Some institutions deal with water quality and some with quantity, which sometimes leads to regulatory and programmatic inconsistencies.

These findings suggest that it is far more practical to express these research opportunities in terms of questions to be answered than in terms of categories of social objectives or problems. Almost all of the papers in this report look at hydrologic research in this way.

Several specific fundamental and applied research questions are cited as examples of the scientific issues that need to be addressed.

1. Can the concept of scale--temporal and spatial--be included in hydrology so that some of the processes can unambiguously and efficiently be studied and managed at one level and some at another? The drainage basin or aquifer has been the customary unit for hydrologic inquiry, and this unit has indeed been satisfactory for short-run projections. However, the watershed is only a local or regional unit in the global moisture cycle, which has branches in the continents, the oceans, and

the atmosphere. A change made by man of the land surface alters the natural balance through changes in the transfer of heat and moisture between the land and the atmosphere. Determination of the new balance raises many questions that tax present hydrologic understanding. The long-term alteration of permeability of confining materials around a nuclear-waste repository involves time scales heretofore neglected in conventional hydrologic inquiry. These questions, in which the temporal and spatial scales are quite different from those of the watershed or aquifer, are at the bottom of uncertainties about the long-run effects of global land use change and of climate change (Chapter 2).

2. How can chemistry and biology be coupled with our understanding of the physics of water movement? How can this understanding be applied to the management of groundwater and surface-water quality (Chapters 5 and 6), especially to the resolution of problems of waste transport and disposal? Do institutional specilizations require reorientation to apply this coupling?

3. What is the impact on the food web--and, ultimately, on human life--of toxic substances in water (Chapters 5 and 6)? Do the long-time horizons explicit in the migration of toxic substances impose the need for a new calculus of risk assessment to establish optimal regulatory standards?

4. Given various theoretical models, what advances and techniques are needed to measure and define statistically stable estimates of the requisite parameters? Are existing data sets adequate for both causal and statistical usage? Do new data sources from the other sciences suggest improved techniques for parameter measurement and/or estimation, either for existing formulations or for new models motivated by analyses of these data? What information needs to be collected to meet the requirements of both theoretical and applied hydrology? Which institutions should collect, archive, and disseminate these data? Is hydrologic research constrained by institutional structures and needs, which themselves represent the accumulated scientific knowledge of earlier generations?

5. Can models of sequentially dependent processes (e.g., pollution discharge, transport, dilution, exposure, and physiologic consequence) be linked satisfactorily to form a causal chain of understanding leading to optimal control? Given such uncertainties of each process and of their linkages, can such a causal chain have the predictive power necessary to address the relevant societal issues? Can existing data bases be manageably integrated to formulate and calibrate such chains of models and their parameters?

6. What is the proper role of statistics in developing and assessing models? How far should we press for predictive accuracy and precision? When is a model "good enough" for scientific understanding or engineering applications? At what scale should we strive for understanding hydrologic phenomena, and at what scale is empiricism adequate? How can we encourage the utilization of better understanding by the client institutions?

7. What algorithms should be used to design protective works (e.g., reservoirs, flood walls, levees) and establish nonstructural defenses against hydrologic extremes. Planners often express the need to know "the true density function" of these extremes, but in fact there is little to be gained in this achievement because the short-term sampling instability of rare events is so great that wide errors in estimation are to be expected. The concept of regret should be incorporated. Thus the real question is not "What is the 500-yr flood?" but "We know that an event of size x has occurred. Do we protect against other potential events of size x, $2x$, $5x$, . . .? What are the socioeconomic consequences of protection to these levels?" The definition and estimation of hydrologic extremes and their incorporation into the planning process need improvement.

SCIENTIFIC UNDERSTANDING OF ELEMENTS OF HYDROLOGY

The panel addresses (in the remainder of this volume) the current status of causal or theoretical hydrologic science in terms of its contribution to the management of water resources. There is general agreement that for most surface-water and most groundwater hydrology our understanding of the physics of basic hydraulic processes is reasonably well developed. On the other hand, enhancement of the predictability of related hydrologic components, such as weather and climatic feedbacks (Chapter 2; NRC Geophysics Study Committee, 1977) and runoff relationships (Chapter 1), would be beneficial. In addition, chemical and biological information must be further integrated with our basic understanding to help formulate and apply hydrologic models for predicting hydrologic fluxes and societal impacts.

New research would improve our understanding of *water quality*. Conventional investigations involving indicators and standards of common pollutants, biological well-being, and human health have not typically resulted in fundamental understanding of the water system or the role of the works of man. Improved understanding of the basic chemical and biological processes (see Chapters 5 and 6) would increase our predictive capabilities in addressing widespread and long-term effects of anthropogenic perturbations and natural phenomena and would help bring regulatory activities into focus. It would be marvelous if hydrologists could track a large number of molecules of water and entrained substances to ascertain how they arrived where they are and how long they resided in various storage compartments along the way. It would be even better if their future trajectories could be established, but we recognize the impossibility of conducting such an "interview." However, a careful study of past trajectories might enable us to determine probabilistically some potential anticipated long-term trajectories (in other words, to make predictions).

Despite the firmly developed understanding of the occurrence of groundwater, some misconceptions (e.g., that the water budget of an area determines the magnitude of possible groundwater development) continue to pervade practice with adverse consequences on public water policy (Chapter 4). Our water-management institutions generally do not take into account the interconnections between groundwater and surface water. Further, with respect to the design of groundwater systems for supply or waste disposal, the time horizon is so distant that a customary period of observation or monitoring (several years or decades) will not generally be meaningful for predicting long-term effects. Such predictions, inherently uncertain, must depend on an understanding of the hydrogeologic system as well as the processes--the nature of the transport with its accompanying dispersion and chemical reactions. Because the variances are large, these are high-risk predictions from a statistical and operational viewpoint. Although we have good models for many conventional flow patterns, a particular concern in groundwater is how to treat flow and chemical transport in fractured geologic media--an important issue in the search for adequate methods and sites for waste disposal. These questions need to be addressed if we are to cope with societal matters related to significant usage (i.e., mining) of groundwater.

The *vadose zone*--and several other components of the hydrologic cycle--offers a particular class of problems. The equations governing flow and transport in the vadose or unsaturated zone are of a type for which measurements are especially difficult, so that calibration of various models and of their estimated parameters is extremely tenuous and fragile. The inherent difficulty (see Chapter 3)--or perhaps impossibility in the foreseeable future--of measuring the physical properties within the vadose zone suitably to calibrate theoretical models of its behavior is explicitly recognized. In the absence of such calibration, it is currently necessary, and frequently acceptable for engineering purposes, to accept some associative (lumped-parameter) models in order to make predictions within this domain.

HYDROLOGIC SCALE

One of the most difficult problems in hydrology is that of scale. Current societal problems confronting hydrology are of large temporal and spatial scales heretofore of little practical interest. Hydrologic thinking needs to be extended beyond small catchments or even major watersheds, beyond conventional design horizons and institutional and budgetary constraints. It is necessary to identify the dimensions of time and space along which explanations and predictions must hold if they are to serve satisfactorily.

Much of what we know, and much of what could be proposed to be learned, deals with understanding the various microscale elements in the hydrologic cycle. One of the assumptions frequently made is that our understanding of the microscale elements and processes can, with minor modifications, be extrapolated in principle to an understanding of the macroscale environment, thus enabling reliable predictions to be made by linking the solutions to form a causal chain. Unfortunately, it seldom happens that way. Sooner or later, at some scale or characteristic dimension, mechanistic explanation breaks down and is necessarily replaced by unverified causal hypotheses or statistical representations of the processes. Several papers in this report deal with this scale problem.

Studies tend to focus on hydrologic representations of the fundamental fluxes that connect a small set of storage reservoirs, inflows, and outflows within the watershed; the watershed is typically taken as a convenient unit. It is probably unreasonable to expect that the same set of equations and parameters that "explain" the performance of a small watershed will serve equally well for a large one, let alone the global hydrologic system. We cannot apply to the *whole* the same scientific formulations that are optimal for designing each of its parts. We need some measure of hydrologic scale (analogous to the Froude and Reynolds numbers in hydraulics) to help discriminate among alternative forms of analysis.

We cannot be comfortable with our understanding of critical hydrologic or hydraulic processes because the temporal and spatial scales of our problems are such that even small prediction errors might be levered into catastrophic effects when magnified and extended by the systems as actually constructed, operated, and impacted by unforeseen events. There is therefore an understandable or even laudable tendency to examine in minute detail some of the processes that are presumed to govern these massive systems, the justification being that understanding the microhydrologic processes is a necessary and perhaps sufficient condition for making usable predictions of the performance and effects of macroscopic systems. The problem is that hydrologic analysis has concentrated on systems that are small compared to real or potential perturbations to the hydrologic cycle. Thus it must be asked if causal models are a necessary condition for formulating working hypotheses about macroscopic systems. Perhaps not. This is not to argue against understanding components of a causal chain but rather to suggest that advances are made by an orderly alternation between theory and careful observation.

But changes on the watershed scale exert global hydrologic effect. A change in local evaporation, for example, caused by a change in vegetation cover, results in compensating changes in precipitation elsewhere. Questions of this order raise difficult problems of incommensurate spatial and temporal scales among the linkages between the continents, the atmosphere, and the oceans. These questions have long vexed those concerned with atmosphere-ocean modeling; including the land phase in these models would introduce even more complexity.

A finer scale of inquiry is needed to answer some critical questions, such as nonpoint sources of pollution (Chapter 1), than is now provided by most lumped-parameter watershed models. In these models, total precipitation over the watershed is related to observed runoff and gives results that are satisfactory for many short-run predictions of future runoff or of the flow of nearby ungauged streams. The models neglect considerations of some internal mechanisms and flow paths that govern both the flow and the water quality; and it is argued that these must be understood to predict the long-run effects of changes in land and water use. Fiering and Kuczera (Chapter 7) take an intermediate view, arguing that coupled equations, and fitting procedures designed for them, are efficient predictors.

Some hydrologic processes have time scales on the order of minutes or hours (e.g., atmospheric precipitation), others centuries (e.g., groundwater); some are scaled in meters (e.g., channel width and depth), others in kilometers (e.g., river distances). All of these are linked, and yet it is infeasible for a single model to accommodate range of temporal and spatial scales. The challenge of scale, coupled with the challenge of appropriate utilization of new data commensurate with the scale of inquiry, constitutes an important and exciting frontier in hydrology. The computer is essential in this respect. With an increase of about one million times in the computing speed, we can begin to ask questions for which solutions were essentially unapproachable under former computational constraints. We must begin to think in large-scale, global terms with utilization of as much data as are applicable to provide new theoretical and empirical insights to the solution of hydrologic problems.

HYDROLOGIC MODELING

One branch of hydrologic modeling deals with the frequency of extremes--floods, droughts, and major storms--for which statistical techniques are clearly essential. Decision making in water-resource planning involves a class of operations-research models for guidance in optimal utilization of resources; appropriate use of these models is not challenged here. The division between causal and statistical models does, however, become an issue in considering the prediction or retrospection of the movement of water through a basin and its chemical and biological effects. At one end lie full, causal (conceptual) explanations, based on recognized first principles of the basic sciences. At the other end lie statistical empirical explanations based on analysis of data collected in the basin.

Empirical models of hydrologic processes differ from conceptual models principally in the dimension of the system under study. For example, in the typical empirical representation of the rainfall-runoff process, the characteristic dimension (i.e., the catchment) is large. The selected scale of the representation serves to emphasize relationships between input quantities (current and antecedent precipitation, basin characteristics, groundwater storage, and evaporation potential) and outputs (runoff, discharge, and aquifer flow). The results derived from these models often do not fit the observations well enough for the predictive purposes at hand; therefore, random components might be introduced to accommodate residual variation. In contrast, the potential modification of transmissivity of aquitards and aquicludes over thousands of years of exposure to heat and radiation that could exist in the vicinity of a nuclear-waste repository is an example for which the characteristic spatial dimension is small.

In managing our water resources, we resort to empirical models of phenomena and to a variety of decision-oriented optimization models (many of questionable scientific value) because of imperfect understanding. Some of these are deeply rooted in institutional tradition. This has inevitably led to occasional overenthusiastic reliance on operational mathematics and optimization techniques and a plethora of rainfall-runoff relationships that harbor inconsistencies and incompatibilities with respect to possible physical systems, with inadequate consideration (e.g., the use of linear models for nonlinear processes) of the interactions of hydrology and its related disciplines of geology, chemistry, biology, and meteorology [a form of Gresham's law might be operating (Comar, 1978)].

Empirical model building is valued by many respected hydrologists. The difficulty with such empirical analysis inheres in what is excluded rather than in what is included. For example, global runoff accounts for approximately one third of the incident precipitation, with infiltration and evapotranspiration accounting for the remainder. When the latter two fluxes are ignored in rainfall-runoff relationships, the models are almost certain to fail as predictive tools except in local and short-term applications, where they can be fine tuned to a particular data set. Predictive failure occurs because the omitted flux is about twice as large as that which is included. Moderate errors in the omitted portion are levered into much more substantial errors in the runoff. Perhaps what is needed is systematic use of lumped-parameter watershed models that strive for robust representation of basin performance.

The proliferation of models has led to many inconsistencies (Chapter 8) and incompatibilities among the assumptions incorporated in the models. Moreover, many of the values of hydrologic parameters used in conventional empirical models are generated by an optimization technique. Such estimated parameters may bear little or no resemblance to the physical parameters in the system. Fiering and Kuczera (Chapter 7) offer possible improvements here, but the coupled problems of model specification and parameter estimation and measurement will remain pervasive in hydrology.

Philip (1975) recognized two important points with respect to modeling: (1) that hydrologists deal with models that specialize in, and attempt to relate, processes of enormously different (temporal and spatial) scale and (2) that " . . . hydrologists do not always make good use of what is known." Whereas this appears to be a tautology, Philip suggested the inherent difficulty or perhaps impossibility of developing a widely applicable catchment-runoff scheme based on a few axioms and the potential for use of existing regional data. Philip (1975) suggested a compromise with respect to physical models:

> Any such model suffers . . . from a lack of "robustness of identifiability" and, in consequence, the parametric representation of the system is inefficient and will tend to be unreliable. In these circumstances it may well happen that, when there are sufficient external reasons to regard the system as stationary, a seemingly mindless use of a prediction system based on a regression (with no pretense at a physical model) may well represent a most efficient procedure . . . On the other hand, for dynamic systems where changes in catchment characteristics are known or suspected, a regression-type prediction system is powerless to take account of catchment changes. What, then, can be done? All tnat one can counsel is that the pseudo-physical model should be confirmed and supported by as much "ground-truth" as possible; and that continuous monitoring of catchment characteristics be undertaken. Presumably one must keep the model under continual review and also must maintain a programme of direct physical observations on the catchment.

The difficulty in hydrology is that, if each element of a long causal string of interconnected models requires an enormous amount of insight, work, and data to be validated within the conventional scientific framework, there is still the nagging problem of how to put the pieces together in a meaningful way. This is essentially to divide problems into microscopic and macroscopic scales; the issue is where to make the cut and still preserve the essence of the problem, and this might not coincide with the existing institutional structure. So long as there is the possibility that one of these links will remain weak, we must seek some middle ground that utilizes as much as we know. However, if unifying concepts enable complicated pieces to be linked in an operationally useful way, obsolete formalisms need to be replaced by these new insights.

Operationally useful models are neither black nor white but are varying shades of gray. To suggest otherwise frustrates the decision makers who, when faced with the apparent hopelessness of the situation, might turn to frauds and charlatans with "divining rods" to seek advice. Decisions must be made, and in the absence of a mediating mechanism that allows for an interplay in the scientific process and periodic assessments, incalculable harm could be done. Such a mechanism and the role of periodic assessment and monitoring are critical.

Continued reference to models reflects the inductive nature of hydrologic research whose advances are judged in terms of predictive capability (Chapter 11). Matalas *et al.* note in Chapter 11 that there is no generally accepted definition of causality and prefer to speak in terms of the connectedness between events in space and time. Within this framework the principle of determination admits statistical laws, as well as causal laws.

This Overview argues that what hydrology needs is to (1) integrate the available knowledge; (2) guide programs for data collection--particularly as remote sensing and satellite imagery are beginning to show utility in hydrology; (3) develop techniques for accessing and integrating the various data bases; and (4) make incisive models of those hydro-

logic, chemical, and biological processes that are needed for acceptable predictions. The operational test of a prediction is not necessarily its accuracy or precision, but rather the extent to which it leads to good decisions.

REFERENCES

Comar, C. (1978). Bad science and social penalties (editorial), *Science 200*, June 16.

Fiering, M. B. (1976). Reservoir planning and operation, in *Stochastic Approaches to Water Management, Vol. II*, H. W. Shen, ed., H. W. Shen, Fort Collins, Colo., pp. 17:1-17:21.

Gumbel, E. J. (1954). *Statistical Theory of Extreme Values and Some Practical Applications*, National Bureau of Standards, American Mathematical Society.

Lord, W. B., L. Adelman, P. Wehr, C. Brown, R. Crews, B. Marvin, and M. Waterstone (1979). *Conflict Management in Federal Water Resource Planning*, Institute of Behavioral Science, The University of Colorado, Boulder, Colo., 114 pp.

NRC Geophysics Study Committee (1977). *Climate, Climatic Change, and Water Supply*, National Academy of Sciences, Washington, D.C., 132 pp.

NRC Geophysics Study Committee (1978). *Geophysical Predictions*, National Academy of Sciences, Washington, D.C., 215 pp.

NRC Water Resources Research Review Committee (1981). *Federal Water Resources Research: A Review of the Proposed Five-Year Program Plan*, National Academy of Sciences, Washington, D.C., 75 pp.

Philip, J. R. (1975). Some remarks in science and catchment prediction, in *Third National Symposium on Hydrology; Prediction in Catchment Hydrology*, T. G. Chapman and F. X. Duxin, eds., Australian Academy of Sciences, Canberra.

U.S. Water Resources Council (1977). *Guidelines for Determining Flood Flow Frequency*, Bulletin 17A of the Hydrology Committee, Washington, D.C.

BACKGROUND

Models of Runoff Processes and Their Significance

1

THOMAS DUNNE
University of Washington

INTRODUCTION

Rainwater and meltwater can follow one of several surface or subsurface paths down hillsides to stream channels. The migration of water along one of these paths is called a runoff process, and more than one such process usually operates at any given site. The relative magnitude of flows along each path depends on local climatic, soil, and geologic characteristics and particularly on vegetation cover, as is discussed later.

Runoff processes are the central issue in land-surface hydrology because they affect the volume, rate, and timing of streamflow and therefore the generation of floods and critically low dry-weather flows, the prediction of which has been the subject of much effort. These efforts have responded mainly to design needs in water-resource development and flood control, and their aims were limited. Consequently, it was sufficient to ignore the complex set of processes generating streamflow and to treat the problem of runoff prediction as a statistical problem in which measured output (streamflow) was related to measured input (rainfall or snowmelt) and various drainage-basin characteristics such as area and slope. The resulting empirical relationship could be used for predicting future streamflow in the measured drainage basin or on unmeasured catchments. This method has yielded useful results that formed the basis of a great deal of successful engineering. However, the statistical approach has several limitations.

1. Most hydrologic records are short and unlikely to sample extreme events; the most important design events must often be estimated through extrapolation from an inadequate set of data.
2. Climatic fluctuations and human influences in the hydrologic cycle constantly alter the rainfall-runoff relationship in some regions, so that even long records do not provide a homogeneous set of data for statistical analysis.
3. The statistical approach provides little or no information about the physical characteristics or spatial distribution of runoff within a drainage basin. Recent interest in the scientific aspects of hydrology and in the application of hydrologic studies to a wider range of societal problems have focused attention on these runoff characteristics.

Runoff from hillsides entrains and transports into channels the sediment, chemical, organic debris, bacteria, viruses, and other materials that affect the characteristics of streamwater. On its way to the channel, runoff is associated with various degrees of soil wetness and pore pressure, which affect such important land characteristics as suitability for agriculture or waste disposal, trafficability, and stability against landsliding.

Human activity frequently alters the intensity of runoff mechanisms or introduces a process that did not occur in the area before disturbance. Thus, changes in land use are often accompanied by increases in the size of floods on small streams, waterlogging of soils, and landsliding or high concentrations of sediment, chemicals, or biological pollutants in streamwater. The disturbances spread and their effects accumulate while subject to a random array of meteorological events. Consequently, it is exceedingly difficult to predict the influences of human activity on the basis of empiricism alone, because it is not possible to compile adequate data sets through repetition of controlled experiments. Various federal agencies have expended a great deal of effort on such experiments in small drainage basins, but the statistical limitations of the experimental designs and the inherent difficulty of interpreting spatially distributed processes on the basis of measurements at the basin outlet alone severely restrict the conclusions that can be drawn from such experiments. Thus, after 30 yr of such research there is still controversy and confusion about such issues as the influence of forest management on flood peaks, soil nutrient status, sediment production, and water quality.

Many questions concerning land and water management are likely to be of this type in the next few decades. For example, the problem of predicting nonpoint sources of phosphorus, nitrogen, or bacteria in streams and lakes or of the wash-off of heavy metals and other pollutants in urban storm drainage requires that the amount and path of runoff and its hydraulic characteristics can be predicted. A statistical relationship that would predict peak runoff is not adequate. The sources and characteristics of the runoff and what it can entrain and transport to the stream need to be known. A model of these processes would need to account for spreading of the urbanized area in a drainage basin and the impact of strategies for modulating runoff rates.

Another example is the problem of predicting some effects of large-scale exploitation of tropical landscapes, such as the Amazon Basin. Concerns such as the potential for physical and chemical degradation, the alteration of biogeochemical cycles, and the stream transport of organic materials involve runoff. Yet, only one small experimental investigation of runoff processes has been conducted in the 6.1-million km^2 Amazon Basin (Northcliff *et al.*, 1979) and only several in tropical forests in other parts of the world (Bonell and Gilmour, 1978; Leigh, 1978). We have virtually no basis for making predictions about the hydrologic consequences of disturbance; about the probable distribution of accelerated erosion; or of leaching, waterlogging, or any other impact.

In this and many other situations the consequences are likely to be predicted on the basis of simple statistical relationships that incorporate coefficients to predict volumes or peak rates of runoff, amounts of soil loss, or other requirements. These conceptually simple relationships are often applied to small portions of drainage basins and added together or otherwise manipulated by high-speed computers and presented as complex models. Close examination of such calculations frequently indicates that the application is fundamentally wrong. The relationships are often developed with the wrong runoff process or erosion and transport mechanism in mind, or the parameters are unreasonable in light of field observations or physical possibilities. Despite the fact that such calculations may not describe what is happening in the field, they are often defended with such arguments as: "the procedure can be calibrated against field measurements so that it yields correct answers," or "we have to have a simple uniform procedure that can be applied in all situations to meet the requirements of the law," or "we do not have the resources to train our prediction personnel in field observation or data collection."

If prediction efforts continue along these lines they will lead to expensive failures and mismanagement of land and water resources. Errors are most likely to appear in the prediction of extreme events or disruptions of kinds that have not been well documented. There will also be a lack of intelligent approaches to problems other than prediction of some streamwater characteristic at the basin outlet. In many cases the manager or policymaker may be less interested in a precise calculation of (say) peak runoff than in anticipating the characteristics of certain zones of a landscape under various hydrologic scenarios or where to collect samples to monitor changes. Answers to this kind of question depend on a knowledge of the physics of the land phase of the hydrologic cycle.

During the past 20 yr an alternative approach to hydrologic prediction has emerged through a combination of field experiments and mathematical models. These developments concentrate on the physics of runoff and of the erosion and transport for which it is responsible. They permit predictions not only of basin outflows but of the spatial distribution and characteristics of runoff. This chapter reviews the current status of field observation and physically based modeling capability.

DESCRIPTION OF RUNOFF PROCESSES

Horton Overland Flow

The paths that water can follow to a stream are indicated schematically in Figure 1.1. An important separation occurs at the soil surface, which can absorb water at a certain maximum rate known as the infiltration capacity. This rate is relatively high at the onset of rainfall but declines rapidly to an approximately constant value after about 0.5-2 h. If rainfall intensity at any time during a rainstorm exceeds the infiltration capacity of the soil, water accumulates on the surface, fills small depressions, and eventually spills over to flow downslope as an irregular sheet. This runoff is known as Horton overland flow (path 1 in Figure 1.1) after Robert E. Horton (1933, 1940), who first developed a theory of the relationship between infiltration and runoff and their consequences for land and water management and who also conducted some of the earliest experiments on these processes. The occurrence of Horton overland flow depends mainly on the surface characteristics that control infiltration, the most important of which are vegetation and soil conditions. Dense vegetation cover protects the soil surface from the packing effect of raindrops, provides organic material, and promotes biological activity that forms open stable soil aggregates. Coarse-textured soils can absorb rainfall at a higher rate than silty or clayey soils. Land use often involves the reduction of vegetation cover and the breakdown of soil aggregates, with a consequent lowering of infiltration capacity and an increase in the frequency and amount of Horton overland flow. In the extreme case of urban land use, the infiltration capacity of large fractions of the land surface is reduced to zero.

Horton overland flow tends to produce a major portion of the total runoff in arid and semiarid landscapes and in disturbed zones of humid landscapes, such as cultivated fields, paved areas, mine spoils, construction sites, and rural roads. These areas lack a dense vegetation cover and well-aggregated topsoil. In areas where Horton overland flow is the dominant producer of storm runoff, large areas of hillside commonly have a low infiltration capacity; but because infiltration may vary from place to place over a surface, the runoff is not distributed uniformly. An important problem in runoff prediction is the determination of whether overland flow will occur on a certain portion of the landscape in a particular rainstorm, as well as the amount that will be generated. Identification of runoff-producing zones is also important for recognizing areas that are susceptible to erosion and the washing off of pollutants. Betson (1964) referred to the nonuniform generation of Horton overland flow as "partial-area runoff."

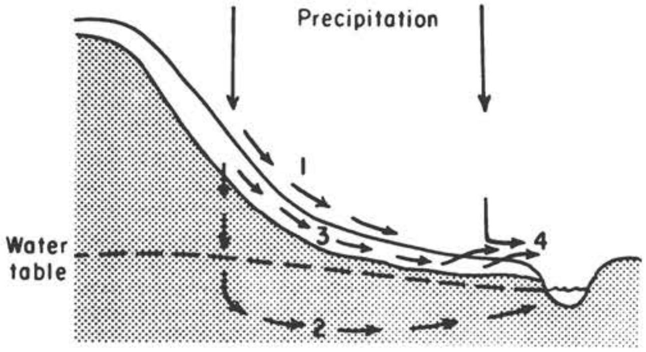


FIGURE 1.1 Paths of runoff from hillsides.

Horton overland flow typically travels over hillslope surfaces at velocities of 10-500 m/h, so that the whole of a 100-m-long hillside can contribute runoff to a stream channel in rainstorms of 0.2-10 h duration. This rapid influx to the channel is responsible for a rapid response of streamflow to rainfall and high peak rates of stream discharge relative to those from basins of similar size but subject to different runoff processes. Prediction of response time is another important goal of runoff modeling.

Subsurface Flow

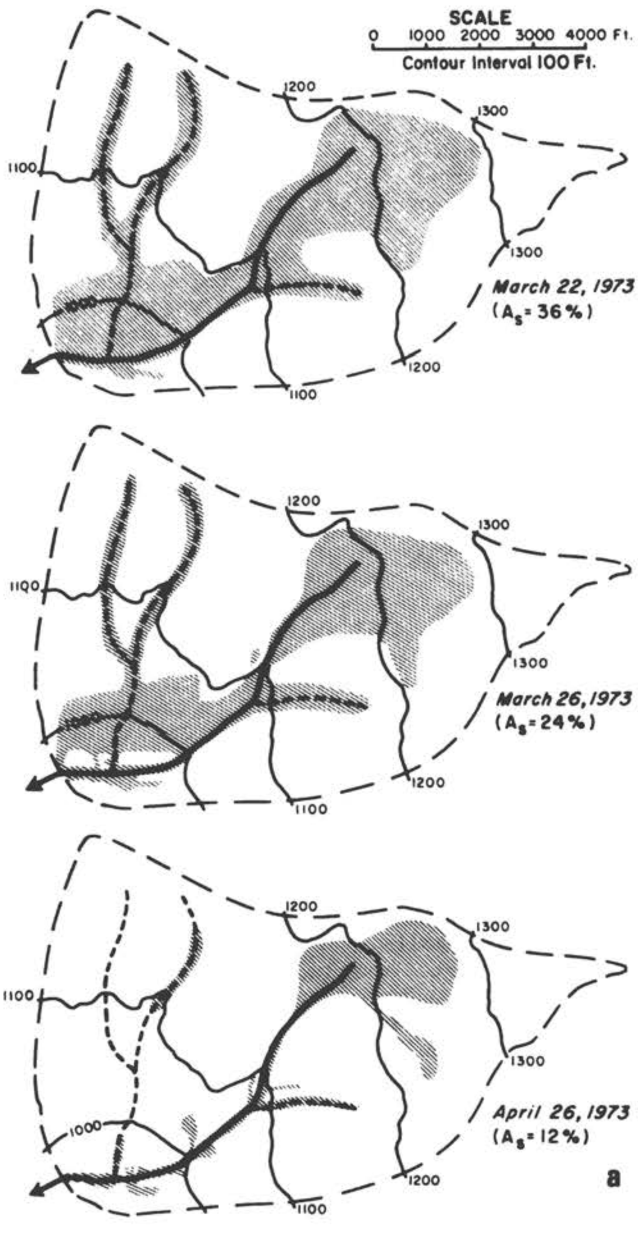
In densely vegetated, humid regions the infiltration capacity is usually high enough to absorb all but the rarest, most-intense storms. If the soil and underlying rock are deep and permeable, the infiltrated water percolates through the soil into the vadose zone (see Chapter 3) and then follows a curving path through the phreatic zone to stream channels. Typical rates of this groundwater flow (path 2 in Figure 1.1) are many orders of magnitude slower than velocities of overland flow, and the subsurface paths are generally long, so that groundwater supplies dry-weather streamflow and is a less important con-tributor to storm runoff. (The characteristics, significance, and modeling of this runoff process are examined in Chapters 4 and 5.)

In many soils or rocks, water percolating vertically encounters an impeding horizon and is diverted laterally. It migrates downslope through the soil and displaces into the stream channel water that is in storage and has been migrating slowly downslope during the preceding days or weeks. If this shallow subsurface flow (path 3 in Figure 1.1) is generated under a low rainfall intensity onto a highly permeable soil, the water may travel downslope without saturating the soil. Flow is confined to the narrower, intergranular pores, and its velocity is usually of the order of 10^{-4} m/h or less. Under intense rainstorms the soil often becomes saturated at some depth. Water is then able to migrate through the largest pores, rootholes, wormholes, and structural openings, and its velocity may increase up to about 0.2 m/h in highly permeable forest soils on steep slopes but to much lower values in most soils. Although it travels more slowly than Horton overland flow, some of this runoff arrives at the channel quickly enough to contribute to floods and is classified as subsurface stormflow (Whipkey, 1965). But small streams fed dominantly by this process respond an order of magnitude more slowly to rainfall than those with similar drainage area receiving Horton overland flow, and their peak rates of runoff tend to be more than an order of magnitude lower (Dunne, 1978). After a rainstorm, shallow subsurface flow declines and may cease altogether in some soils. Hewlett (1961a, 1961b) showed, however, that in a mountainous region underlain by impervious rock, slow drainage of water from unsaturated soils may be responsible for dry-weather flow also. Recent important work has been reported by Harr (1977), Anderson and Burt (1977, 1978), and Anderson and Kneale (1980).

Saturation Overland Flow

Vertical and horizontal percolation may cause the soil to become saturated throughout its depth on some parts of a hillside. The soil saturates upward from some restricting layer, and when saturation reaches the ground surface, rainfall cannot infiltrate but runs over the surface as *saturation overland flow*. Some of the water moving slowly through the topsoil may emerge and flow overland to the channel as *return flow* (Musgrave and Holtan, 1964). Thus, *direct precipitation onto the saturated soil*, with or without return flow, generates *saturation overland flow* (path 4 in Figure 1.1), which is intimately related to subsurface runoff and soil conditions. The process occurs most frequently on gentle footslopes and hollows with shallow, wet soils (Dunne, 1970) but can also occur on other parts of a hillslope with thin or wet soils. During a storm the saturated zone may spread upslope and especially into topographic hollows and zones of shallow, wet, or less permeable soil. Slower fluctuations of the zone producing overland flow result from seasonal changes in wetness. Figure 1.2 provides field examples of such fluctuations. Saturation overland flow usually moves more slowly (0.3 m/h to less than 100 m/h) than Horton overland flow because it travels through dense ground vegetation on low gradients. It generates streamflow with a lag time and peak discharge that are intermediate between those characteristic of Horton overland flow and of subsurface stormflow.

It is now generally agreed that in densely vegetated humid regions, most storm runoff is generated by a combination of shallow-subsurface stormflow and saturation overland flow, which in turn consists of various proportions of



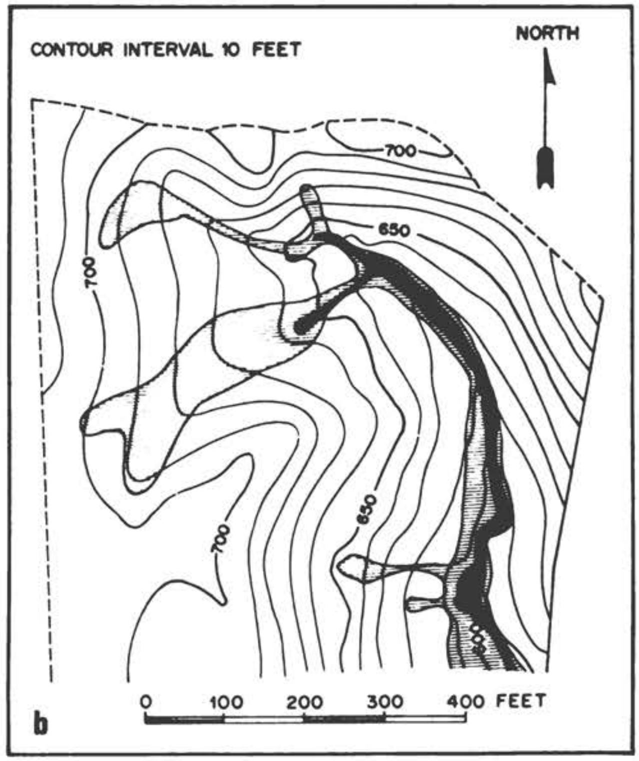


FIGURE 1.2 (a) Seasonal variation of the saturated zone in a catchment at Randboro, Quebec. (b) Expansion of the saturated area during a 46-mm rainstorm in a catchment in northern Vermont. The solid area indicates the saturated zone at the beginning of the storm; the lined area indicates that zone at the end of the storm (from Dunne *et al.*, 1975).

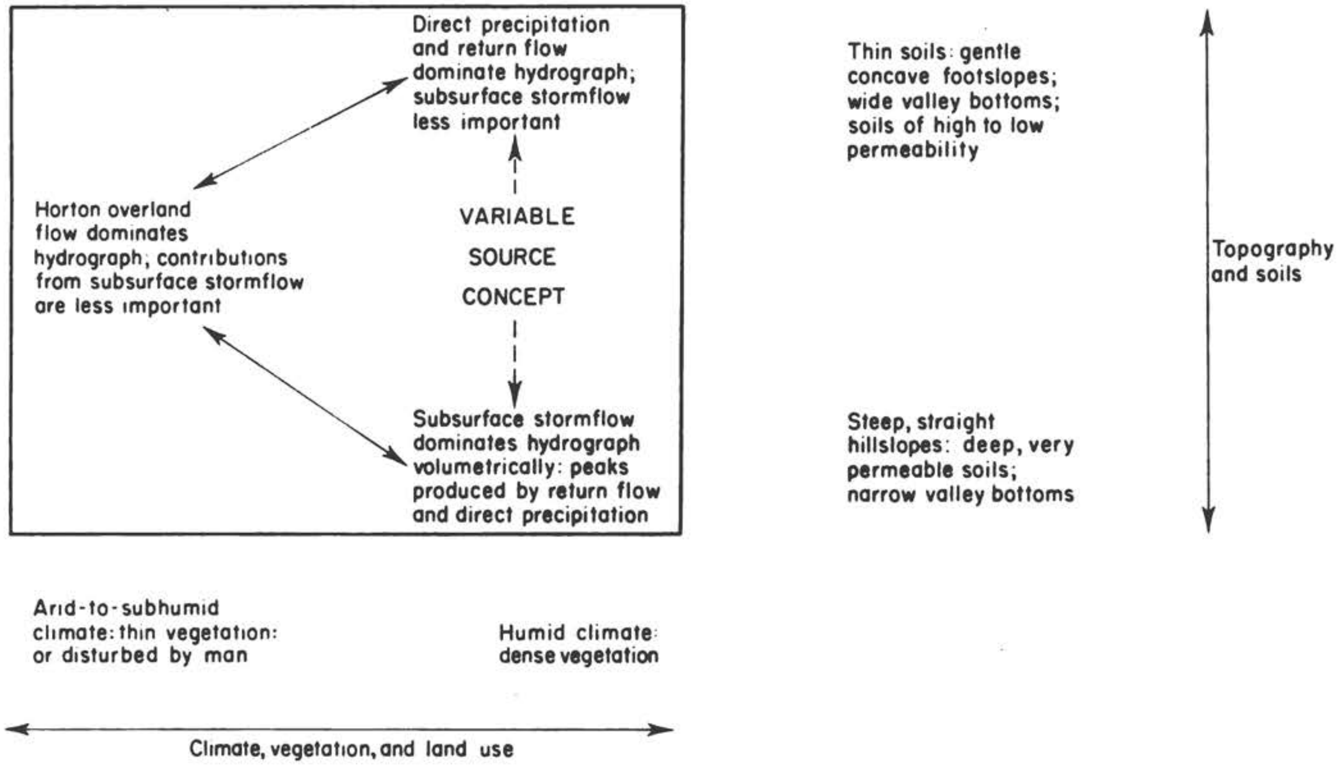


FIGURE 1.3 Schematic summary of the occurrence of runoff processes in relation to their major controls (after Dunne, 1978).

return flow and direct precipitation onto the saturated area. The relative contributions of subsurface stormflow and overland flow vary with soil and topographic conditions, and it is important that mathematical models be able to separate these contributions not only for prediction of the stream hydrograph but because the processes are associated with differences in soil drainage, entrainment of pollutants, and other facets of hillslope hydrology that are important in resource management. Where deep, highly permeable soils border narrow-valley floors, almost all of the infiltrated water can travel beneath the soil surface to the stream. The stream hydrograph is dominated by subsurface stormflow, although the initial rapid response of streamflow to rainfall is often supplied by saturation overland flow from the saturated valley floor. Hewlett and Nutter (1970) and Harr (1977) have described conditions in forested mountains of the southeastern and western United States where subsurface flow is the dominant runoff mechanism. At the other end of the spectrum in humid landscapes are those regions that have long, gentle, concave footslopes covered with shallow, wet soils of low-to-moderate hydraulic conductivity. In these areas, soils become saturated more frequently and the area generating saturation overland flow may expand from less than 10 percent of the basin to more than 50 percent during a storm or a wet season. Figure 1.3 summarizes the conditions affecting the dominant runoff processes in a region.

The temporal and spatial variation of zones producing storm runoff by subsurface flow led early investigators to refer to their conceptual models of the process as the *variable-source-area concept* (Hewlett and Hibbert, 1967) and the *dynamic-watershed concept* (Tennessee Valley Authority, 1964). Most fieldwork on hillslope hydrology in humid regions is now focused on this concept, and a summary of progress up to about 1974 is provided in the book edited by Kirkby (1978).

Snowmelt Runoff

In cold-temperate and subarctic regions, snowmelt runoff generates floods, recharges groundwater, and renders large areas of land temporarily uncultivable. When the economic and social impacts of snowmelt runoff are considered, as well as the sophistication of research into the physics of the melting process, it is almost inconceivable that so little effort has been spent on investigations of runoff processes during snowmelt. Yet differences in the path of runoff can have much greater impacts on the timing and peak rates of runoff than do many of the melt parameters that have been studied in detail.

At a recent state-of-the-art conference on modeling of snow-cover runoff (Colbeck and Ray, 1978), only 3 papers (out of 41) reported on field investigations of runoff processes. The difficulty and discomfort of making the necessary measurements probably contribute to this dearth, but the importance of the subject was attested to by several other papers that dealt with snowmelt-runoff prediction without the assistance of field observations.

Major advances have been made recently in understanding the physics of snowpack metamorphism and the percolation of meltwater through the pack (Colbeck, 1971, 1978). After meltwater reaches the ground surface it follows one or more of the paths outlined earlier, but several factors may alter the relative importance of the various runoff processes from the situation under rainfall at the same site. Snowmelt releases large quantities of water more slowly than in most important flood-producing rainstorms, so that the probability of the infiltration capacity being exceeded is low. Under some circumstances, however, a dense layer of ice may accumulate in the topsoil, rendering it almost impermeable. The equivalent of Horton overland flow occurs as saturated percolation through the lower layers of the snowpack. Horton (1938) was the first to recognize the importance of this process and to propose an approach to its computation and formal study. Colbeck (1974) presented a more refined theory, which was applied to a field investigation by Dunne *et al.* (1976). The opposite extreme of complete infiltration and the generation of subsurface stormflow was studied by Stephenson and Freeze (1974), and Dunne and Black (1971) documented an intermediate case of both surface and subsurface flow. The current state of knowledge on this subject was summarized by Wankiewicz (1978a) and Price *et al.* (1978), but much remains to be done, particularly in areas of ephemeral soil freezing or saturation.

PHYSICALLY BASED MATHEMATICAL MODELS

General Statement

Freeze (1974) pointed out that physically based mathematical models of runoff processes involve solving boundary-value problems based on the partial differential equations that describe the various kinds of low described in the preceding section. Definition of the problem (Freeze, 1974, p. 634) requires that the following be available:

1. Size and shape of the flow region;
2. Equation describing flow in the region;
3. Boundary conditions around the margins of the region and their spatial and temporal distributions;
4. Initial conditions and their spatial distribution;
5. Spatial and temporal distributions of the hydraulic parameters that control flow;
6. A mathematical method of solving the flow equations.

If a runoff problem involves more than one process it is necessary to specify the region over which each process occurs and to ensure that the boundary conditions in each region are compatible with those of the adjoining region. Current computer capacity severely constrains the treatment of three-dimensional, time-varying flow, but fortunately most hillside and channel flow regions can be treated as one- or two-dimensional units.

The simplest runoff problems, such as the calculation of Horton overland flow from a planar hillside with uniform surface characteristics, can be solved analytically using a technique known as the kinematic-wave approximation (see Woolhiser, 1975, for a review). Simplifications of this type are useful, for example, in urban storm-drainage design and even for computations on some natural hillsides. Most runoff problems require that solutions be obtained by numerical techniques that can be used for transient processes within irregular boundaries and for surfaces or media with heterogeneous hydraulic properties (Freeze, 1974, 1978; Beven, 1975). The models are capable of specifying not only the outflow from a hillside or stream channel but also the spatial and temporal distribution of overland flow, soil moisture content, pore pressure, and several other hydrologic consequences of rainfall or snowmelt.

Infiltration

In the first half of the twentieth century, many empirical equations were used to describe the frequently observed temporal variation of infiltration capacity during a rainstorm. At the beginning of an intense storm the infiltration rate is high but declines rapidly after the onset of rain and approaches a constant value. The most widely used of these empirical equations is that of Horton (1933), which has proven adequate for many simple tasks. The three parameters of the equation can be obtained from a relatively simple field test under artificial rainfall, although they are often highly variable between tests on a single soil type.

The need to explain such variability and to understand the effects of land treatment and other controls of infiltration such as rainfall intensity and antecedent moisture has stimulated the development of equations describing the physics of the process. The earliest such attempt was by Green and Ampt (1911), who proposed that under ponded conditions the infiltration rate depends on the saturated hydraulic conductivity of the soil, the depth of ponded water, the depth of the zone of saturated soil, and the pore pressure in the water at the wetting front. Because the underlying physical model was a crude approximation of the correct process, only the general form of the equation was used as a guide for determining two empirical constants from field measurements. However, the simplicity of the equation and the fact that it has some physical basis make it attractive for incorporation into models of runoff generation.

Therefore, in recent years, the Green-Ampt equation has received detailed study, and important advances have been made in interpreting the equation on the basis of a better understanding of soil physics (e.g., Mein and Farrell, 1974; Mein and Larson, 1973; Morel-Seytoux and Kanji, 1974; Smith and Parlange, 1978). There is a need for careful, detailed field measurements of infiltration and soil moisture behavior in order to evaluate the performance of these equations over a wide range of field conditions. Morel-Seytoux (1976, 1978) examined the combined effects of air and water movement during infiltration and developed a relatively simple and practical physically based method of predicting infiltration during rainstorms of varying intensity. The equation will probably be used in many runoff models in the near future.

The most ambitious attempts to develop physically based infiltration models involve solution of the equation for one-dimensional unsaturated flow, which combines the continuity equation with Darcy's law specifying the velocity of flow under a potential gradient in a porous medium. In the latter equation, the hydraulic conductivity and pore pressure in the porous medium depend on moisture content and, consequently, vary during the rainstorm. Philip's (1957-1958) classic works on the physics of infiltration provide some analytical solutions to the unsaturated flow equation and a practical prediction tool for relatively simple conditions.

Following the work of Rubin (1966), many infiltration models are now based on numerical solutions of the flow equation, and it is possible to analyze more complicated and realistic field situations such as infiltration into layered soils under varying rainfall intensity. Freeze (1969), Smith (1972), and many others have computed infiltration and the resulting profiles of moisture and pressure into uniform soils and have explained how the intake rate declines as a result of capillary suction as water is stored in the topsoil. Eventually, the limiting rate of infiltration declines to the saturated hydraulic conductivity of the topsoil. The computations are usually based on laboratory measurements of the relations between moisture content, pore pressure, and hydraulic conductivity of the soil. Some field measurements of these relations are also available (Davidson *et al.*, 1969). The measurements are time-consuming to make and are not available for most soils, but Brooks and Corey (1966) compiled statistical summaries of soil properties from which it is possible to make a rough estimate of the hydraulic properties of a soil from its texture. Computation of infiltration into layered soils by numerical methods increases the data requirements. Another complication arises where soil properties vary systematically or randomly along a hillslope profile (Freeze, 1980; Philip, 1980).

However, two important effects are ignored in these calculations. The first is the influence of plant root-holes and other structural openings that are not usually sampled in laboratory measurements of conductivity. In many soils these effects are probably small and do not invalidate or reduce the utility of models of porous-media flow. However, the effect of flow down large voids in cracking clay soils (Bouma *et al.*, 1978; Dunne and Dietrich, 1980) is not dealt with in current infiltration models. A second complication involves the development of crusts of exceedingly low conductivity that are formed on the surface of thinly vegetated soils by raindrop dispersal of soil aggregates and the plugging of the surface by washed-in clay and silt particles (McIntyre, 1958a, 1958b). Although it is possible with a great deal of difficulty to obtain field measurements of the necessary hydraulic parameters to model these processes, it is unlikely that such models will be used routinely for the infiltration component of runoff predictions. It is easier to make a direct field measurement of the infiltration capacity using simple equipment. In this case the primary value of the physically based models lies in formalizing our understanding of the infiltration mechanism and its controls and in guiding field-measurement programs, interpreting results, and extrapolating them to unmeasured sites.

Horton Overland Flow

The depth, velocity, and discharge of Horton overland flow are calculated on the basis of equations that describe flow in a shallow channel. These equations involve statements of the conservation of mass and of momentum (the "shallow-water equations") and a relationship, such as the Darcy-Weisbach equation, between the velocity and depth and slope of the water surface. These last three variables are related by an empirical coefficient (the Darcy-Weisbach friction factor) that represents the flow resistance over the particular hillslope surface. The friction factor is an important parameter that varies with the Reynolds number (discharge per unit width of hillside divided by the kinematic viscosity of water) and therefore ranges over two orders of magnitude or more along a hillside. The friction factor also varies with the rainfall intensity impinging of the flowing sheet, the vegetation cover, the texture of the soil surface, and the gradient. The flow equations are solved numerically, either in their complete form or after simplification by means of the kinematic-wave approximation described in detail by Woolhiser and Liggett (1967). This approximation involves the assumption, reasonable under most circumstances, that the rates of change of water depth and velocity head along the hillside are negligibly small compared with the ground-surface slope.

In some applications of these runoff models, serious difficulties can arise in specifying the geometry of the flow region and the hydraulic resistance. The geometry of the hillslope is approximated as a plane or as a set of contiguous planes (called a "cascade," see Figure 1.4), which have various lengths, gradients, and hy-

FIGURE 1.4 Schematic illustration of a kinematic cascade used for computing overland flow from a convexo-concave hillslope. The hillslope indicated by the dashed line, and its approximation as a series of planes, is represented by the lower solid line. The shaded area shows the sheet of overland flow.

draulic resistances. On these surfaces, Horton overland flow is assumed to move directly downslope as a laterally uniform sheet. On some short smooth hillsides, field observations indicate this to be a reasonable approximation, and it is possible to conduct plot experiments under simulated rainfall and to measure flow depth, velocity, and slope and to compute the friction factor over a range of Reynolds numbers for a variety of hillslope surfaces (T. Dunne and W. E. Dietrich, University of Washington, manuscript). On other hillsides, microtopography and clumpy vegetation cause the water to converge, diverge, and meander to such an extent that it is not possible to define a meaningful average depth in the field. Emmett (1970) conducted a valuable study of the hydraulics of overland flow in hillsides in Wyoming and documented the variability of flow characteristics but was unable to generalize about the hydraulic characteristics of the sheetflow in a form that could be used in a mathematical model and transposed to other sites. Other natural hillslopes that have been represented by smooth, regular surfaces are intricately dissected by gullies (Singh, 1976). When such approximations are made, it is obviously not possible to specify the average hydraulic resistance at some distance along a hillslope, except in an artificial way. It must be assumed that for the scale of the drainage basin under study, flow conditions are laterally homogeneous, even if locally irregular, at some distance from a major ridgetop. The most common way to obtain an answer is to choose by trial values of the friction factor that provide a good fit between computed and measured discharge from a hillside or small watershed (Woolhiser *et al.*, 1970). The friction factor thus obtained may be a single value for an entire watershed or may vary with the Reynolds number along a hillslope profile.

Several writers have suggested that this is the best that can be done in the computation of overland flow, and Woolhiser (1975) has proposed that overland flow be defined as that phenomenon as represented mathematically by the shallow-water equations and a fitted resistance law. Such an approach is an adequate basis for studying the physics of runoff that is well represented as a sheetflow and is adequate for the purpose of computing storm hydrographs, although it has led to several notorious examples in the literature of subsurface stormflow being modeled as overland flow with a high resistance coefficient. Such mistakes could be avoided by a modest amount of field experience on the part of the user of such a model. For rougher hillslopes or surfaces with rills and gullies, the shallow-water equations provide only an abstraction that is useful for computing the aggregate discharge from many small channels. There is probably little to be gained from formulating more detailed models of hydraulic resistance and lateral variation of flow if one is concerned only with runoff rates. However, this artificial representation is not adequate for a detailed, physical understanding of such channelized runoff or for certain applications of runoff models in geomorphology and soil erosion research or the study of nonpoint-source pollution. For these purposes, the degree of flow convergence, or even rilling and gullying, is important, as are the hydraulic properties of the highly nonuniform flow. Useful description and summary of these characteristics will probably be difficult, but some progress is necessary on this important topic (Foster and Meyer, 1975). This is a matter on which computational techniques are clearly ahead of understanding and field measurements, and new concepts need to be developed on the basis of new field observations.

Smith and Woolhiser (1971) combined a mathematical model of infiltration described earlier with a model of overland flow based on the kinematic approximation of the shallow-water equations. Their method successfully reproduced runoff hydrographs from a natural hillside plot, but a considerable amount of estimation and parameter fitting was required to obtain the necessary hydraulic parameters for both the infiltration and overland-flow components. Yet the hillside was on a research station and had already been the subject of a better-than-average soil-measurement program. The complexity and expense of developing numerical, physically based models does not reduce their value in the investigation of the fundamentals of runoff processes nor in certain important engineering applications. The complexity allows meaningful representation of complex processes and allows one to deal with spatial variation in types and intensity of processes. However, for application to field sites, these models require more data than are usually available. One response to this paucity should be the collection of more extensive, relevant data on the hydraulic properties of soils and the resistance characteristics of hillslope surfaces, for example. These data should be related to easily observed textural and morphological features. An alternative strategy, used successfully by Smith and Woolhiser (1971) and others for predictions of hillslope and watershed discharge, is to back-calculate the necessary parameters through trial-and-error fittings of predictions to runoff records. The estimated parameters can then be used for the prediction of discharge in other storms.

A simpler model of infiltration and overland flow was described by Engman and Rogowski (1974). They used Philip's (1969) approximate, physically based method of computing infiltration when the soil surface is suddenly ponded. The Philip equation requires knowledge of the soil-moisture content at the beginning of rainfall and the relationship between hydraulic conductivity and moisture content. Spatial variations of infiltration and therefore of overland-flow generation were allowed, in keeping with Betson's (1964) partial-area concept of runoff referred to earlier. If the initial soil moisture of soil hydraulic properties varies through the basin, the model accounts for expansion of the area contributing Horton overland flow. Rainwater accumulating on the soil surface is routed downslope using the kinematic wave method with a fixed hydraulic resistance coefficient obtained from the literature. The Engman-Rogowski model is designed to be used with data obtained in routine soil surveys, which suggests the direction in which much applied modeling research needs to be focused in the immediate future, although even with this approach large amounts of data are required for field application, and collection of such data is slow and expensive.

Subsurface Flow

The prediction of subsurface flow to supply both stormflow and dry-weather streamflow involves numerical solution of the continuity equation and Darcy's law describing time-varying, unsaturated-saturated flow in a heterogeneous, anisotropic porous medium, such as a layered soil or rock. The models require that the geometry and hydraulic properties of the soil horizons or geologic strata be known in considerable detail. The computer capacity required for the computations is much larger than that needed for modeling overland flow, and only two-dimensional subsurface cases are tractable with present computer storage capacity, although the three-dimensional case has been addressed by Freeze (1971). Output from these models includes the rate of subsurface flow to streams, the configuration of the water table, and the distributions of water content and hydraulic head within the porous medium.

Models of this kind were developed by Freeze (1971, 1972a, 1972b), who used finite-difference methods of solution, and by Beven (1975), who employed the finite-element method. The models were applied to a wide variety of hypothetical situations and contributed to a unified interpretation of many field experiments on runoff processes in different environments. A particularly enlightening application was that by Stephenson and Freeze (1974), who attempted to simulate the runoff mechanism and hydraulic potential field resulting from snowmelt on an instrumented hillslope in Idaho. The results were encouraging in some respects but cautionary in others. The two investigators, closely involved with the field-data collection and with the details of model construction, had to use considerable ingenuity and experience to estimate some of the model parameters, even though the site has already undergone detailed investigation. Most of the problems involved specifying the geometry of the subsurface flow region and the hydraulic properties of the soil and rock.

Stephenson and Freeze (1974) concluded that the application of the currently most sophisticated, physically based model of subsurface flow was useful for interpretation of field measurements and for defining mechanisms of flow so that simpler, more approximate physically based models could be based on a firmer understanding of processes. Although numerical models of saturated groundwater flow are used routinely (see Chapter 4), Stephenson and Freeze argued that the transient unsaturated-saturated model of runoff processes is not useful for routine applications. They pointed out limitations in the theoretical development, the difficulties of model calibration, the constraints imposed by the storage capacity of the current generation of computers, and the enormous task of collecting sufficiently detailed field data.

Simpler models of subsurface flow have been introduced, for example, by Beven and Kirkby (1979), but they represent runoff processes by rudimentary analogies governed by highly simplified relationships. For example, in the case referred to above, it is proposed that subsurface flux is simply an exponential function of soil water content, and the boundaries of the subsurface flow region are not well defined. Parameters for the model must be obtained from extensive but simple field measurements. These models are, however, useful for some tasks, such as in those cases where the need for simple computations of whether a portion of a watershed will yield significant subsurface flow is more important than obtaining a correct physical representation of that process.

Saturation Overland Flow

Because of its close association with shallow subsurface flow, saturation overland flow has been computed by the transient unsaturated-saturated model of Freeze (1972b). The predicted water table rises to the surface over an expanding zone at the base of a hillside and creates a seepage face from which soil water emerges as return flow (see Figure 1.5). The width of the seepage face perpendicular to the contour determines the saturated area on which direct precipitation contributes to saturation overland flow. Precise definition of the occurrence and spread of saturation overland flow requires a closely spaced network of nodes in the numerical model of the topsoil and, therefore, increases the required computer capacity, computation time, and data requirements.

The relative importance of saturation overland flow to the hillslope hydrograph (Figure 1.5) depends on the ratio of rainfall intensity to the saturated hydraulic conductivity of the soil, the duration of rainfall, and the volume and geometry of the subsurface flow region, including the effects of layers. The Freeze

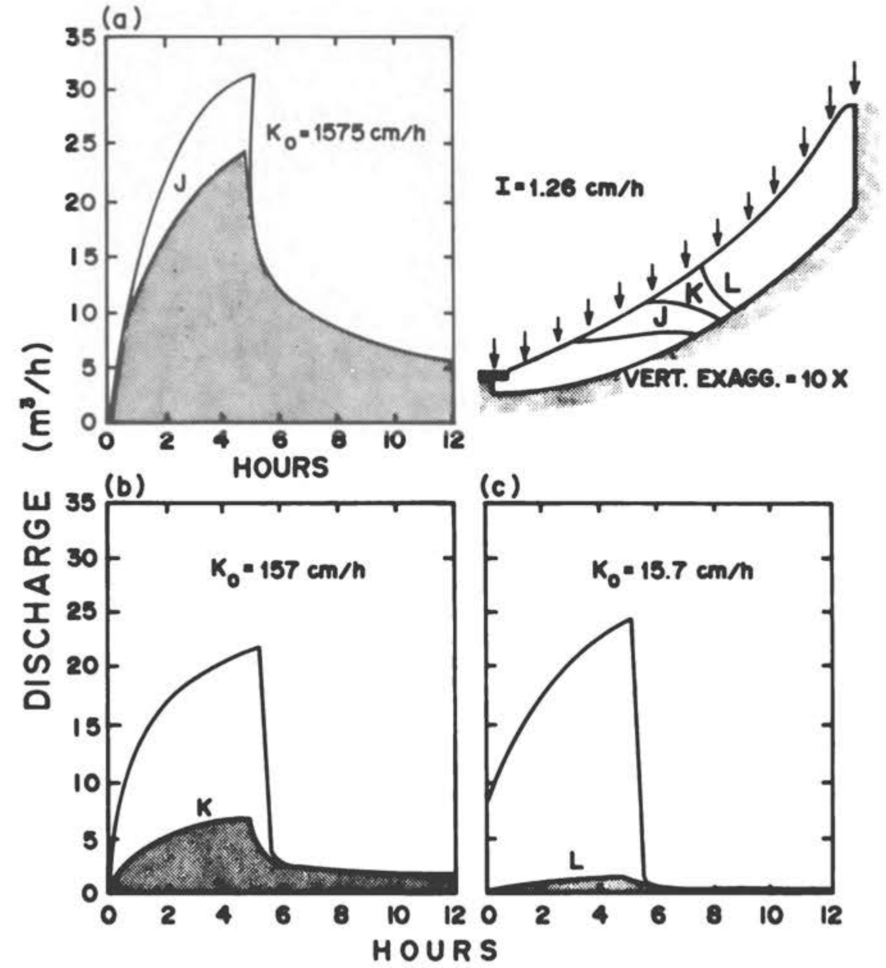


FIGURE 1.5 Calculated hydrographs of subsurface stormflow and return flow (shaded zone) and direct precipitation onto saturated areas (unshaded zone) for a 5-h rainstorm with an intensity of 1.26 cm/h, on soils with various hydraulic conductivities (K_0). The hillslope and the position of the water table at the end of rainfall are shown for each case in the inset (after Freeze, 1972b, reproduced with permission of the American Geophysical Union).

(1972b) hydrograph calculations shown in Figure 1.5 do not separate return flow over the saturated zone from subsurface stormflow, although such a separation is within the capabilities of the model. This is an important distinction that such models should make because of the ability of return flow to wash soil and pollutants from the hillside. Field measurements indicate that return flow may comprise up to at least 50 percent of the peak rate of saturation overland flow (Dunne, 1970). Refinements are also needed in the calculations of velocities of saturation overland flow. In the calculations by Freeze the hypothetical hillslopes were so steep and the saturated zones so narrow that it could be reasonably assumed that overland flow is evacuated from the surface as soon as it is generated. Some saturated zones on natural hillsides have widths exceeding 100 m (Dunne *et al.*, 1975) and gradients of only 1 or 2 percent and are densely vegetated. Flow velocities are often less than 10 m/h under such circumstances, and there may be a considerable lag between a burst of rainfall and runoff.

For reasons referred to above, the detailed physically based model of saturation overland flow is unlikely to be used routinely in the foreseeable future. The simpler approach of Beven and Kirkby (1979) involves dividing a basin into small units (areas of 200-2000 m^2) on the basis of soil and topography. On each unit, runoff may be generated by Horton overland flow, subsurface flow, and direct precipitation onto saturated areas, and the model predicts the time-varying spatial distribution of each process. Each runoff rate is calculated from a simple empirical relationship between the amount of water stored in some portion of the soil and a parameter that must be estimated or measured. Thus, subsurface stormflow is taken to be an exponential function of the volume of soil-moisture storage. Each parameter may be obtained from field experiments in which small plots are irrigated with a spray nozzle and measurements are made of soil-moisture changes and the volume, rate, and velocity of runoff. Runoff is routed to the channel using a velocity that varies with local gradient and soil type (and therefore probably vegetation cover) but not with flow depth. It is then routed down the channel network under the assumption that flow velocity varies with discharge at a station on the channel but not along the channel. The authors were able to make satisfactory predictions of storm hydrographs from an 8-km^2 drainage basin using parameters based only on simple field measurements.

Although it contains several important compromises with reality, and several parameters without precisely definable physical meaning, the model has certain advantages over the current generation of statistical-empirical models. It describes explicitly, though approximately, several processes that have been recognized in the field, and it predicts their time-varying spatial distribution. Yet it retains enough simplicity to be usable on a routine basis for predicting hydrographs, zones of a basin that are subject to erosion or the removal of pollutants, and the wetness of soils between storms. It is not yet clear whether the model provides a description of overland flow that is adequate for a deterministic model of erosion and transport of solids, although an empirical, statistical approach to these matters would probably still be necessary. Such work is in the early stages of development and with refinement of its physical basis will probably be the basis of most routine predictions of runoff in areas to which the variable-source concept applies.

To improve physical insights and computational efficiency in variable-source hydrology, there is a need for carefully conceived field experiments coupled with modeling efforts. As in other areas of hydrology, it is an unfortunate fact that field researchers and modeling specialists traditionally work separately. Large quantities of data might be collected, but some measurement crucial to understanding the physics or to making an acceptable mathematical approximation to the physics might be overlooked. Mathematical modelers then make assumptions and approximations to the disappointment and scorn of some field investigators. This state of affairs can be improved through cooperative investigations of processes aimed at understanding the physics of variable-source runoff and at making field results generalizable and transposable to other sites. Government agencies and the major funding bodies should encourage this cooperation in preference to solitary efforts in field measurement or computation. For example, there is much to be learned about how to compute efficiently the spatial patterns of soil drainage in a catchment between rainstorms, the effect of large openings ("macropores") in the soil, the effects of soil anisotropy and variability, and the hydraulics of shallow runoff through dense ground vegetation.

Snowmelt Runoff

Major advances in our understanding of water transmission through snowpacks are mainly the result of the work of Colbeck, who in a series of papers summarized in his 1977 and 1978 review articles has outlined theories of snow metamorphism ("ripening") and its effect on the hydraulic properties of snow, of vertical unsaturated percolation in homogeneous and heterogeneous snowpacks, of unsaturated-saturated flow in packs containing discontinuous ice lenses, and of lateral saturated flow within the base of the snowpack. Wankiewicz (1978a) published experimental work on the hydraulic properties of snow and has reviewed (1978b) the important concepts dealing with snowmelt runoff models. The models treat snow as a porous medium with hydraulic properties (which have been measured in the laboratory and in the field) similar to those of a coarse sand. Thus, modeling is based on Darcy's law and the continuity equation, and solutions of the resulting equations are obtained for the simplest cases by a simplified kinematic approximation (Colbeck, 1971, 1974). Irregular flow-region geometry and other complications, such as introducing more precise conductivity-moisture relationships, require numerical solutions.

The Colbeck models of vertical percolation and lateral overland flow are simplifications of the previously discussed physically based models of infiltration and subsurface flow through a porous medium. The hydraulic properties of snow and the geometry of the flow region (particularly the small thickness of the basal saturated layer compared with the thickness of the pack) allow such simplifications. Dunne *et al.* (1976) demonstrated that these models could be used easily to predict hillslope runoff hydrographs from homogeneous ripe snow if hourly melt rates were known. However, the complexity of the problem escalates rapidly if the snowpack contains ice lenses or even snow layers with differing hydraulic properties as most snowpacks do. The situation becomes similar to that discussed earlier for subsurface flow; the data requirements concerning the thickness, permeability, and continuity of ice lenses are beyond present routine monitoring capabilities, and there is little prospect of incorporating the modern physical concepts into the current generation of lumped models used to forecast snowmelt floods. Other major uncertainties about the accumulation, distribution, and depth of the snowpack; the spatial variation of melting; and topographic effects on runoff also preclude routine application of physically sound snowmelt runoff models. Routine prediction of the lag and attenuation of the diurnal wave of meltwater released at the surface of the pack will continue for some time to be based on quasi-physical, empirical relationships that are based on an analogy between storage and transmission of the meltwater wave through the pack and the storage and transmission of a flood wave through a lake or a stream channel. The latter methods require at least a few historical records, but they work reasonably well for most purposes.

The contributions of Colbeck and others have, however, cast light on several features of the timing of snowmelt runoff that were puzzling heretofore, and they promise to stimulate major refinements in field observations of runoff and of statistical tools for its prediction at the basin scale. They also allow computations of extreme conditions that lie beyond the range of the historical record and of runoff during times when the snowpack characteristics are changing rapidly. Thus, relatively few applications of the detailed physical models based on limited measurements of snowpack characteristics or even plausible assumptions can extend the usefulness and reliability of the empirical techniques currently in use. The applicability of physically based models of snowmelt runoff to hydrological forecasting was reviewed by Colbeck (1977).

The scientific study of soil-snow interactions is not well developed. Current research is focused on the coupling of equations describing heat and water movement at the interface (Guymon, 1978). In some regions, a requirement of snowmelt runoff models is the ability to predict whether the topsoil is frozen (Cary *et al.*, 1978) and the hydraulic properties of the frozen soil. These issues are strongly affected not only by weather and snow depth but also by vegetation cover (Post and Dreibelbis, 1942; Trimble *et al.*, 1958); however, little or no effort has been expended on modeling these effects or on predicting them by other means. This gap in our present knowledge is only one aspect of the general lack of concern for runoff processes in models designed to predict streamflow from snowmelt (see the papers on this topic in the symposium proceedings edited by Colbeck and Ray, 1978).

SUMMARY AND PROSPECT

Most hydrologic predictions are still based on the empirical correlation approach in which measured variables such as flood peaks, mean annual floods, low-flow statistics, or characteristics of unit hydrographs are correlated with various geologic, geomorphic, and climatic variables, and the resulting empirical equation is used to estimate future hydrologic events or statistics in the measured basin or in ungauged watersheds. The approach is adequate for prediction of streamflow at a basin outlet during moderate-sized events and is still the preferred tool of many hydrologists, as indicated by the papers presented at a recent symposium of the International Association of Scientific Hydrology (1975). Limitations of the statistical predictions include the difficulty of estimating extreme events through extrapolation from short records and the fact that methods that treat the drainage basin as a lumped system fail to describe important characteristics of runoff processes that vary spatially. These spatial characteristics are required for rational prediction of other processes, such as the entrainment and transport of pollutants and impeded soil drainage, that may render fields impassable to farm equipment or reduce the suitability of a site for liquid-waste disposal.

Within the last two decades, field experiments have enhanced understanding of the physics and spatially varying characteristics of runoff processes. The work has stimulated physically based mathematical modeling of the spatially distributed processes. These efforts have in turn provoked better field research by providing unified interpretations of data from disparate environments and by suggesting critical measurements that need to be made. This synergism bodes well for scientific hydrology and particularly for field research that seems likely to be conducted selectively in conjunction with physically based mathematical models of each process in order to verify that the model is an adequate description of the process and that the model incorporates correct values of important parameters. Closer cooperation between fieldwork and modeling efforts should encourage formulation of better field experiments and discourage much of the aimless collection of data that sometimes passes for field hydrology, especially at the watershed scale. A vigorous effort is necessary to improve the quality of hydrologic field studies along the lines suggested above. Publication trends in major hydrologic journals suggest that the most vigorous and sophisticated current developments in hydrology are due to the efforts of researchers concerned with physically based mathematical models. However, this expanding frontier will be hollow unless it is matched by equally sophisticated field experiments to discover unexpected hydrologic phenomena, to develop new concepts about familiar processes, and to guide the development of mathematical models based on sound physical insights into field conditions.

The most sophisticated physically based mathematical models of runoff processes are currently competitive with and more accurate than statistical models when the parameters of the former are derived carefully from field measurements. In the simplest cases, such as Horton overland flow from nearly straight, ungullied hillsides, approximate mathematical solutions of the flow equations can be used with great value for routine prediction. Even for more complicated overland-flow situations, computer programs involving numerical solutions of the flow equations are usable for predicting discharge but not yet for predicting the other characteristics of runoff or its ability to transport soluble and solid materials.

Physically based models involving subsurface flow (including infiltration) are much more difficult to apply routinely than are computations of Horton overland flow. They tax computing capacity because of the need to accommodate two- and three-dimensional layered flow regions and to obtain and store information on the complex hydraulic properties of unsaturated porous media. The data requirements of these models are also extremely expensive and time-consuming to satisfy, and it is often not easy to check all aspects of the predictions. They are of value first because in conjunction with field experiments they contribute to the understanding of fundamental hydrologic processes. They can also provide information about certain economic or social concerns that are so important that the resources are made available to overcome the difficulties listed above. Thus, one could envisage questions about the consequences of an accidental spillage of hazardous liquid waste being answered with the aid of an infiltration-runoff model or with a model of unsaturated-saturated subsurface flow; either model would have to be coupled with a transport model to compute the dispersion and chemical reactions of the pollutant. Calculations of maximum probable snowmelt floods in some basins may warrant the use of a sophisticated numerical model of snowpack metamorphism, water percolation, and runoff. Or a two- or three-dimensional computer model of subsurface flow may be of use in analyzing the stability of a large landslide that threatens a dam or other structure. Another potential use of the sophisticated models might be the calculation of nomographs based on relatively few runs of the model. The resulting nomographs could be used for routine predictions by interpolation (graphically or through the use of multivariate statistical methods).

In the longer term it is likely that the usefulness of the sophisticated physically based models will be extended by improvements in computer technology, in methods of collecting the necessary input data, and in simplifications of theory. However, in the foreseeable future most routine predictions of stream discharge, soil moisture, and other hydrologic variables of importance to land and water management will be made partly on the basis of empirical statistical methods and partly by means of approximate physically based models that account for the spatially distributed nature of runoff processes and their controls. Such approximate physical models are now being vigorously developed. It is hoped that they will incorporate parameters that can be obtained by direct field measurement or by estimation on the basis of field conditions rather than by fitting of predictions to measured results. The latter approach is acceptable when observations confirm that the model represents field conditions well. But in the hands of some users the fitting approach has proven simply to be a retrogression to empirical methods of the past and could be improved on through careful field observations.

ACKNOWLEDGMENTS

An earlier draft of this manuscript was improved considerably by the reviews of R. A. Freeze, R. E. Smith, and D. A. Woolhiser.

REFERENCES

Anderson, M. G., and T. P. Burt (1977). Automatic monitoring of soil moisture conditions in a hillslope spur and hollow, *J. Hydrol. 33*, 27-36.

Anderson, M. G., and T. P. Burt (1978). Toward more detailed field monitoring of variable source areas, *Water Resour. Res. 14*, 1123-1131.

Anderson, M. G., and P. E. Kneale (1980). Topography and hillslope soil-water relationships in a catchment of low relief, *J. Hydrol. 47*, 115-128.

Beven, K. (1975). Hillslope hydrographs by the finite element method, *Earth Surface Processes 2*, 13-28.

Beven, K. J., and M. J. Kirkby (1979). A physically-based variable-contributing-area model of basin hydrology, *Hydrol. Sci. Bull. 24*, 27-53.

Betson, R. P. (1964). What is watershed runoff? *J. Geophys. Res. 69*, 1541-1551.

Bonell, M., and D. A. Gilmour (1978). The development of overland flow in a tropical rainforest catchment, *J. Hydrol. 39*, 365-382.

Brooks, R. H., and A. T. Corey (1966). Properties of porous media affecting fluid flow, *Am. Soc. Civil Eng. J. Irrig. Drainage Div. 92, ID2*, 61-88.

Bouma, J., L. W. Dekker, and J. H. M. Wösten (1978). A case study on infiltration into dry clay soil. II, Physical measurements, *Geoderma 20*, 41-51.

Cary, J. W., G. S. Campbell, and R. I. Papendick (1978). Is the soil frozen or not? An algorithm using weather reports, in *Modeling of Snow Cover Runoff*, S. C. Colbeck and M. Ray, eds., U.S. Army, CRREL, pp. 319-327.

Colbeck, S. C. (1971). One-dimensional theory of water flow through snow, *U.S. Army, CRREL Res. Rep. 296*, 16 pp.

Colbeck, S. C. (1974). Water flow through snow overlying an impermeable boundary, *Water Resour. Res. 10*, 119-123.

Colbeck, S. C. (1977). Short-term forecasting of water runoff from snow and ice, *J. Glaciol. 19*, 571-588.

Colbeck, S. C. (1978). The physical aspects of water flow through soil, *Adv. in Hydrosci. 11*, 165-206.

Colbeck, S. C., and R. M. Ray, eds. (1978). *Modeling of Snow Cover Runoff*, U.S. Army Cold Regions Research and Engineering Laboratory, Hanover, N.H., 432 pp.

Davidson, J. M., L. R. Stone, D. R. Nielsen, and M. E. Larue (1969). Field measurement and the use of soil-water properties, *Water Resour. Res. 5*, 1312-1321.

Dunne, T. (1970). Runoff production in a humid area, *U.S. Dept. of Agric. ARS-41-160*, 108 pp.

Dunne, T. (1978). Field studies of hillslope flow processes, in *Hillslope Hydrology*, M. J. Kirkby, ed., Wiley-Interscience, New York, 227-293.

Dunne, T., and R. D. Black (1971). Runoff processes during snowmelt, *Water Resour. Res. 7*, 1160-1172.

Dunne, T., and W. E. Dietrich (1980). Experimental investigation of Horton overland flow on tropical hillsides, *Z. Geomorphol. Suppl. 35*.

Dunne, T., T. R. Moore, and C. H. Taylor (1975). Recognition and prediction of runoff-producing zones in humid regions, *Hydrol. Sci. Bull. 20*, 305-327.

Dunne, T., A. G. Price, and S. C. Colbeck (1976). The generation of runoff from subarctic snowpacks, *Water Resour. Res. 12*, 677-685.

Eagleson, P. S. (1972). Dynamics of flood frequency, *Water Resour. Res. 8*, 878-898.

Emmett, W. W. (1970). The hydraulics of overland flow on hillslopes, *U.S. Geol. Surv. Prof. Pap. 662-A*.

Engman, E. T., and A. S. Rogowski (1974). A partial area model for storm flow synthesis, *Water Resour. Res. 10*, 464-472.

Foster, G. R., and L. D. Meyer (1975). Mathematical simulation of upland erosion by fundamental erosion mechanics, *U.S. Dept. of Agric. ARS-5-50*, 190-207.

Freeze, R. A. (1969). The mechanism of natural groundwater recharge and discharge, 1, One-dimensional, vertical, unsteady, unsaturated flow above a recharging or discharging groundwater flow system, *Water Resour. Res. 5*, 153-171.

Freeze, R. A. (1971). Three-dimensional, transient, saturated-unsaturated flow in a groundwater basin, *Water Resour. Res. 7*, 347-366.

Freeze, R. A. (1972a). Role of subsurface flow in generating surface runoff, 1, Baseflow contributions to channel flow, *Water Resour. Res. 8*, 609-623.

Freeze, R. A. (1972b). Role of subsurface flow in generating surface runoff, 2, Upstream source areas, *Water Resour. Res. 8*, 1272-1283.

Freeze, R. A. (1974). Streamflow generation, *Rev. Geophys. Space Phys. 12*, 627-647.

Freeze, R. A. (1978). Mathematical models of hillslope hydrology, in *Hillslope Hydrology*, M. J. Kirkby, ed., Wiley-Interscience, New York, pp. 177-226.

Freeze, R. A. (1980). A stochastic-conceptual analysis of rainfall-runoff processes on a hillslope, *Water Resour. Res. 16*, 391-408.

Green, W. H., and G. A. Ampt (1911). Studies on soil physics. I: The flow of water and air through soils, *J. Agric. Sci. 4*, 1-24.

Guymon, G. L. (1978). A review of soil-snow interactions, in *Modeling of Snow Cover Runoff*, S. C. Colbeck and M. Ray, eds., U.S. Army CRREL, pp. 297-303.

Harr, R. D. (1977). Water flux in soil and subsoil on a steep forested slope, *J. Hydrol. 33*, 37-58.

Hewlett, J. D. (1961a). Soil moisture as a source of baseflow from steep mountain watersheds, *U.S. For. Serv. Southeast. For. Exp. Stn. Res. Pap. 132*, 11 pp.

Hewlett, J. D. (1961b). Some ideas about storm runoff and baseflow, *U.S. For. Serv. Southeast. For. Exp. Stn. Ann. Rep.*, pp. 62-66.

Hewlett, J. D., and A. R. Hibbert (1967). Factors affecting the response of small watershed to precipitation in humid areas, in *Forest Hydrology*, W. E. Sopper and H. W. Lull, eds., pp. 275-290.

Hewlett, J. D., and W. L. Nutter (1970). The varying source area of streamflow from upland basins, in *Symposium on Interdisciplinary Aspects of Watershed Management*, Am. Soc. Civil Eng., pp. 65-83.

Horton, R. E. (1933). The role of infiltration in the hydrologic cycle, *EOS: Trans. Am. Geophys. Union 14*, 446-460.

Horton, R. E. (1938). Phenomena of the contact zone between the ground surface and a layer of melting snow, *Internatl. Assoc. Sci. Hydrol. Publ. 23*, pp. 545-561.

Horton, R. E. (1940). An approach towards a physical interpretation of infiltration capacity, *Soil Sci. Soc. Am. Proc. 5*, 399-417.

International Association of Hydrological Sciences (1975). *The Hydrological Characteristics of River Basins*, IAHS Publ. 117.

Kirkby, M. J., ed. (1978). *Hillslope Hydrology*, Wiley-Interscience, New York, 389 pp.

Leigh, C. H. (1978). Slope hydrology and denudation in the Pasoh Forest Reserve. I, Surface wash; Experimental techniques and some preliminary results, *Malay. Nat. J. 30*, 179-197.

McIntyre, D. S. (1958a). Permeability measurements on soil crusts formed by raindrop impact, *Soil Sci. 85*, 185-189.

McIntyre, D. S. (1958b). Soil splash and the formation of surface crusts by raindrop impacts, *Soil Sci. 85*, 261-266.

Mein, R. G., and D. A. Farrell (1974). Determination of wetting front suction in the Green-Ampt equation, *Soil Sci. Soc. Am. Proc. 38*, 872-876.

Mein, R. G., and C. L. Larson (1973). Modeling infiltration during a steady rain, *Water Resour. Res. 9*, 384-394.

Morel-Seytoux, H. J. (1976). Derivation of equations for rainfall infiltration, *J. Hydrol. 31*, 203-219.

Morel-Seytoux, H. J. (1978). Derivation of equations for variable rainfall infiltration, *Water Resour. Res. 14*, 561-568.

Morel-Seytoux, H. J., and J. Kanji (1974). Derivation of an equation of infiltration, *Water Resour. Res. 10*, 795-800.

Musgrave, G. W., and H. N. Holtan (1964). Infiltration, in *Handbook of Applied Hydrology*, V. T. Chow, ed., McGraw-Hill, New York.

Northcliff, S., J. B. Thornes, and M. J. Waylen (1979). Tropical forest systems: A hydrological approach, *Amazoniana 6*, 557-568.

Philip, J. R. (1957-1958). The theory of infiltration, *Soil Sci. 83*, 345-357; and *84*, 163-177, 257-264, 329-339.

Philip, J. R. (1969). Theory of infiltration, in *Advances in Hydroscience*, Academic Press, New York, pp. 215-296.

Philip, J. R. (1980). Field heterogeneity: Some basic issues, *Water Resour. Res. 16*, 443-448.

Post, F. A., and F. R. Dreibelbis (1942). Some influences of frost penetration and micro-climate on the water relationships of woodland, pasture, and cultivated soils, *Soil Sci. Soc. Am. Proc. 7*, 95-104.

Price, A. G., L. K. Hendrie, and T. Dunne (1978). Controls on the production of snowmelt runoff, in *Modeling of Snow Cover Runoff*, S. C. Colbeck and M. Ray, eds., U.S. Army, CRREL, pp. 257-268.

Rubin, J. (1966). Theory of rainfall uptake by soils initially drier than their field capacity and its applications, *Water Resour. Res. 2*, 739-749.

Singh, V. P. (1976). A distributed converging overland flow model, 3, Application to natural watershed, *Water Resour. Res. 12*, 902-908.

Smith, R. E. (1972). The infiltration envelope: Results from a theoretical infiltrometer, *J. Hydrol. 17*, 1-21.

Smith, R. E., and J. Y. Parlange (1978). A parameter-efficient hydrologic infiltration model, *Water Resour. Res. 14*, 533-538.

Smith, R. E., and D. A. Woolhiser (1971). Overland flow on an infiltrating surface, *Water Resour. Res. 7*, 899-913.

Stephenson, G. R., and R. A. Freeze (1974). Mathematical simulation of subsurface flow contributions to snowmelt runoff, Reynolds Creek, Idaho, *Water Resour. Res. 10*, 284-294.

Tennessee Valley Authority (1964). Bradshaw Creek-Elk River: A pilot study in area-stream factor correlation, *Res. Pap. No. 4, Office of Tributary Area Development*, Tennessee Valley Authority, 64 pp.

Trimble, G. R., R. S. Sartz, and R. S. Pierce (1958). How type of frost affects infiltration, *J. Soil Water Conserv. 13*, 81-82.

Wankiewicz, A. (1978a). A review of water movement in snow, in *Modeling of Snow Cover Runoff*, S. C. Colbeck and M. Ray, eds., U.S. Army, CRREL, pp. 222-252.

Wankiewicz, A. (1978b). Water pressure in ripe snowpacks, *Water Resour. Res. 14*, 593-599.

Whipkey, R. Z. (1965). Subsurface stormflow from forested slopes, *Intl. Assoc. Sci. Hydrol. Bull. 10*, 74-85.

Woolhiser, D. A. (1975). Simulation of unsteady overland flow, in *Unsteady Flow in Open Channels*, K. Mahmood and V. Yevjevich, eds., Water Res. Pubs., Fort Collins, Colo., pp. 485-508.

Woolhiser, D. A., C. L. Hanson, and A. R. Kuhlman (1970). Overland flow on rangeland watersheds, *N. Z. J. Hydrol. 9*, 336-356.

Woolhiser, D. A., and J. A. Liggett (1967). Unsteady, one-dimensional flow over a plane: The rising hydrograph, *Water Resour. Res. 3*, 753-771.

Hydrology and Climate

2

PETER S. EAGLESON
Massachusetts Institute of Technology

INTRODUCTION

The global moisture cycle has branches in both the earth and the atmosphere. These branches are dynamically coupled across the earth's surface (at widely different time and space scales), so that a change in one branch produces a change in the other.

Hydrologists and atmospheric scientists have achieved a first-order understanding of this complex system through independent study of their respective, decoupled land-surface and atmospheric components. Emerging problems associated with anthropogenic alteration of both components demand a higher order of understanding.

Anthropogenic alterations in the land surface disrupt the natural balance of the moisture cycle by causing changes in the transfer of heat and moisture between land and air. A new balance must ultimately be reached, but how and where will it differ from the old? A reduction in local evaporation such as that resulting from deforestation or swamp drainage or an increase in local evaporation such as that occurring with extensive irrigation must produce compensating changes in precipitation or evaporation elsewhere. The question is where? What other changes will result from the accompanying changes in land-surface temperature? Gradually increasing atmospheric concentrations of carbon dioxide and particulates are certain to bring future modification of the hydrologic cycle. What will be the nature, magnitude, location, and time scale of these changes?

Such questions are of increasing importance to food production, water supply, energy demand, and public health. Their answers require the development and verification of complex mathematical models of the global atmosphere-land surface-ocean system.

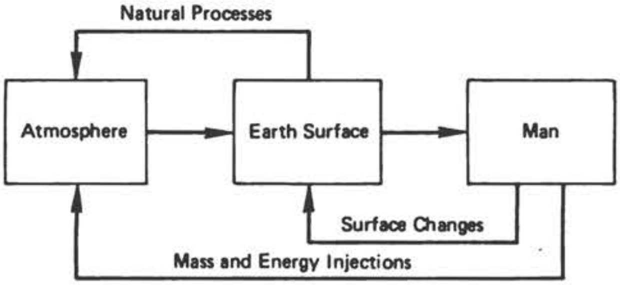
BACKGROUND

Since the birth of modern hydrology a century ago, water scientists and engineers have focussed their attention on the fundamental hydrologic unit, the catchment. Until relatively recently, perhaps the last two decades, interest has been on small-scale problems such as providing for adequate water supply and protecting man against the extremes of streamflow. This limited viewpoint, while stressing the sensitivity of catchment output to land-surface state, has encouraged the convenient assumption that there is no feedback to the climate from either the state or the hydrologic output of the catchment. This simple conception is illustrated schematically in Figure 2.1a. This has proved adequate for the purposes stated above primarily because of incommensurate atmospheric and land-surface scales. In the middle latitudes of the developed world, atmospheric processes are dominated by cyclonic motions having horizontal scales an order of magnitude larger than those of the individual catchments being studied. *Given the local atmospheric inputs* and the configuration and state of the local land surface, hydrologists have developed complicated causal models of such systems (Crawford and Linsley, 1966), which, with adequate local calibration, do an acceptable job of generating, for example, the desired streamflows and soil moistures.

Nevertheless, the current literature presents considerable evidence implicating man in hydrologic anomalies through his changes to the land surface. As stated by the World Meteorological Organization (WMO, 1979), these changes are of four basic types:

1. Changes of surface roughness, which governs the exchange of momentum and energy of air motion between the atmosphere and ground;
2. Changes of surface albedo, which governs the fraction of solar radiation converted to surface heating;
3. Changes in thermal behavior of the ground and in heat transfer to the atmosphere; and
4. Changes of moisture-holding capacity of the surface (and of vegetation), which causes the flow of water substance between the surface and the atmosphere to be redistributed in space and time.

a. Former Viewpoint



b. Current Viewpoint

FIGURE 2.1 Evolution of understanding of man's role in atmosphere-earth surface coupling.

The most easily recognizable of these anomalies are small scale where the hydrologic effect occurs in the vicinity of the land-surface change, the most familiar of which are urban related (Landsberg, 1956, 1974). Table 2.1 summarizes the known urban-rural climatic differences as averaged for a number of cities. The

TABLE 2.1 Climate Changes Resulting from Urbanization (from Chagnon, 1977)

	Average Changes, as a Percentage of Rural Conditions		
Variable	Annual	Cold Season	Warm Season
Solar radiation	-22	-34	-20
Temperature	+2	+4	+1
Humidity (relative)	-6	-2	-8
Visibility (frequence)	-26	-34	-17
Fog (frequency)	+60	+100	+30
Wind speed	-25	-20	-30
Cloudiness (frequency)	+8	+5	+10
Rainfall (amount)	+14	+13	+15
Snowfall (amount)	±10	±10	--
Thunderstorms (frequency)	+16	+5	+29

observed increase of urban-induced precipitation results not from increased local evaporation, for this is normally *decreased* by urbanization, but rather from the thermodynamic changes and from the presence of additional condensation nuclei. The urban-produced anomalies are not confined to the urban areas themselves but have been felt in rural areas as far as 100 km downwind (Changnon, 1977), where their economic impact on agriculture can be important.

Observations also suggest that the well-known direct amplification of flood flows by the surface changes accompanying urbanization (Terstriep *et al.*, 1976) may be augmented *indirectly* by an urban-induced increase in precipitation intensities.

Extensive irrigation in the Great Plains (Schickedanz, 1976) and in the Columbia Basin (Stidd, 1967, 1975) has been shown to increase the rainfall in the vicinity of the irrigated areas during the irrigation months. It is thought that the increase results not only from the changed (here increased) atmospheric moisture but also from thermodynamic side effects produced by a cool, moist "dome" over the irrigated area. Optimum water use suggests the need for understanding this feedback process and anticipating it during project design.

As we begin to consider problems of regional scale, and as development increases in the low latitudes (where thermal convection may dominate the moisture cycle), the atmospheric and surface scales become commensurate and the feedback effects of land-surface state on local atmospheric processes become important.

Increasing atmospheric pollution (i.e., carbon dioxide) as well as serious public consideration of such macroengineering projects as drainage of the White Nile swamps, flooding of the Qattara depression, and deforestation of the Amazon Basin have prompted growing scientific interest in the climatic as well as the hydrologic consequences of large-scale anthropogenic influences (Sagan *et al.*, 1979). There is even some concern (Lorenz, 1976) that the coupled earth-atmosphere system may have one or more alternate equilibrium states to which it might move under the proper natural or anthropogenic perturbation.

Theoretical analyses indicate that the removal of vegetation may cause a change in the precipitation regime. Charney (1975) showed this for the arid Sahel, where albedo increase due to overgrazing can apparently cause sinking atmospheric motion that amplifies the dryness. Lettau *et al.* (1979) demonstrated that radical deforestation of the Amazon Basin should decrease precipitation in the western, downwind portion of the basin.

Morton (1968) and others demonstrated the direct regulation that regional evapotranspiration exerts on regional humidity.

To study such problems it is necessary to consider the coupled behavior of the atmosphere and the earth's surface with respect to the storage and flux of both heat and water vapor. This feedback system is illustrated schematically in Figure 2.1b. Current lack of understanding of the nature, strength, sensitivity, and scale of the many feedback loops of this system (Kellog and Schneider, 1974) inhibits our dealing with these questions. Most specific and many general hydroclimatological questions may never be answered without global analytical models that include the oceans as well as the atmosphere and the land surface. Such analytical models are one of the goals of the WMO-ICSU Global Atmospheric Research Program (GARP, 1975). Other more limited questions may be answered at least qualitatively through the use of judicious idealizations (Lettau, 1969; Sasamori, 1970; Deardorff, 1978; Eagleson, 1978) although the counterintuitive nature of the complex physical interactions makes such system simplifications a dangerous expedient.

IMPORTANCE OF EARTH-ATMOSPHERE COUPLING

Philip (1977) characterized the global hydrologic cycle as "a gigantic distillation scheme" that forms a critical part of the earth's heat engine. He reminds us that the intensity of this cycle, and hence life as we know it, is a result of a particularly favorable juxtaposition of earthly temperatures and water properties. In particular, the heat capacity and latent heat of vaporization of water are high (in fact, greater than for most known substances), and earth's temperatures are moderate enough to ensure the presence of either ice or liquid water in addition to water vapor, yet at these temperatures the saturated density of water vapor is great enough to produce significant rates of evaporation and condensation. The combination of these factors produces a close coupling of our planet's energy and water cycles, leading to latent-heat fluxes that are the major component of earthly energy redistribution.

The average annual heat and water balances are given in Tables 2.2 and 2.3, respectively, for the planet as a whole and for its separate ocean and land-surface portions. We see from these that globally 83 percent of the net radiational energy is involved in evaporation. For ocean surfaces where the water supply is unrestricted, this process uses 90 percent of the locally available energy, while for land surfaces, where surface moisture supply is limited either by precipitation or by soil properties, slightly more than half (51 percent) of the available energy is used in evaporative processes.

The oceans of the earth are thus certainly the key element in this heat engine, being a net source of atmospheric moisture and hence of energy at least on an annual basis. We must anticipate global hydrologic consequences from any significant change in their surface temperatures.

The land surfaces, however, must not be dismissed as a negligible and passive climatic component. Fully 10 percent of all solar energy absorbed *globally* goes into evaporative processes on the *land* surfaces, and there is evidence (Benton and Estoque, 1954) that in the

TABLE 2.2 Annual Heat Balance in kcal cm^{-2} (from Budyko, 1974)

Region	Net Radiation	Latent Heat	Sensible Heat
Ocean (71%)	82 (100%)	74 (90%)	8 (10%)
Land (29%)	49 (100%)	25 (51%)	24 (49%)
Earth (100%)	72 (100%)	60 (83%)	12 (17%)

summer half-year the hydrologic cycle is reversed, with the land surfaces becoming a net *source* of atmospheric moisture. As our natural land-surface systems become more highly stressed by man, the risk of catastrophe increases and a higher order of physical understanding becomes essential.

TABLE 2.3 Annual Water Balance in Millimeters (from Philip, 1977)

Region	Precipitation	Evaporation	Runoff
Ocean	1270	1400	-130
Land	800	485	315
Earth	1133	1133	0

PHYSICAL DESCRIPTION OF THE VERTICAL HYDROLOGIC COUPLING

The atmosphere is hydrologically coupled to the earth through the exchange of thermal energy and water mass across the ocean and land surfaces. The *primary one-dimensional* elements of this coupling are indicated schematically in Figure 2.2 for ocean surfaces, in Figure 2.3 for snow and ice surfaces, and in Figure 2.4 for natural land surfaces. In these illustrations the principal-state variables are indicated by the labeled rectangles. The primary energy and moisture fluxes that either lead to or are produced by these states are indicated by the labeled arrows. The unlabeled arrows represent the effect of one state variable on another. Plus and minus signs at the arrowheads indicate the sense of the change in the "directee" variable produced by an increase in the "director" variable.

Atmospheric dynamics and energetics produce concentrations of atmospheric water vapor in the form of clouds, the density of which serves to regulate the local inputs to the surface of both precipitation and net solar radiation. The other primary-atmospheric state variables are temperature and humidity (or vapor pressure). These last two have their counterparts on the earth side of the interface, only here the state of the water is represented by a different measure, depending on the medium. The local states of the systems on each side of the interface, as determined by their respective temperatures and moisture concentrations, control the amount of water (and hence the amount of latent heat) that is returned locally to the atmosphere through the surface cooling (or self-limiting), as indicated in Figure 2.2 by the negative sign on the feedback loop of the evaporation process.

Ocean Surfaces

At oceanic surfaces (Figure 2.2), the local moisture concentration is, of course, unity regardless of local precipitation. Local evaporation thus takes place independently of local precipitation and according to the local vapor transport capacity of the atmosphere. This "potential"-evaporation rate is an increasing function of the sea-surface temperature, which in turn depends on the evaporative cooling as well as on the radiant and sensible heat transport in both media. The negative feedback of evaporative cooling illustrates its potential for thermostatic regulation of surface temperature.

Snow and Ice Surfaces

At snow and ice surfaces (Figure 2.3), there also is an unrestricted supply of moisture, although the low-vapor pressures accompanying the frozen surface greatly suppress evaporation (i.e., sublimation). In contrast to liquid water, snow and ice metamorphose with age, and their state is only crudely described in terms of the two variables shown. Most important of the omitted variables is the very high albedo, which reduces to secondary importance the radiational term of the vertical-energy balance.

Not shown in this one-dimensional idealization is the areal extent of the snow and ice surface. For water surfaces, this variable plays an integral role in a lateral positive-feedback process sequentially involving water surface cooling, ice formation, albedo increase, surface super-cooling, air cooling, and spreading further water-surface cooling. The sensitivity of this process, first suggested by Budyko (1962), has aroused concern over the stability of the earth's climate (Schneider and

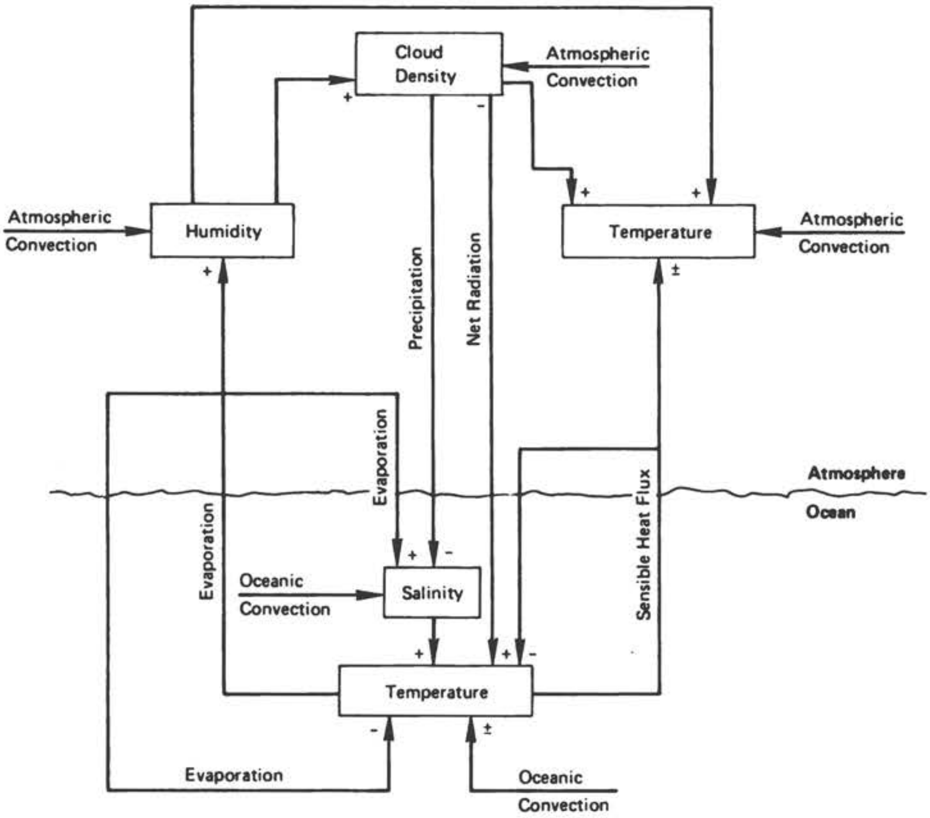


FIGURE 2.2 Primary moisture and energy couplings across the ocean surface.

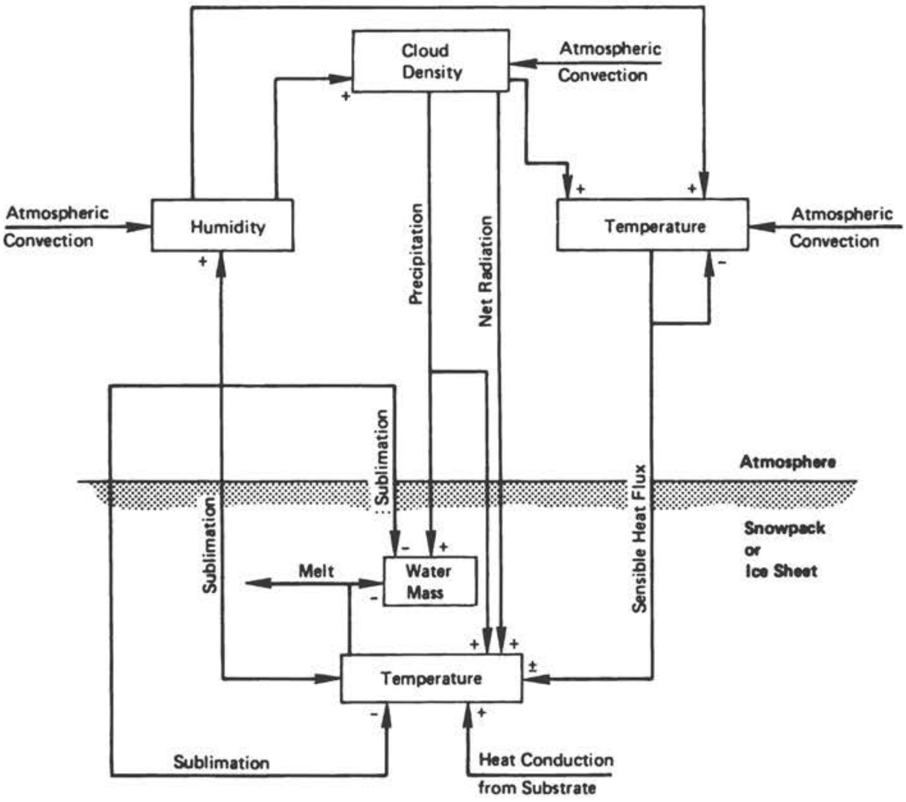


FIGURE 2.3 Primary moisture and energy couplings across a snow or ice surface.

Gal-Chen, 1973) in the face of changing average surface temperatures in the polar regions.

It should be pointed out that when considering formation of ice, the "melt" of Figure 2.3 becomes "accretion" and the sense of its feedback to water mass and the sense of the sensible heat flux will both be reversed.

Land Surfaces

At land surfaces (Figure 2.4) the couplings are much more complex. Here, the water-vapor flux from earth to atmosphere has components in both the bare soil (evaporation) and in the vegetation (transpiration). Their sum, evapotranspiration, depends on and is limited by the atmospheric-vapor-transport capacity, which is strongly influenced by land-surface temperature.

Within certain temperature bounds the vegetation canopy will develop to a natural equilibrium (i.e., "climatic climax") density governed by the available energy, nutrients, and soil moisture. The vegetation canopy plays the central role in feedback loops on both sides of the interface.

Finally, the local land-surface evaporation (in contrast to local oceanic evaporation) is bounded by the local precipitation.

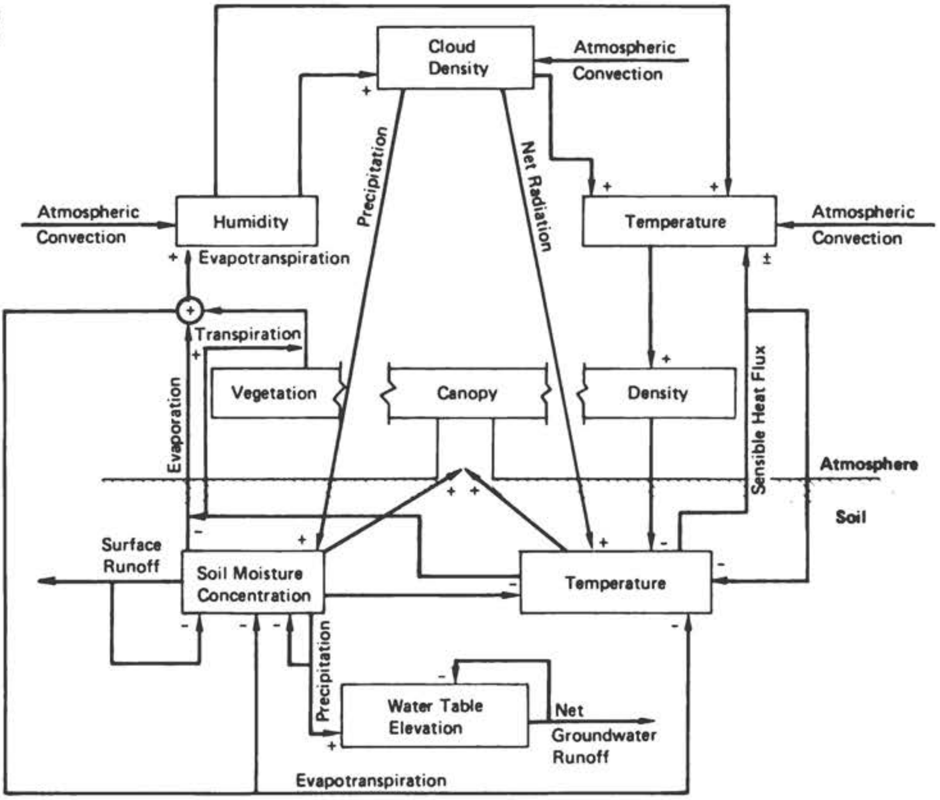
For the land surface it is not only the precipitation climate (i.e., statistics of annual totals) that is of interest but also the intensity, duration, and spacing (both temporal and spatial) of its episodes, because these are critical to the establishment and maintenance of soil moisture levels.

We see from this that evaporation from land surfaces may be controlled by either the climate or the soil. As evaporation falls away from its potential value with increasing climatic aridity, its thermostatic control of land-surface temperature is correspondingly weakened.

Gravity continually pulls soil moisture downward (i.e., "percolation"), where it "puddles" to a water-table elevation that is sufficient under steady-state conditions to force a groundwater runoff equal to the gravitational accretion. Where such lateral runoff is sharply constrained geologically, the water table may rise to the surface, forming a swamp. Under these extreme but not uncommon conditions, the earth-atmosphere coupling reverts essentially to that shown in Figure 2.2 without the subsurface convection but with the addition of vegetation.

Another lateral process not seen in this one-dimensional simplification and that has overriding hydrologic importance in arid climates results from the small spatial scale of the convective-precipitation events characteristic of such areas. In these cases, local precipitation becomes local surface runoff according to the local increase in soil moisture during the storm. This local runoff moves laterally to a dry spot and infiltrates, to be largely evaporated later.

FIGURE 2.4 Primary moisture and energy couplings across a land surface.



Characteristic Times and Horizontal Scales

In thinking about the behavior of this "engine," it helps to have a feeling for the intensities and scales of the major components.

Dividing values of volume and turnover rate, Philip (1977) estimated the global average residence times for the various components of the earth's waters. These are reproduced here as Table 2.4, which clearly shows the incommensurate time scales of the separate water reservoirs of the atmosphere, hydrosphere, and lithosphere. Proportional differences in spatial scales also exist, which presents difficulty when modeling the coupled system to answer questions of large scale. To represent all processes in their dynamic majesty would require models having time resolutions of minutes and spatial grid spacings of a few kilometers. At present this is clearly unfeasible, both computationally and economically.

One solution to this problem has been to model only the atmosphere dynamically and to represent the earth-surface boundary by simple, uncoupled heat and moisture conservation relationships. Some inferences from such models are summarized in the following section.

Another resolution of this problem may be found through the development of statistical-dynamic models (Leith, 1975). In such models the system is divided into two parts, one having long-period or slow fluctuations and the other containing the rapidly varying components. The scales of the latter system are left unresolved, and this component is represented statistically. These statistics appear in a relatively large-scale dynamical representation of the former system. Eagleson (1978) took a first step in this direction hydrologically by representing the annual water balance in terms of the soil moisture. The soil moisture responds dynamically to atmospheric excitation, which is statistically expressed.

When assessing man's impact on climate and hydrology we need to know the lateral scale of the feedback loops shown in Figures 2.2, 2.3, and 2.4. This knowledge will help to define the geographical influence function (i.e., the environmental impact radius) of a proposed land-surface change and will lead to an estimate of the percentage of local precipitation derived from local evapotranspiration. Very little is known about these scales.

TABLE 2.4 Global Average Residence Times for the Waters of the Earth (Philip, 1977)

Component	Global Average Residence Time
Ocean	2600 yr
Ice caps and glaciers	1100 yr
Groundwater	700 yr
Lakes	13 yr
Soil water	0.5 yr
Rivers	13 days
Atmosphere	8.2 days
Biological water	3.4 days

We may approximate the horizontal scale of the evaporation precipitation process by considering the longitudinal travel of an atmospheric parcel in one residence time. This will vary from nearly zero in radiative-thermal, convective-type (i.e., vertical process) atmospheres to continental scale, where horizontal atmospheric motions predominate. The scale of the land-surface change will have to exceed this atmospheric-moisture replacement distance before the feedback loop can begin to create a downwind amplification of the original disturbance.

The strength of the local coupling between evaporation and precipitation was estimated to be small by Budyko and Drozdov (1953) for the European U.S.S.R., and also by Benton and Estoque (1954) for the Mississippi Valley. However, Lettau *et al.* (1979) found localities in the Amazon Basin where up to 71 percent of the precipitation appears to come from locally evaporated water. As they point out, the apparent underestimation of moisture recycling provided by the Budyko theory results predominately from local showers caused by the diurnal heat cycle and occurring before the locally evaporated moisture gets uniformly mixed throughout the tropospheric column (as Budyko assumed). Evidently, this extremely important environmental question has not yet been resolved, and we look to detailed dynamical models for the answer.

SOME HYDROLOGIC INFERENCES FROM CLIMATE MODELING

In recent years, dynamic general circulation models (GCMs) have been developed and exercised numerically (Mason, 1979). They have proved capable (Manabe and Hollaway, 1975; Gates and Schlesinger, 1977; Stone *et al.*, 1977) of simulating the gross features of the global distribution of various hydrologic quantities, particularly annual precipitation and annual evaporation. Early models employed primitive (i.e., prescribed) hydrologic boundary conditions in the interests of computational economy and in the naive belief that the land surface, at least, was a passive system having relatively constant local thermal and moisture states. Modeled hydrologic cycles using these boundary conditions proved much too intensive and led to numerical experiments in which the effect of the boundary representation was explored. For example:

1. Manabe (1975) demonstrated that prescription of all continental soils to be constantly saturated causes a large increase in precipitation when compared with model runs having no such constraint.
2. Charney *et al.* (1977) found significant sensitivity in simulated atmospheric circulation and precipitation to the particular formulation used for surface evapotranspiration.

3. Rowntree and Bolton (1978) showed the monthly rainfall of central Europe to be extremely sensitive to the soil moisture of that region at the beginning of the prior month.

These and improved models have also been used to explore some of the likely hydrologic consequences of anthropogenic influence. For example:

1. Manabe and Weatherald (1975) showed that doubling the carbon dioxide concentration of the atmosphere significantly increases the intensity of the hydrologic cycle.
2. Charney *et al.* (1977) demonstrated that with or without local evaporation an increase of surface albedo causes a decrease of precipitation.
3. Walker and Rowntree (1977) found for the Sahel region of Africa that increasing or decreasing the soil moisture (while keeping the albedo constant) produced a positive feedback, which, respectively, increased or decreased the precipitation.
4. Washington and Chervin (1979) found that the intensity of thermal energy expected to be injected into the atmosphere by the U.S. East Coast megalopolis in A.D. 2000 may produce large and significant precipitation changes in that region.

SOME CLIMATIC INFERENCES FROM HYDROLOGIC MODELING

At land surfaces the soil column responds dynamically to the climatic sequences of precipitation and evapotranspiration events. It accepts part of the delivered moisture during periods of precipitation (shedding the rest as surface runoff), pumps some of this back to the surface during evaporative periods, and rejects the remainder to the water table more or less continuously. The surface moisture exchange and thus the surface heat exchange depend critically on the physical properties of the soil and vegetation as well as on the weather conditions during the alternate periods of precipitation and evapotranspiration. The quantitive relation among the long-term averages of this partition is called the "water balance."

The single most important measure of mean hydroclimatological behavior is the *evaporation efficiency*, the ratio of the average annual actual evapotranspiration to the average annual potential evapotranspiration of a given surface (Budyko, 1974). This ratio approaches 1 as the critical evapotranspiration parameters (soil moisture, soil properties, time between storms, water-table elevation, vegetation species and density, and potential rate of bare soil evaporation) combine (Eagleson, 1978) to indicate an increasing ability of the climate-soil-vegetation system to deliver moisture to the surface from below. At this extreme the actual evapotranspiration is said to be "demand limited" or "climate controlled," since its value is restricted only by the capacity of the atmosphere to remove moisture from the surface. Such systems are classified as "humid." At the other extreme, as the above parameters combine to indicate a decreasing ability of the system to deliver water to the surface from below, the evaporation efficiency approaches zero. For low efficiencies we say the system is "supply limited" in that the actual evapotranspiration is controlled either by the supply of water to the soil surface from above (i.e., precipitation) or by the soil properties that regulate the supply of water from below.

The next most important hydroclimatological parameter is the *potential humidity*, which is defined as the ratio of the mean annual precipitation to the mean annual potential evapotranspiration. It has been shown (Eagleson, 1978) that when this ratio is less than 1, the mobility of soil moisture controls the evaporation efficiency and that systems in this range are highly sensitive to those land-surface changes that alter the vegetation or the effective soil properties. Such systems may be classified as "arid." Systems having potential humidities greater than 1 are *potentially* humid in that their evaporation efficiency is not limited by atmospheric moisture supply. They have the climatic capability of high-evaporation efficiencies, and this efficiency is constrained to a decreasing degree by soil-moisture immobility. For high potential humidities, the evaporation efficiency approaches 1, giving again the case in which evaporative behavior is limited by the atmospheric-vapor transport capacity. The hydrologic behavior of such humid systems is highly sensitive to land-surface changes that alter the albedo and is insensitive to land-surface changes that alter the vegetation and/or the effective soil properties (unless these also modify the albedo).

Practical interpretations of the above behavior must take account of another feedback loop not yet pictured or discussed--the effect of vegetation on soil properties (Eyre, 1968). The presence of vegetation assists in the processes of mechanical and chemical weathering. More important to this discussion, however, is the role of vegetation in modifying soil structure through the addition of humus. The clay-humus complex facilitates the movement of soil moisture and has the potential for holding plant nutrients in chemical bond. Availability of water and nutrients promotes growth of the vegetation canopy. The canopy protects the soil from the nutrient-leaching action of unimpeded precipitation as well as from the structure-changing drying and chemical reactions induced by direct sunlight. When we consider the effect of deforestation, for example, we must realize not only that the albedo may change (particularly in the arid climate where the soils are dry and light in color), and that the soil moisture will change, but also that the hydraulic properties of the soil may change in such a way as to inhibit soil-moisture movement.

SUMMARY

The atmosphere, oceans, and land surface (including vegetation) are coupled dynamically at widely different time and space scales, forming a complex global system for the transport of heat and water. The behavior of this hydroclimatological system has a large-random component, thus its state variables must be defined probabilistically.

The system responds selectively to change in its parameters, with the marginal (i.e., highly stressed) ecosystems being particularly sensitive. We have seen this recently in the Sahel, but we do not yet understand the sense, strength, and scale of the sensitivities. Parameter change alters the variance (if not the mean) of certain state variables. Anthropogenic change often removes the buffering of natural systems (for example, the removal of temperature-modulating vegetation by urbanization or deforestation), leading to a variance increase. Many of man's life-support systems, particularly agriculture, water supply, and space heating are sensitive to change in the existing hydroclimatological regime, thus it is essential that we develop the capability of anticipating the quality, quantity, and location of the anthropogenic effects.

The hope for such a distributed, quantitative understanding of the interrelationship between climate and hydrology lies with the development of sophisticated models--models with the capability of providing at least the second moments of the state variables. To do this at global scale without the prohibitive expense of repetitive Monte Carlo simulation presents a primary challenge to climatological modelers.

Current physical understanding of the separate system components seems to provide an adequate basis for the necessary modeling effort. The problems lie with the modeling; that is, isolation of the critical physical processes and representation of these at dynamically adequate (yet computationally economical) time and space intervals while bridging the disparity of scale between coupled system elements.

To verify these models, global data sets are needed that specify current land-surface state (i.e., soils, vegetation, land use, water surface, and snow cover) and give the global distribution of the water-balance elements. These data sets should give monthly averages and must include the oceans.

Finally, as Eriksson (1975) pointed out, the interactive role of hydrology in climate will be mastered only with an uncommon degree of cooperative endeavor from a broad range of the earth sciences. Perhaps this will be the most difficult problem of all.

REFERENCES

Benton, G. S., and M. A. Estoque (1954). Water vapor transfer over the North American continent, *J. Meteorol. 11*, 462-477.

Budyko, M. I. (1962). Polar ice and climate, *Izv. Akad. Nauk SSSR Ser. Geogr. Geofiz. No. 6*, 3-10. (Available in English as Res. Mem. RM-5003-PR, 1966, 9-23, Rand Corp., Santa Monica, Calif.).

Budyko, M. I. (1974). *Climate and Life*, (English edition by D. H. Miller) Academic Press, New York, 508 pp.

Budyko, M. I., and O. A. Drozdov (1953). Characteristics of the moisture circulation in the atmosphere, *Izv. Akad. Nauk SSSR Ser. Geogr. Geofiz. No. 4*, 5-14.

Changnon, S. A., Jr. (1977). Impacts of urban-modified precipitation on man's activities, *J. Weather Modification 9*, 8-18.

Charney, J. G. (1975). Dynamics of deserts and drought in the Sahel, *Q. J. R. Meteorol. Soc. 101*, 193-202.

Charney, J. G., W. J. Quirk, S. M. Chow, and J. Kornfield (1977). A comparative study of the effects of albedo change on drought in semi-arid regions, *J. Atmos. Sci. 34*, 1366-1385.

Crawford, N. H., and R. K. Linsley, Jr. (1966). Digital Simulation in Hydrology: Stanford Watershed Model IV, *Tech. Rep. 39*, Dept. of Civil Engineering, Stanford University, Stanford, Calif., 210 pp.

Deardorff, J. W. (1978). Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, *J. Geophys. Res. 83*, 1889-1903.

Eagleson, P. S. (1978). Climate, soil and vegetation, Parts 1-7, *Water Resour. Res. 14*, 705-776.

Eriksson, E. (1975). Properties of land surfaces and their influence on atmospheric processes, in *The Physical Basis of Climate and Climate Modelling*, GARP Publ. Ser. No. 16, WMO-ICSU, Geneva, 198-200.

Eyre, S. R. (1968). *Vegetation and Soils*, Edward Arnold Ltd., London, 328 pp.

GARP (1975). *The Physical Basis of Climate and Climate Modelling*, GARP Publ. Ser. No. 16, WMO-ICSU, Geneva, 265 pp.

Gates, W. L., and M. E. Schlesinger (1977). Numerical simulation of the January and July global climate with a two-level atmospheric model, *J. Atmos. Sci. 34*, 36-76.

Kellog, W. W., and S. H. Schneider (1974). Climate stabilization: For better or for worse? *Science 186*, 1163-1172.

Landsberg, H. E. (1956). The climate of towns, in *Man's Role in Changing the Face of the Earth*, W. L. Thomas, ed., University of Chicago Press, Chicago, Ill., pp. 584-603.

Landsberg, H. E. (1974). Inadvertent atmospheric modification through urbanization, in *Weather and Climate Modification*, W. N. Hess, ed., John Wiley & Sons, Inc., New York, pp. 726-763.

Leith, C. E. (1975). The design of a statistical-dynamical climate model and statistical constraints on the predictability of climate, in *The Physical Basis of Climate and Climate Modelling*, GARP Publ. Ser. No. 16, WMO-ICSU, Geneva, pp. 137-141.

Lettau, H. (1969). Evaporation climatonomy: A new approach to numerical prediction of monthly evapotranspiration, runoff, and soil moisture storage, *Mon. Weather Rev. 97*, 691-699.

Lettau, H., K. Lettau, and L. C. B. Molion (1979). Amazonia's hydrologic cycle and the role of atmospheric recycling in assessing deforestation effects, *Mon. Weather Rev. 107*, 227-237.

Lorenz, E. N. (1976). Nondeterministic theories of climate change, *Quat. Res. 6*, 495-506.

Manabe, S. (1975). A study of the interaction between the hydrological cycle and climate using a mathematical model of the atmosphere, *Proc. Conf. on Weather and Food*, Endicott House, Massachusetts Institute of Technology, 9-11 May 1975, 10 pp.

Manabe, S., and J. L. Holloway, Jr. (1975). The seasonal variation of the hydrologic cycle as simulated by a global model of the atmosphere, *J. Geophys. Res. 80*, 1617-1649.

Manabe, S., and R. T. Weatherald (1975). The effects of doubling the CO_2 concentration on the climate of a general circulation model, *J. Atmos. Sci. 32*, 3-15.

Mason, B. J. (1979). Some results of climate experiments with numerical models, Overview Pap. No. 9, *Proceedings of World Climate Conference*, WMO, Geneva, 32 pp.

Morton, F. I. (1968). *Evaporation and Climate: A Study in Cause and Effect*, Scientific Ser. No. 4, Inland Waters Branch, Dept. of Energy, Mines and Resources, Ottawa, Canada, 31 pp.

Philip, J. R. (1977). Water on earth, in *Water--Planets, Plants and People*, A. K. McIntyre, ed., Australian Academy of Science, Canberra, pp. 35-59.

Rowntree, P. R., and J. A. Bolton (1978). Experiments with soil moisture anomalies over Europe, in *The GARP Programme on Numerical Experimentation. Research Activities in Atmospheric and Oceanic Modelling*, R. Asselin, ed., GARP Rep. No. 18, August 1978, Geneva, p. 63.

Sagan, C., O. B. Toon, and J. B. Pollack (1979). Anthropogenic albedo changes and the earth's climate, *Science 206*, 1363-1368.

Sasamori, T. (1970). A numerical study of atmospheric and soil boundary layers, *J. Atmos. Sci. 27*, 1122-1137.

Schickedanz, P. T. (1976). The effect of irrigation on precipitation in the Great Plains, *Final Report on NSF Grant GI:-43871*, Illinois State Water Survey, Urbana, Ill., 105 pp.

Schneider, S. H., and T. Gal-Chen (1973). Numerical experiments in climate stability, *J. Geophys. Res. 78*, 6182-6194.

Stidd, C. K. (1967). Local moisture and precipitation, Univ. *Nev. Desert Res. Inst., Preprint No. 45A*, 34 pp.

Stidd, C. K. (1975). Irrigation increases rainfall? *Science 188*, 279-280.

Stone, P. H., S. Chow, and W. J. Quirk (1977). The July climate and a comparison of the January and July climates simulated by the GISS general circulation model, *Mon. Weather Rev. 105*, 170-194.

Terstriep, P. E., M. L. Voorhees, and G. M. Bender (1976). Conventional urbanization and its effect on storm runoff, *Illinois State Water Survey, Report on Contract No. 47-26-84-390*, Urbana, Ill., 68 pp.

WMO (1979). *The Influence of Society on Climate Change and Variability*, Report of Working Group II, World Climate Conference, Geneva, 8 pp.

Walker, J., and P. R. Rowntree (1977). The effect of soil moisture on circulation and rainfall in a tropical model, *Q. J. R. Meteorol. Soc. 103*, 29-46.

Washington, W. M., and R. M. Chervin (1979). Regional climatic effects of large-scale thermal pollution: Simulation studies with the NCAR general circulation model, *J. Appl. Meteorol. 18*, 3-16.

Implications of the Vadose Zone to Water-Resource Management

3

DONALD R. NIELSEN *and* **J. W. BIGGAR**
University of California, Davis

The vadose zone is usually considered the portion of the earth's crust that is positioned between the soil surface and the groundwater table, which is usually composed of soil and unconsolidated material and is not completely saturated with water.

The vadose zone is inextricably involved in many aspects of hydrology--infiltration, evaporation, runoff, groundwater recharge, and water quality. Despite its role in the hydrologic cycle, the nature and behavior of the vadose zone remain obscure compared with other facets of hydrology. The importance of the vadose zone has either been overlooked or considered of marginal importance relative to rainfall-runoff, flood distribution, and groundwater analyses. During the past decade the vadose zone has received greater recognition--an outgrowth of the need to improve and protect the quality of water supplies.

A major factor leading to an increased interest in the vadose zone has been the disposal of solid and liquid wastes in the land and the desire to maintain water quality and to prevent environmental degradation near waste-disposal sites. Laws and regulations at both the local and federal levels have focused attention on such problems. The relatively slow movement of constituents within the water through and away from disposal sites into the vadose zone and eventually into surface waters and groundwaters is beginning to be appreciated. Recent research on the chemical, physical, and biological processes associated with waste disposal in treatment plants or in land-disposal sites has demonstrated that some discharge is inevitably transmitted by the vadose zone; therefore, zero discharge as mandated by Public Law 92-500 is a scientific impossibility. There is a general lack of theoretical and experimental understanding of the vadose zone, which would allow prediction and possibly management of the rate of advance of a pollutant through it.

The impetus to understand further the vadose zone stems from needs far greater than just the waste-disposal example cited above, where, although readily perceived by the public as important, the portion of the landscape involved with waste disposal is almost nil compared with the total domain of our land and water resources. Surface waters and groundwaters are linked by the water in the vadose zone. In the past, as well as in the present, surface waters were often diverted and managed without sufficient regard to the impact on groundwater, and vice versa. The desire for conjunctive use of surface waters and groundwaters in basin planning, development, and management is already appreciated by hydrologists, particularly in geographic areas of marginal or deficient rainfall. However, the quality of water as it moves through and over our land forces consideration of the vadose zone and the delineation of its role in water-resource management.

The transfer and retention of water and its constituents in topsoils and soils have traditionally been investigated and managed by agricultural and soil scientists with the primary emphasis on plant and crop production. They have shown little or no concern for the water that passes into the vadose zone (Black, 1968). Two notable exceptions exist. One is the application of additional water to the soil to leach excess salts out of the root zones into the vadose zone (Richards, 1954). The other is in deciding the amount and frequency of the application of pesticides, fertilizers, and other soil additives so that excessive rainfall or irrigation will not leach them out of the area where they would be beneficial to crop production (Taylor and Ashcroft, 1972; Ayers and Branson, 1973). Even in these cases, management decisions are based on the economics of crop production rather than the potential impact on the quality and quantity of water in the vadose zone.

There are two traditional hydrologic views of the vadose zone. One is that the vadose zone is a buffer for runoff and erosion as the topsoil and underlying vadose zone, through the process of infiltration, serve as a sink for water reaching the soil surface during rainfall and irrigation. The other view is that of the vadose zone as a source of water that reaches the water table with a strength equal to the difference between infiltration and evapotranspiration (Eagleson, 1970). Although both viewpoints have provided a rationale for examining the hydrologic cycle with respect to rainfall-runoff and groundwater recharge, it is obvious that such viewpoints ignore those properties of the vadose zone that impinge on the quality of the water. The quality of the water, in turn, has a direct influence on the retention and transmission of water in the vadose zone.

NATURE OF THE VADOSE ZONE

Compared with the vadose zone the water content of the topsoil oscillates over a wide range of values. Oscillations are associated with wet and dry seasons and with each rainfall or irrigation and subsequent drainage into the vadose zone and evaporation of water into the atmosphere. Absorption of water by vegetation causes additional oscillations. Depending on the relative dryness of the topsoil, it has the capacity to allow large quantities of water to infiltrate its surface before runoff begins (Philip, 1957). The infiltration rate depends not only on soil-water content but also on soil texture, pore-size distribution, and the quality of the infiltrating water (e.g., Reeve and Tamaddoni, 1965). On the other hand, the water content of the vadose zone at any one depth is nearly constant regardless of the season or how the topsoil is managed. This near constant behavior of the water content stems from the fact that the water-conducting property of the vadose zone is approximately logarithmically related to the water content (Gardner, 1959). That is, the hydraulic conductivity of a soil is greatest for water-saturated conditions and decreases by about 7 orders of magnitude as the soil-water content diminishes to an air-dry condition. Hence, any appreciable change in the rate at which water drains from the topsoil because of seasonal or management practice is easily accommodated by an increase or decrease in the water conduction rate within the vadose zone without a major change in its water content. The vadose zone has the potential to transmit both water and solutes over a wide range of rates depending on the net rate at which water leaves the topsoil.

In the topsoil and in the vadose zone, water moves predominately in a vertical direction, whereas in the water-saturated aquifer, water generally moves horizontally and at a rate proportional to the difference in elevation and the pore-size distribution of the aquifer. The vadose zone usually consists of layers of different physical and chemical properties that give rise to mixed water-saturated and unsaturated zones with water traveling horizontally even though its primary direction is downward. Transient times for vertical movement for the vadose zone are generally much greater than those for the soil profile and much smaller than those for the aquifer.

The transient time for water or solute to move through the vadose zone is directly related to the thickness of the zone. Its thickness is no greater than that of the topsoil (1 or 2 m) in many locations, especially in the eastern United States, and can extend to several hundred meters in alluvial basins in the western United States. In the eastern United States, where rainfall generally exceeds evapotranspiration and where soils are shallow, water moves quickly to streams and rivers each season. In the western United States, where rainfall is minimal and where the vadose zone is deep, water moves exceedingly slow. Hence, the rate at which a soil additive or its metabolite reaches a groundwater or surface water supply may vary by several orders of magnitude depending solely on local geography.

The vadose zone experiences a geothermal gradient (about 1°C increase in temperature for every 40-m increase in the distance from the topsoil) as well as annual oscillations of temperature. The topsoil experiences both annual and daily temperature oscillations, which enhance chemical weathering. Although there are numerous examples of theoretical and practical considerations of water flow in the presence of thermal gradients, groundwater aquifers are usually analyzed in most textbooks as if they were insensitive to temperature fluctuations (Viesmann *et al.*, 1977). Temperature has its greatest impact in the vadose zone on the viscosity of water and on biological and chemical reactions. Temperature oscillations at the soil surface have an indirect but nevertheless important effect on the vadose zone inasmuch as both chemical and biological processes are dominant within the soil profile. The topsoil has larger quantities of organic matter and an abundance of microbial species that account for the transformation rates of vegetative residues, soil additives, pesticides, solid and liquid wastes, and other inorganic and organic constituents. Because the translocation and retention of inorganic and organic constituents within the vadose zone hinges on the above rates, which to date have not been appropriately ascertained under field conditions, their prediction remains obscure and subject to conjecture. Unless exceedingly heavy volumes of solid or liquid waste are applied and leached rapidly into the vadose zone, both the groundwater aquifer and the vadose zone exhibit only inconsequential traces of organic matter and microbial species.

The vadose zone is characterized further by slow, subtle changes in constituent concentrations of its gaseous phase as compared with the topsoil, caused principally by the differing rates at which air is exchanged within the water-unsaturated profile. The primary constituent is CO_2, which directly affects pH and concentrations of bicarbonate and carbonate in the water phase that enter directly into many chemical reactions in the vadose zone.

We do not suggest that the basic scientific principles of chemical, physical, and biological phenomena within the vadose zone differ from those operating in other sectors of our natural resources, for they are identical. We encourage familiarity with classical textbooks on solution chemistry (Garrels and Christ, 1965), physical processes (Davies and Rideal, 1963), enzyme kinetics (Segal, 1974), and soil microbiology (Alexander, 1961). The recent book by Freeze and Cherry (1979) summarized the various processes that impinge on the quality of the ground water. Nevertheless, most water-resource publications do not treat the vadose zone explicity nor do they provide experimental procedures or methods by which the various phenomena can be quantified for natural field situations.

DESCRIPTIVE THEORETICAL CONCEPTS

What is known theoretically about the vadose zone has been borrowed *a priori* from many scientific and technological disciplines inasmuch as experimentation within the zone below the root depths of vegetation is nearly nonexistent.

The water in the vadose zone is not pure, but is a solution of water and dissolved solid and gaseous constituents. Moreover, vadose water cannot be considered simply as ordinary water with a few dissolved solutes, since its properties are intimately linked to the chemical and physical properties of the solid phase on which it is sorbed (Nielsen *et al.*, 1972). The impact and degree of this linkage hinge on the amount of water that is in the soil and the mineralogical composition and particle-size distribution of the solid phase. Typically, vadose water exists as films having thicknesses on the order of only a dozen water molecules. The physical properties of this water differ markedly from that which completely fills the relatively large pores in groundwater aquifers because water is a strong dipole (dipole moment of 1.87×10^{-18} esu cm) and is readily influenced by the net surface charge density of the soil particles and the numbers and kinds of its dissolved constituents. Ions in the water satisfy the surface charge on the soil particles caused by isomorphous substitution of one element for another in the crystal lattice of clays, ionization of hydroxyl ions at the surface edges of clay particles, and other mechanisms. The charge density distribution of an assemblage of soil particles gives rise to an electrical field that controls cations and ions within the water films, and it changes the configuration and properties of the water. Figure 3.1 shows distributions of cations and anions in the soil solution as a function of distance from the soil-particle surface within a water-saturated soil pore. For the more concentrated solution of 0.1*N* the impact of the electrical field is not evident at distances greater than about 50 A, while that for the dilute solution

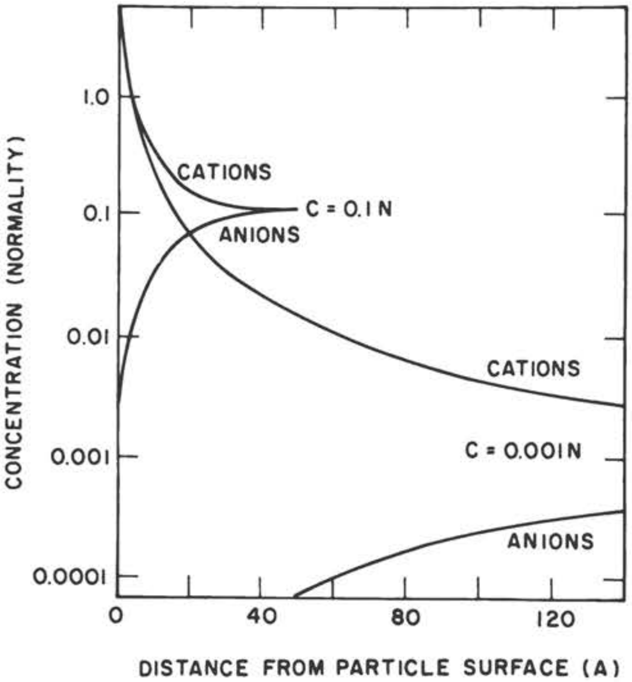


FIGURE 3.1 Concentration distributions of cations and anions in the vicinity of the boundary of a soil pore filled with water having concentrations 0.1*N* and 0.001*N*.

(0.001*N*) extends further than 200 A into the pore. As the soil-water content decreases, the cations and anions are forced to occupy a space limited by the thickness of the films on the soil-particle surfaces. Such a surface-related phenomenon gives rise to swelling pressures (Bolt and Bruggenwert, 1976), streaming potentials (Overbeek, 1952), and salt sieving (Kemper, 1960). It is well known, even for water-saturated conditions, that the hydraulic conductivity of a soil can change an order of magnitude by merely altering the concentration of the soil solution or the kinds of cations associated with the soil-particle charge density (Quirk and Schofield, 1955). Compared with water saturation having pores and lenses with thicknesses on the order of 10^3-10^4 molecules of water, the water-solute-particle surface interactions become increasingly more important as the soil becomes progressively drier and cannot be neglected at those water contents manifested in the vadose zone that are characterized by films that are 10-20 molecules thick.

Because of the complexities described above, a description of the forces acting on the vadose water to predict its retention and movement cannot be restricted as it is for groundwater aquifers to those stemming from the earth's gravitational field. Defining a force as the gradient of a scalar potential, the soil-water potential Ψ_w (ergs/gm) is defined as

$$\Psi_w = \sum_i \Psi_i = \Psi_p + \Psi_m + \Psi_s + \Psi_e + \Psi_z, \qquad (3.1)$$

where Ψ_p is the pressure potential, Ψ_m the metric potential, Ψ_s the solute potential, Ψ_e the electrical potential, and Ψ_z the gravitational potential. The first and last terms on the right-hand side of the equation are usually applied to water-saturated conditions. For the water-unsaturated vadose zone, the first term is not applicable. Assuming that water moves porportionally to the forces acting on it, for steady-state conditions, the rate at which water moves through the vadose would be

$$V_w = - \sum_i K_i \frac{d\Psi_i}{dz}, \qquad (3.2)$$

where V_w is the volumetric flux of water (cm^3/cm^2 sec) and K_i are proportionality coefficients that all depend more or less on the degree of water unsaturation and the temperature. For nonisothermal conditions, Eq. (3.2) becomes extremely complex. Although the magnitude of V_w depends on the local climate and the season, its average value usually oscillates between nil and 1 mm/day. For waste-disposal sites, irrigated agriculture, or humid areas, its value may be much larger.

The obscurity of the vadose zone can be easily explained by the fact that field techniques for measuring Ψ_m, Ψ_s, and Ψ_e remain undeveloped or are limited to special conditions (Black, 1965). Moreover, no direct means have been devised to measure V_w (Cary, 1973; Dirksen, 1972); hence, values of K_i, the transfer coefficients, have not been adequately evaluated or correlated with the composition of the vadose zone. In other words, direct measurements or calculations of the rate at which water moves through the vadose zone, based on its properties, have not yet been devised.

The above discussion of the soil-water flux was based on the assumption that no marked or sudden changes take place at the soil surface that significantly alter the soil-water content or the composition of the soil solution in the vadose zone. For such abrupt, transient conditions the soil-water content, θ, would change both with time t and depth and be described by,

$$\frac{\partial \theta}{\partial t} = - \frac{\partial V_w}{\partial z}. \qquad (3.3)$$

Most field studies to date commonly assume that Ψ_s, Ψ_e, and the temperature are constant and that $\Psi_m + \Psi_z$ equals Ψ_h, the hydraulic potential, giving rise to

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left[K\,(\theta | \Psi_m)\, \frac{\partial \Psi_h}{\partial z} \right], \qquad (3.4)$$

where $(\theta | \Psi_m)$ includes all the uncertainty of neglecting Ψ_s, Ψ_e, and temperature variations. When these terms are neglected, many textbooks and treatises are available that clearly describe the fundamental physical processes of water moving through saturated and unsaturated porous materials (Child, 1969; de Wiest, 1969; Eagleson, 1970; Kirkham and Powers, 1972). It should be emphasized that neglecting such terms for the vadose zone may simplify the analysis but may not be appropriate. For example, Eq. (3.4) applies to transient conditions whereby the soil-water content changes because of extraction or displacement of the soil solution. Such transient conditions give rise to changes in the concentration of the soil solution that in turn affect the magnitude of the hydraulic conductivity. This phenomenon is illustrated in Figure 3.2 for a clay loam whose hydraulic conductivity $K(\theta)$ was measured by Dane and Klute (1977) for two soil-solution concentrations having identical sodium adsorption ratios (Richards, 1954). The data show that the values of K diminish twentyfold when the concentration is reduced from 100 to 10 meq^{-1}. Similarly, for a given soil-solution concentration, it is well known that the value of the hydraulic conductivity depends on the distribution of cations and anions within the soil solution. Hence, analytical or numerical solutions of the above equation may be able to simulate a sequence of events at a given location because of the special manner in which the parameters are evaluated but would not be applicable nor be able to predict future events without a rigorous consideration of the ionic constituents within the soil solution. Research is required to measure Ψ_i and K_i *in situ* and to relate such values to major land forms and geologic strata.

The composition and general quality of the water that enters the vadose zone hinge on the biological and chemical reactions that dominate the topsoil. It has only been during the past quarter century that soil scientists, hydrologists, and engineers have given serious thought to how dissolved constituents in soil and vadose waters behave. Earlier investigations date back at least to the turn of the century, when Schlicter (1905) and others were interested in the dispersion of solutes associated with ground water migration. With the current environmental emphasis on coupling conservation and efficient utilization of both renewable and nonrenewable resources, the pace to understand and manage more effectively the concomitant progress of solutes and water moving through the vadose zone has accelerated. The spectrum of solutes includes fertilizers, pesticides, municipal and industrial wastes, and other organic and inorganic solutes intentionally or inadvertently added or modified in the topsoil.

In the 1950's, initial studies of the principles of leaching, hydrodynamic dispersion, and miscible displacement were confined to rather simple, artificial conditions in the laboratory (e.g., Day, 1956; Rifai *et al.*, 1956). And in large measure, current research still reflects those modest beginnings from myopic viewpoints of isolated scientific disciplines. The general perception and understanding of solute movement in the vadose zone is far less than the little-understood concepts of its water movement. A good understanding requires considerations of chemistry, physics, and microbiology. After having given such considerations to the various processes, a substantial number of parameters are inevitably introduced, giving rise to a theory that can easily be adapted to fit any set of experimental data without necessarily proving anything or providing substantive guidance for future research or technology.

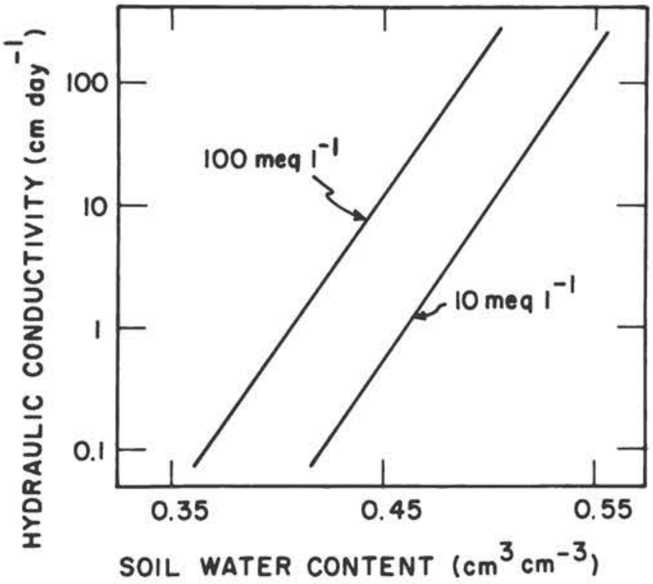


FIGURE 3.2 Soil hydraulic conductivity versus soil-water content for two soil solution concentrations.

Review articles on how solutes and water mix as they migrate through water-saturated soils and groundwater aquifers are now relatively abundant and implicitly identify areas of the unknown (e.g., Fried, 1975). The unknown, we believe, stems from the propensity of soil scientists, hydrologists, and irrigation engineers to limit their consideration and vision to their respective disciplines. For example, irrigation engineers tend to disregard the chemical interactions between soil particles and solutes in the unsaturated vadose zone, whereas soil scientists neglect the physical aspects of the zone. The hydrologist, on the hand, generally acknowledges but disregards the chemical and microbiological processes occurring in the topsoil that are important to the retention and migration of water in the vadose zone. Experimentally, in the laboratory or in the field, each discipline tends to continue to select those conditions and treatments to emphasize a single disciplinary viewpoint without really coming to grips with a more common, realistic situation or one that can be analyzed and transferred to other conditions. We illustrate the above assertions by using the following commonly used but oversimplified equation for describing the C (carbon) concentration of a solute as it is leached through a homogeneous, isothermal vadose zone:

$$\frac{\rho}{\theta}\frac{\partial s}{\partial t} + \frac{\partial C}{\partial t} = D\,\frac{\partial^2 C}{\partial z^2} - \frac{\partial vC}{\partial z} + Q, \tag{3.5}$$

where s is the solute associated with the solid phase, D the apparent diffusion coefficient, v the average pore-water velocity, Q a source or sink term, ρ the bulk density of the soil, and θ the soil-water content. Hydrologists minimizing the impact of the first and last terms utilize solutes to evaluate the behavior of D for different kinds of porous media and flow conditions and suggest improvements to the formulation of similar equations based on geometric considerations of the soil pores (Bear, 1969). Of paramount importance is the magnitude of the pore-water velocity. Virtually all research for ascertaining its value hinges on two concepts: (1) the application of equations to determine the hydraulic properties of the soil or (2) the behavior of a tracer purposely introduced or found naturally in the environment. When the former concept is used, the implications of the water-solute-solid interactions on the value of the hydraulic properties illustrated in Figure 3.2 are seldom included in the derivation of Eq. (3.5). The utility of the latter concept depends on the tracer either being "nonreactive" with the porous material or that the nature and impact of the reaction [first and last terms of Eq. (3.5)] are quantitatively known. There are few, if any, tracers within the thin water films of the vadose zone that do not interact with the soil-particle surfaces, and as the thickness of the water film changes, the relative rates of movements of the tracer and the water also change. This is illustrated in Figure 3.3, which shows the concentration distribution of an anion tracer and the water velocity distribution within a single soil pore. The distance d_s is the idealized portion of the pore containing no tracer, and d_u is the idealized portion of the water that is stagnant (Krupp *et al.*, 1972). The relative rates of water and solute moving through the pore depend on the relative values of d_s and d_u. If, indeed, $d_u < d_s$, salt sieving occurs. The magnitudes of d_s and d_u depend on the concentration and kinds of ionic species in the soil solution, the thickness of the water films, and the soil-water flux. Similar considerations hold for cationic, polar, and nonpolar tracers.

Geochemists emphasize parameters of the first term (exchange and precipitation reactions) at the expense of neglecting the right-hand side of Eq. (3.5). The rate at which a solute interacts or exchanges with the solid phase is of great interest with many issues remaining unresolved (Helfferich, 1962; Lindstrom and Boersma, 1970). Resolution of such issues usually rests on the consistency between the scale at which the phenomenon is perceived and the scale of the observations for which the perceptions are tested. Often overlooked is the fact that the surface charge characteristics of soil colloids are of two types: a constant surface charge and a variable surface potential, and a variable surface charge and a constant surface potential (van Olphen, 1977). For the latter type, because the charge is determined by the nature of the adsorbed ions, the concentration of the ions in solution, and the pH (Keng and Uehara, 1974), any displacement of the soil solution or change in the thickness of the water films will impinge directly on the parameters used to describe the first term in Eq. (3.5).

Microbiologists emphasize the last term of Eq. (3.5), which describes the growth and maintenance of a microbial population in relation to organic materials applied to the soil surface and the decay of roots of vegetation. Soil microbiologists often neglect the third term of the equation (McLaren, 1970). Prediction of the retention and movement of a given solute also hinges on the quantitative description of the solute's precursors (e.g., through hydrolysis, the precursor of ammonium is urea) and its successors (e.g., through oxidation, nitrate is the successor of ammonium). Such a prediction could be conceptually derived from a series of equations analogous to Eq. (3.5) written for each solute within the liquid or gaseous phase of the vadose zone. For example, a soluble organic nitrogeneous material could be leached into the vadose zone, transformed into ammonium, oxidized to nitrate (from oxygen stemming from the gaseous phase diffusing from the soil surface), and reduced to N_2, which moves through the soil water and into the gaseous phase and diffuses eventually back to the soil surface and out into the atmosphere. Such studies, even for the most important elements of C, H, O, S, N, and P associated with industrial and municipal wastes and leachates from plant and animal production in agriculture and ecological investigations, have received only cursory consideration in the laboratory, much less a comprehensive examination in the field.

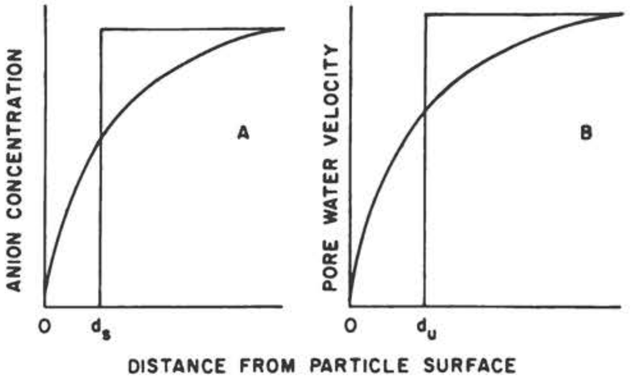


FIGURE 3.3 Anion concentration distribution (A) and pore-water velocity distribution (B) in the vicinity of the boundary of a soil pore. Anion exclusion distance d_s and immobile water distance d_u are shown for idealized step functions.

Finally, regardless of the disciplinary emphasis given in Eq. (3.5), experimentation for water-unsaturated soils has been largely confined to laboratory studies involving rates of water movement of an order of magnitude greater than is usually manifested in the vadose zone and concentrations of solutes and their relative abundances compared with other solutes in excess of realistic connate values.

It should also be recognized that partial, nonlinear, second and higher order equations similar to Eq. (3.5) are now amenable to numerical solutions that were not even attempted as recently as a decade ago (van Genuchten, 1978). Hence, the capability to solve such equations and to simulate the behavior of water and solutes in the vadose zone has surpassed our technology to derive more appropriate equations and to verify their numerical solutions. We illustrate this latter statement by referring to Eq. (3.5) or any one of a number of its homologues in the scientific literature. We refer specifically to the scale of measurement of any of the terms in Eq. (3.5). Reduced to its simplest form for a "nonreactive constituent" being displaced in the vadose zone, Eq. (3.5) becomes

$$\frac{\partial C}{\partial t} = D\,\frac{\partial^2 C}{\partial z^2} - \frac{\partial vC}{\partial z}, \tag{3.6}$$

for which time t and depth z definitely are deterministic and unambiguous compared with C, v, and D. The values C, v, and D hinge on the scale of measurements and the method by which a spatial or temporal average is obtained. At the molecular level or at an aggregation equivalent to the thickness of a water film in the vadose zone, the values of all three terms are dependent on the distance within the film from the soil-particle surface at which the terms are evaluated. At an aggregation one or two orders of magnitude larger, their values hinge on the localized characteristics of the particular volume of soil sampled. At still one or two larger orders of magnitude, their values would differ appreciably compared with those estimated at the molecular level. To date, no publications are available that report the scale of measurement relative to its applicability in terms of Eqs. (3.5) or (3.6), or related formulations.

Regardless of scale considerations, there is a lack of experimentation to ascertain parametric values in Eqs. (3.5) or (3.6). For water-unsaturated soils or granular geologic materials, most experimentation has been conducted in the laboratory. In fact, most laboratory studies have been limited to water-saturated materials or to systems evaluating the simultaneous solutions of simplified equations such as Eqs. (3.4) and (3.6) for infiltration (Kirda *et al.*, 1973; Smiles *et al.*, 1978). Because of the complexity of experimental techniques and the inordinate amount of time needed to deal with water-unsaturated flow conditions, the information used to predict water and solute movement in the vadose zone usually stems from water-saturated experiments. For example, from observations of the migration of dye and other soluble materials added to water tables in field situations as well as observations in water-saturated sand columns in the laboratory, the behavior of D in Eq. (3.6) is thought to be

$$D = D_0 + mv^n, \tag{3.7}$$

where D_0 is the molecular diffusion coefficient within a porous material, m a constant that depends on the properties of the material, and n a constant usually taken as a value between 1 and 2 (Fried and Combarnous, 1971). For unsaturated soils, the values of D_0, m, and n all depend on soil-water content (Yule and Gardner, 1978). The value of D_0 is also known to vary with the distribution of solute species in the water (Kemper and van Schaik, 1966). The second term, commonly called mechanical dispersion (Freeze and Cherry, 1979), remains virtually unknown for water-unsaturated conditions. In the vadose zone, the magnitude of v may be either smaller or larger than D_0, the description of m has not been attempted, and, to date there has been only one comprehensive field study of the magnitude of n for a nearly water-saturated soil. The results of that study (Biggar and Nielsen, 1976) relative to Eq. (3.7) are shown in Figure 3.4 for 358 observations of pairs of values of D and v measured at six depths from the topsoil to 2 m within a 150-ha field. The value of m is two orders of magnitude larger than those obtained from laboratory studies with the value of n being close to that reported by Fried and Combarnous (1971). The striking feature of Figure 3.4 is the extreme scatter of the data with the values of D ranging between 0.5 to 13,000 cm^2/day. The significance of such variability on the accuracy and precision of solutions of Eq. (3.6) was recently demonstrated by Warrick and Amoozegar-Fard (1979).

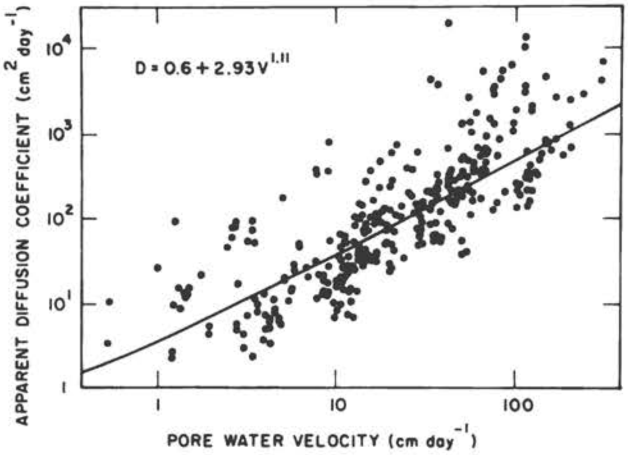


FIGURE 3.4 Apparent diffusion coefficient versus pore-water velocity measured at 358 locations within a 150-ha field.

The formulation of Eq. (3.5) or (3.6) is subject to further investigation as the water velocity distribution within the pores of the soil is not even adressed for different degrees of water saturation or for different magnitudes of v. Field studies conducted more than 80 yrs ago demonstrated that several small applications of water-leached solutes more efficiently per unit water applied than did one large water application. Means and Holmes (1900) suggested that a significant fraction of the soil pores did not conduct water and that the solutes that they contained were leached only after they migrated by diffusion into pores containing water that was mobile. A more recent field study by Miller *et al.* (1965) showed similar results and related the leaching efficiency of chloride to the degree of water saturation and the soil-water flux during the leaching process. They applied a chloride salt to the surface of a soil and observed its movement through the soil profile under two different soil-moisture regimes (Figure 3.5). Applying the water slowly (Figure 3.5A) compared with a more rapid application and at a greater soil-water content (Figure 3.5B) leached the chloride deeper. Comparing Figures 3.5A and 3.5B, it is apparent that 60 cm of water applied slowly leached the salt to a greater depth than did 90 cm of water applied more rapidly. The processes responsible for the differences in behavior may be attributed to the water-solute-solid interactions

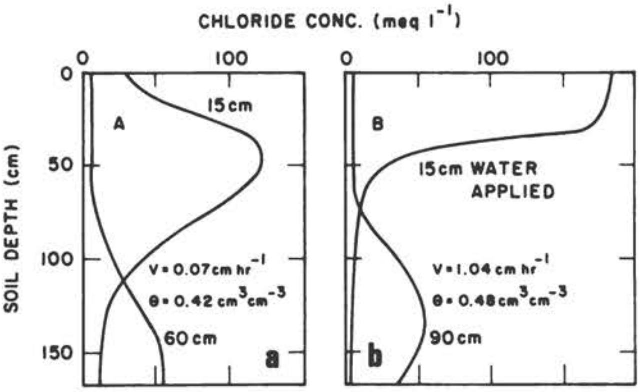


FIGURE 3.5 Chloride distributions within a soil profile stemming from a pulse of surface-applied chloride leached under two different soil-water regimes.

discussed above as well as recognizing that relative amounts of water may be stagnant within soil aggregates (Philip, 1968), dead-end pores (Coats and Smith, 1964), and isolated regions in unsaturated soils (Gaudet *et al.*, 1977). Moreover, several investigators have shown that infiltrating water moves preferentially through a limited number of pore sequences bypassing much of the soil mass (Anderson and Bouma, 1977; Kanchauasert *et al.*, 1978; Quissenberry and Phillips, 1976; Scotter, 1978; Wild, 1972). Equations analogous to Eq. (3.5) that account for zones of mobile and immobile water have also been derived and solved for special restrictive conditions (Coats and Smith, 1964; Villermaux and Swaaij, 1969; Bennet and Goodridge, 1970; van Genuchten and Wierenga, 1976; de Smedt and Wierenga, 1979; and several others). The application of such equations to the vadose zone awaits further investigation.

OUTLOOK FOR FUTURE RESEARCH

The utility of equations similar to Eqs. (3.2), (3.3), and (3.5) to analyze and manage water and solutes moving through the vadose zone hinge, as a minimum, on the development of field technology to measure or estimate the soil water potential and its components, the flux of water and its solutes, and the chemical reactions of the solutes as a function of space and time. We have seen that the value of hydraulic conductivity can be altered by at least twentyfold depending on the concentration and species of the solutes in the unsaturated zone. We also recognize that the net charge density of soil-particle surfaces can be altered and their sign even reversed from negative to positive depending on the concentration of the solutes and the pH of the unsaturated soil-water system. Hence, for vadose zones of more or less well-behaved, solute-water-soil matrix interactions it would not be unexpected that the relative movements of water and solute would vary on the order of 20 percent provided the flux of water could be appropriately ascertained. Because of the uncertainty of the hydraulic conductivity in the vadose zone, coupled with little or no verification of the magnitude of the solute-water-soil matrix interactions, standard modeling approaches are unfounded and rendered ineffective without improved experimental observations. It is perhaps satisfying to some to simulate the retention and migration of water and solutes in the vadose zone, but it is also equally disturbing to others that field verification of such simulations have seldom been attempted much less been successful or utilized by groundwater managers in predicting the impact of soil-surface events and vadose-water travel times on the arrival of constituents at groundwater interfaces. Even with the development of acceptable methods for measuring water and solute fluxes and concentrations, it is also becoming clear that there is a need to analyze and simulate vadose-water regimes using stochastic approaches rather than the deterministic systems used traditionally by agriculturists and hydrologists (e.g., Bresler and Dagan, 1979). The primary reason for the stochastic approach is the realization that the vadose zone is exceedingly heterogeneous. This heterogeneity gives rise to large uncertainties in both inputs and outputs of numerical and analytical models. Spectral analyses, Monte Carlo simulations, perturbation methods, and deterministic equations containing regionalized parameters and variables need to be developed for the vadose zone.

How should a vadose soil-water property be defined or measured to represent a distinct region of a watershed? Do particular values measured at one site represent the watershed? How many sites need to be measured to represent the watershed? How meaningful are average values determined for a few sites if the fluctuations are large over the watershed? These questions cannot be answered using statistical analyses that assume observations are independent of their spatial locations unless shown experimentally. We expect geostatistical techniques used for in the vadose zone to provide answers relative to the number, size, and spacing of observations in relation to morphological criteria for the characterization of soil and geologic mapping units. Identifying appropriate scales of observations for the unsaturated zone remains a major task.

In the immediate future we expect progress to be made primarily through the following approaches. First, we envisage miscible displacement techniques currently being used on surface soils where biological, chemical, and physical processes are integrated concomitantly to be extended to *in situ* vadose zones with depths of 10-30 m where travel times are on the order of a year or two. Such field investigations could use existing technology in locations judiciously selected on the basis of accurate land-management history that is typically available in agricultural and watershed experiment stations. Second, samples of unconsolidated sediments and materials described and categorized geologically could be brought to the laboratory and analyzed in relatively large columns subjected to soil-water contents and fluxes prevailing in the vadose. At matrix potentials on the order of -100 mb the various chemical-physical reactions that control and transport processes could be ascertained for major geologic formations. Third, we foresee the re-examination of geologic and hydrologic borehole loggings in relation to soil-surface events ranging from management decisions to climatological data. Emphasis would be given to the dynamics of the transport processes rather than the capacity or storage characteristics of the profiles. Closely associated with this approach is the need to re-examine the migration of radioactive materials from nuclear testing, where distances and travel times are effectively known and could be related to other solutes.

Finally, progress toward the understanding and management of the vadose zone will be substantially improved when principles of chemistry, physics, microbiology, and hydrology transcend the curricula of agriculturists, hydrologists, and engineers. The microbiologically and chemically induced reactions in the soil that control the quality of the recharge must be understood and coupled with the physical-chemical processes that control the transient nature of the vadose water and its constituents. With the above pedagogic elements common to the various professions, each professional viewpoint of the vadose zone should become more clearly focused and eventually would lead to a melding of the various viewpoints into rational planning and management acceptable to public policies.

ACKNOWLEDGMENT

The authors wish to thank T. Usselman for his efforts in improving the manuscript.

REFERENCES

Alexander, M. (1961). *Introduction to Soil Microbiology*, John Wiley & Sons, Inc., New York.

Anderson, J. L., and J. Bouma (1977). Water movement through pedal soils: 1. Saturated flow, *Soil Sci. Soc. Am. J. 41*, 413-418.

Ayers, R. S., and R. L. Branson (1973). Nitrates in the upper Santa Ana Basin in relation to groundwater pollution, *Calif. Agr. Exp. Sta. Bull. 861*.

Baumol, W. J., and W. E. Oates (1971). The use of standards and prices for protection of the environment, *Swed. J. Econ. 1*, 42-54.

Bear, J. (1969). Hydrodynamic dispersion, in *Flow through Porous Media*, R. J. M. de Wiest, ed., Academic Press, New York.

Bennet, A., and F. Goodridge (1970). Hydrodynamic and mass transfer studies in packed adsorption columns: 1. Axial liquid dispersion, *Trans. Inst. Chem. Eng. 48*, 232-244.

Biggar, J. W., and D. R. Nielsen (1976). Spatial variability of the leaching characteristics of a field soil, *Water Resour. Res. 12*, 78-84.

Black, C. A., ed. (1965). *Methods of Soil Analysis Part 1*, American Soc. Agronomy, Madison, Wisc.

Black, C. A. (1968). *Soil-Plant Relationships*, John Wiley & Sons, Inc., New York.

Bolt, G. H., and M. G. M. Bruggenwert (1976). *Soil Chemistry, Part A: Basic Elements*, Developments in Soil Science Vol. 5A, Elsevier Scientific Publishing Co., Amsterdam.

Bresler, F., and G. Datan (1979). Solute dispersion in unsaturated heterogeneous soil at field scale: II. Applications, *Soil Sci. Soc. Am. J. 43*, 467-472.

Cary, J. W. (1973). Soil water flow meters with thermocouple outputs, *Soil Sci. Soc. Am. Proc. 37*, 176-181.

Childs, E. C. (1969). *The Physical Basis of Soil Water Phenomena*, Wiley-Interscience, New York.

Coats, K. H., and B. D. Smith (1964). Dead-end pore volume and dispersion in porous media, *Soc. Petrol. Eng. J. 4*, 73-84.

Dane, J. H., and A. Klute (1977). Salt effects on the hydraulic properties of a swelling soil, *Soil Sci. Soc. Am. J. 41*, 1043-1049.

Davies, J. T., and E. K. Rideal (1963). *Interfacial Phenomena*, Academic Press, New York.

Day, P. R. (1956). Dispersion of a moving salt water boundary advancing through a saturated sand, *Trans. Am. Geophy. Union 37*, 595-601.

de Smedt, F., and P. J. Wierenga (1979). Mass transfer in porous media with immobile water, *J. Hydrol. 41*, 59-67.

de Wiest, R. J. M. (1969). *Flow through Porous Media*, Academic Press, New York.

Dirksen, C. (1972). A versatile soil water flux meter, *Proc. Second Symp. on Fundamentals of Transport Pnenomena in Porous Media 2*, IAHR, ISSS, University of Guelph, Ontario, pp. 425-442..

Eagleson, P. S. (1970). *Dynamic Hydrology*, McGraw Hill Book Co., New York.

Freeze, R. A., and J. A. Cherry (1979). *Groundwater*, Prentice-Hall, Inc., Englewood, N.J.

Fried, J. J. (1975). Groundwater pollution, in *Developments in Water Science 4*, Elsevier Scientific Publishing Co., Amsterdam.

Fried, J. J., and M. A. Combarnous (1971). Dispersion in porous media, in *Advances in Hydroscience*, Ven Te Chow, ed., pp. 170-282.

Gardner, W. R. (1959). Solutions of the flow equation for drying of soils and other porous media, *Soil Sci. Soc. Am. Proc. 23*, 183-185.

Garrels, R. M., and C. L. Christ (1965). *Solutions, Minerals and Equilibria*, Harper and Row, New York.

Gaudet, J. P., H. Jegat, G. Vachaud, and P. J. Wierenga (1977). Solute transfer, with exchange between mobile and stagnant water, through unsaturated sand, *Soil Sci. Soc. Am. J. 41*, 665-671.

Helfferich, F. (1962). *Ion Exchange*, McGraw-Hill Book Co., Inc., New York.

Kanchanasut, P., D. R. Scotter, and R. W. Tillman (1978). Preferential solute movement through larger soil voids. II. Experiments with saturated soils, *Aust. J. Soil Res. 16*, 269-276.

Kemper, W. D. (1960). Water and ion movement in thin films as influenced by the electrostatic charge and diffuse layer of cations associated with clay mineral surface, *Soil Sci. Soc. Am. Proc. 24*, 10-16.

Kemper, W. D., and J. C. van Schaik (1966). Diffusion of salt in clay-water systems, *Soil Sci. Soc. Am. Proc. 30*, 535-540.

Keng, J. C. W., and G. Uehara (1974). Chemistry, mineralogy and taxonomy of oxisols and utisols, *Soil Crop. Sci. Soc. Fla. Proc. 33*, 119-126.

Kirda, C., D. R. Nielsen, and J. W. Biggar (1973). Simultaneous transport of chloride and water during infiltration, *Soil Sci. Soc. Am. Proc. 37*, 339-345.

Kirkham, D., and W. L. Powers (1972). *Advanced Soil Physics*, Wiley-Interscience, New York.

Krupp, H. K., J. W. Biggar, and D. R. Nielsen (1972). Relative flow rates of salt and water in soils, *Soil Sci. Soc. Am. Proc. 36*, 412-417.

Lindstrom, F. T., and L. Boersma (1970). Theory of chemical transport with simultaneous sorption in a water saturated porous medium, *Soil Sci. 110*, 1-9.

McLaren, A. D. (1970). Temporal and vectorial reactions of nitrogen in soil: A review, *Can. J. Soil Sci. 50*, 97-109.

Means, T. H., and J. G. Holmes (1900). Soil Survey Around Fresno, California, in *U.S. House Document 107*, U.S. Department of Agriculture, Field Operations of the Division of Soils, No. 526., pp. 333-384.

Miller, R. J., J. W. Biggar, and D. R. Nielsen (1965). Chloride displacement in Panoche clay loam in relation to water movement and distribution, *Water Resour. Res. 1*, 63-72.

Nielsen, D. R., R. D. Jackson, J. W. Cary, and D. D. Evans (1972). *Soil Water*, American Society of Agronomy, Madison, Wisc.

Overbeek, J. T. G. (1952). Electrokinetic phenomena, in *Colloid Science, Vol. 1*, H. R. Kruyt, ed., Elsevier Scientific Publishing Co., Amsterdam, pp. 194-207.

Philip, J. R. (1957). The theory of infiltration: 5. The influence of the initial moisture content, *Soil Sci. 84*, 329-339.

Philip, J. R. (1968). Diffusion, dead-end pores, and linearized absorption in aggregated media, *Aust. J. Soil Res. 6*, 21-30.

Quirk, J. P., and R. K. Schofield (1955). The effect of electrolyte concentration on soil permeability, *J. Soil Sci. 6*, 163-178.

Quissenberry, V. L., and R. E. Phillips, (1976). Percolation of simulated rainfall under field conditions, *Soil Sci. Soc. Am. Proc. 40*, 484-489.

Reeve, R. C., and G. H. Tamaddoni (1965). Effect of electrolyte concentration on laboratory permeability and field intake rate of a sodic soil, *Soil Sci. 99*, 261-266.

Richards, L. A., ed. (1954). *Diagnosis and Improvement of Saline and Alkali Soils*, USDA Agriculture Handbook 60, U.S. Government Printing Office, Washington, D.C.

Rifai, M. N. E., W. J. Kaufman, and D. K. Todd (1956). *Dispersion Phenomena in Laminar Flow througn Porous Media*, Rep. No. 3, I.E.R. Ser. 90, Sanitary Eng. Res. Lab., University of California, Berkeley.

Schlicter, C. S. (1905). Field measurements of the rate of movement of underground waters, *U.S. Geol. Surv. Water Supply and Irrigation Pap. No. 140*, Washington, D.C.

Scotter, D. R. (1978). Preferential solute movement through larger soil voids. I. Some computations using simple theory, *Aust. J. Soil Res. 16*, 257-267.

Segal, J. H. (1974). *Enzyme Kinetics*, Wiley-Interscience, New York.

Smiles, D. E., J. R. Philip, J. H. Knight, and D. E. Elrick (1978). Hydrodynamic dispersion during absorption of water by soil, *Soil Sci. Soc. Am. J. 42*, 229-234.

Taylor, S. A., and G. L. Ashcroft (1972). *Physical Edaphology*, W. H. Freeman and Co., San Francisco.

Todd, D. K. (1959). *Ground Water Hydrology*, John Wiley & Sons, Inc., New York.

van Genuchten, M. T. (1978). Mass transport in saturated-unsaturated media: One-dimensional solutions, *Water Resour. Prog. Rep. 78-WR-11*, Princeton University, Princeton, N.J.

van Genuchten, M. T., and P. J. Wierenga (1976). Mass transfer studies in sorbing media: 1. Analytical solutions, *Soil Sci. Soc. Am. J. 40*, 473-480.

van Olphen, H. (1977). *An Introduction to Clay Colloid Chemistry*, John Wiley & Sons, Inc., New York.

Viessmann, W., Jr., J. W. Knapp, G. L. Lewis, and T. E. Harbaugh (1977). *Introduction to Hydrology*, IEP-A Dun-Donnelley Publisher, New York.

Villermaux, J., and W. P. M. Swaaij (1969). Modele representatif de la distribution des temps de sejour dans un reacteur semi-infini a dispersion axiale avec zones stagnantes. Application a l'ecoulement ruisselent dans des colomes d'anneaux Rasching, *Chem. Eng. Sci. 24*, 1097-1101.

Warrick, W. A., and A. Amoozegar-Fard (1979). Dissipation of soil solutes for spatially-variable, apparent diffusion coefficients, *Agronomy Abstracts*, Soil Sci. Soc. Am., Ft. Collins, Colo., p. 135.

Wild, A. (1972). Nitrate leaching under bare fallow at a site in northern Nigeria, *J. Soil Sci. 23*, 315-324.

Yule, D. F., and W. R. Gardner, (1978). Longitudinal and transverse dispersion coefficient in unsaturated Plainsfield sand, *Water Resour. Res. 14*, 582-588.

Groundwater: The Water-Budget Myth

4

JOHN D. BREDEHOEFT
U.S. Geological Survey

STEPHEN S. PAPADOPULOS
S. S. Papadopulos and Associates, Inc.

H. H. COOPER, JR.
U.S. Geological Survey

INTRODUCTION

Water-resource scientists are concerned that some basic principles are being overlooked by water managers. Rather than discuss the scope of groundwater hydrology, we have chosen to focus on a common misconception to illustrate the point.

Perhaps the most common misconception in groundwater hydrology is that a water budget of an area determines the magnitude of possible groundwater development. Several well-known hydrologists have addressed this misconception and attempted to dispel it. Somehow, though, it persists and continues to color decisions by the water-management community. The laws governing the development of groundwater in Nevada as well as several other states are based on the idea that pumping within a groundwater basin shall not exceed the recharge. It is the intent of this paper to re-examine the issue.

HISTORICAL PERSPECTIVE

Theis (1940) addressed the subject:

> Under natural conditions . . . previous to development by wells, aquifers are in a state of approximate dynamic equilibrium. Discharge by wells is thus a new discharge superimposed upon a previously stable system, and it must be balanced by an increase in the recharge of the aquifer, or by a decrease in the old natural discharge, or by loss of storage in the aquifer, or by a combination of these.

Brown (1963) attempted to illustrate these points by demonstrating that (1) under virgin conditions the height of the water table is a function of the recharge and transmissivity, and recharge is balanced by discharge from the aquifer; (2) the effects of groundwater development are superimposed upon these virgin conditions; and (3) the rate at which the hydrologic system reaches a new steady state depends on the rate at which the natural discharge (in his example to a stream) can be captured by the cone of depression. Brown's argument, which was highly technical, was essentially ignored by many hydrologists.

Bredehoeft and Young (1970) re-examined the issue and restated Theis's conclusions:

> Under virgin conditions, steady state prevails in most groundwater systems, and natural re-

charge is equal to the natural discharge. We can write the following expression for the system as a whole

$$R_0 - D_0 = 0, \qquad (4.1)$$

where R_0 is the mean recharge under virgin conditions and D_0 is the mean discharge under virgin conditions.

Some disturbance of the system is necessary to have a development. At some time after the start of pumping we can write the following expression:

$$(R_0 + \Delta R_0) - (D_0 + \Delta D_0) - Q + \frac{dV}{dt} = 0, \qquad (4.2)$$

where ΔR_0 is the change in the mean recharge, ΔD_0 the change in the mean discharge, Q the rate of withdrawal due to development, and dV/dt the rate of change in storage in the system. From Eqs. (4.1) and (4.2) we can obtain

$$\Delta R_0 - \Delta D_0 - Q + \frac{dV}{dt} = 0. \qquad (4.3)$$

Assuming water-table conditions we can them compute an average drawdown for the system as a whole in the following manner:

$$S_a = \Delta V_t/(S_y \cdot A_b), \qquad (4.4)$$

where s_a is average basinwide drawdown, ΔV_t the volume removed from storage at time t, S_y the specific yield of the aquifer, and A_b the area of the basin. Such an input-output analysis treats the system just as we would treat a surface water reservoir. The response of the system is assumed to take place rapidly with effects equally distributed throughout the basin. In most groundwater systems the response is not equally distributed.

RESPONSE OF GROUNDWATER SYSTEMS

In groundwater systems the decline of water levels in a basin because of withdrawal will occur over a period of years, decades, or even centuries. Some water must be taken from storage in the system to create gradients toward a well. There are two implications to be gathered from these facts: (1) some water must always be mined to create a development, and (2) the time delays in a groundwater system differ from those in surface-water systems.

It is apparent from Eq. (4.3) that the virgin rates of recharge R_0 or discharge D_0 are not of paramount importance in groundwater investigations. For the system to reach some new equilibrium, which we define as $dV/dt = 0$, there must be some change in the virgin rate of recharge and/or the rate of discharge D_0. It is these changes, ΔR_0 and ΔD_0, that are interesting.

The response of groundwater systems depends on the aquifer parameters (transmissivity and storage coefficient), the boundary conditions, and the positioning of the development within the system.

Lohman (1972a), referring to the High Plains of Texas and New Mexico, made the point again. The following discussion is a synopsis of Lohman's argument taken from Bulletin 16 of the U.S. Water Resources Council (1973):

> Withdrawals cannot exceed the rates of recharge or discharge for a prolonged period of time without resultant "mining" of ground water. Adjustments in recharge and discharge rates as a result of pumping can be referred to as capture, and, inasmuch as sustained yield is limited by capture and cannot exceed it, estimates of capture are fundamentally important to quantitative groundwater analysis and planning for long-term water supply.
>
> Decline of water levels in response to sustained withdrawal may continue over a long period of time. At first, some water must be taken from storage in the system to create gradients toward pumping wells. Two important implications of these statements concerning a long-term water supply are that (1) some water must be removed from storage in the system to develop a groundwater supply, and (2) time delays in areal distribution of pumping effects in many groundwater systems demonstrate that balanced (equilibrium or steady-state) conditions of flow do not ordinarily exist. In the clearest examples, water levels decline drastically, and some wells go dry long before the system as a whole reaches a new equilibrium balance between replenishment and natural and imposed discharge rates.
>
> The most well-known example of such a condition of nonequilibrium is the major groundwater development of the southern High Plains of Texas and New Mexico. Water is contained in extensive deposits (the Ogallala formation) underlying the plains (Figure 4.1). Average thickness of these deposits is about 300 feet. They consist of silt, sand, and gravel and form a groundwater reservoir of moderate permeability. The reservoir rests on relatively impermeable rock and constitutes the only large source of groundwater available to the area.
>
> The southern High Plains slope gently from west to east, cut off from external sources of water upstream and downstream by escarpments,

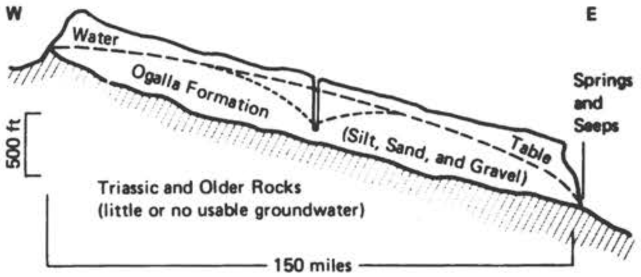


FIGURE 4.1 Development of groundwater in the southern High Plains of Texas and New Mexico. Withdrawal has resulted in a pronounced decline of water levels in the middle of the southern High Plains, but it has had little effect on the gradient to the east (natural discharge) or on natural recharge.

as illustrated in Figure 4.1. Replenishment is dependent on the scanty precipitation, and total recharge in the southern High Plains is extremely small in comparison with the enormous imposed discharge (pumping for irrigation). Total recharge is equivalent to only a fraction of an inch of water per year over the whole of the High Plains. The natural discharge, of the same order as the recharge, continues from seeps and springs along the eastern escarpment.

Withdrawal by pumping has increased rapidly in the past 50 years and at present amounts to about 1.5 trillion gallons per year (4.6 million acre-feet per year). The withdrawal has resulted in a pronounced decline of water levels in the middle of the Plains, where pumping is heaviest (and where the increase in cost of pumping has been greatest). Little additional natural recharge can be induced into the system because the water table lies 50 feet or more beneath the land surface in most of the area, the unsaturated volume of aquifer available for possible recharge is more than ample.

Nor has natural discharge been salvaged by the lowered water levels. As may be noted in Figure 4.1, the hydraulic gradient, or water-table slope, toward the eastern escarpment has been virtually unchanged. Even if all discharge could be salvaged by pumping, however, the salvaged water would be only a small percentage of present pumpage.

THE CIRCULAR ISLAND

Perhaps the easiest way to illustrate our point further is to consider pumping groundwater on an island situated in a freshwater lake. The situation is shown schematically in Figure 4.2. An alluvial aquifer overlies bedrock of low permeability on the island. Rainfall directly on the island recharges the aquifer. Under virgin conditions, this recharge water is discharged by outflow from the aquifer into the lake. The height of the water table beneath the island is determined by the rate of recharge, the area of the island, and the transmissivity.

Under virgin conditions, we can determine a water balance for the island. From our previous notation, recharge to the island is

$$\int^{A} r_a \, dA = R_0,$$

where r_a is the average rate of recharge and A is the total area of the island. Discharge from the island is and

$$\int^{L} kh \left.\frac{\partial h}{\partial s}\right|_s dL = D_0,$$

where k is the hydraulic conductivity of the aquifer, h the height of the water table defined to be equal to the hydraulic head), and

$$\left.\frac{\partial h}{\partial s}\right|_s$$

the gradient in hydraulic head taken at the shoreline of the island (defined to be normal to the shoreline), L the total length of the shoreline, and $R_0 = D_0$.

We drill a well and begin to pump water from the aquifer on the island. A cone of depression develops and expands outward from the well. Figure 4.3 shows this cone of depression a short time after pumping has begun.

If we look at the periphery of the island, we see that until the pumping causes a significant change in the gradient in head at the shoreline the discharge continues unchanged. Gradients in hydraulic head, or saturated thickness, must be changed at the shoreline in order to change the discharge.

If we write the system balance for the entire island, at some time before the cone expands to the shoreline, we see that

$$R_0 - D_0 - Q \neq 0,$$

where Q is the rate of pumping. As neither the recharge, R_0, nor the discharge, D_0, has changed from its initial value, the water pumped, Q, is balanced by the water removed from storage.

The cone of depression will eventually change the gradients in hydraulic head at the shoreline significantly. At this time, discharge from the system begins to change. This is shown schematically in Figure 4.4. The discharge can be changed by pumping so that the system is brought into balance. At some time

$$R_0 - \int^{L} kh \left.\frac{\partial h}{\partial s}\right|_s dL = Q.$$

Since the virgin rate of recharge, R_0, equals the virgin rate of discharge, D_0, we can write

$$D_0 - \int^{L} kh \left.\frac{\partial h}{\partial s}\right|_s dL = Q,$$

where the quantity

$$R_0 - \int^{L} kh \left.\frac{\partial h}{\partial s}\right|_s dL = D_0 - \int^{L} kh \left.\frac{\partial h}{\partial s}\right|_s dL = \Delta D,$$

which we define as the "capture." The system is in balance when the capture is equal to the water pumped, i.e., $Q = \Delta D$.

The term capture is defined and discussed in *Definitions of Selected Ground-Water Terms--Revisions and Conceptual Refinements* (Lohman, 1972b):

> Water withdrawn artificially from an aquifer is derived from a decrease in storage in the aquifer, a reduction in the previous discharge from the aquifer, an increase in the recharge, or a combination of these changes. The decrease in discharge plus the increase in recharge is termed capture. Capture may occur in the form of decreases in the groundwater discharge into

streams, lakes, and the ocean, or from decreases in that component of evapotranspiration derived from the saturated zone. After a new artificial withdrawal from the aquifer has begun, the head in the aquifer will continue to decline until the new withdrawal is balanced by capture.

For the island system chosen, we can induce flow from the lake into the aquifer. In fact, the capture can be greater than the virgin recharge of discharge

$$\int^{L} kh \left.\frac{\partial h}{\partial s}\right|_{s} dL > D_0$$

or

$$\int^{L} kh \left.\frac{\partial h}{\partial s}\right|_{s} dL > R_0.$$

In fact, the magnitude of pumpage that can be sustained is determined by (1) the hydraulic conductivity of the aquifer and (2) the available drawdown, which are independent of other factors (Figure 4.5).

At first glance, this island aquifer system seems much too simple for general conclusions; however, the principles that apply to this system apply to most other aquifer systems. The ultimate production of groundwater depends on

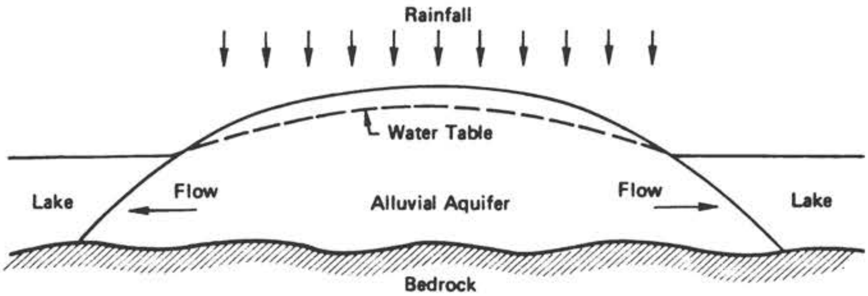


FIGURE 4.2 Cross section of an alluvial aquifer, underlain by bedrock of low permeability, on an island in a freshwater lake.

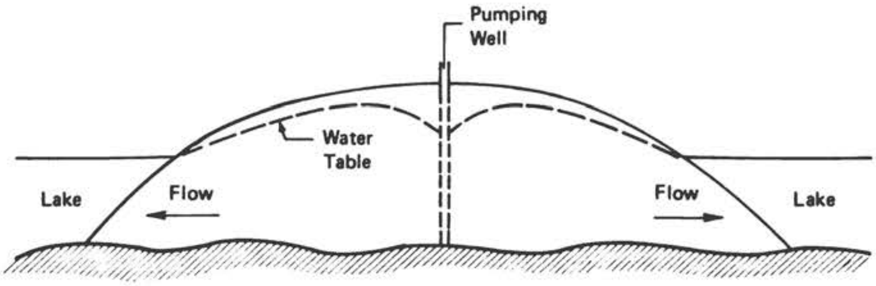


FIGURE 4.3 Cross section of the island depicting the cone of depression soon after pumping has begun.

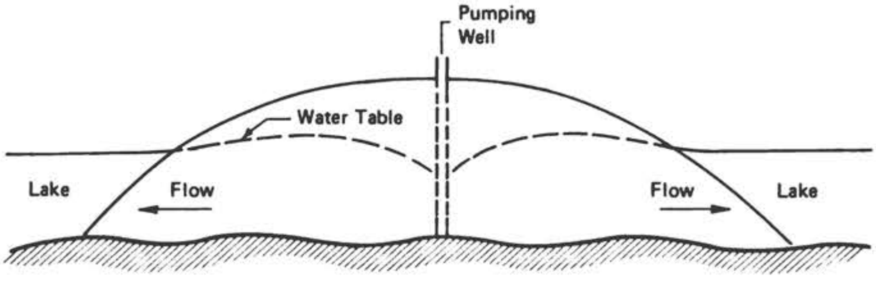


FIGURE 4.4 Cross section of the island aquifer system when the influence of pumping has reached the shoreline.

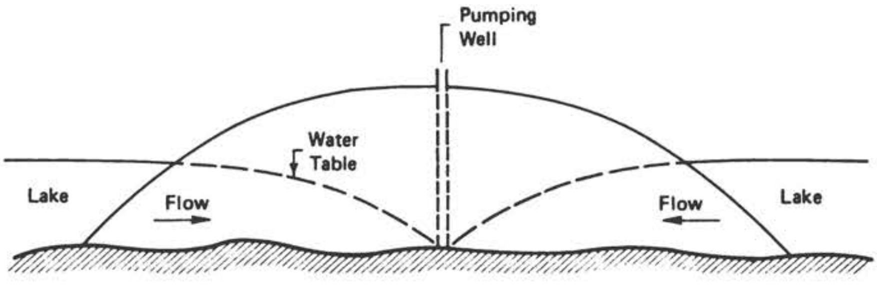


FIGURE 4.5 Schematic cross section of the island aquifer system, which illustrates that the magnitude of pumpage from this system is dependent on the available drawdown, aquifer thickness, and the hydraulic conductivity of the aquifer.

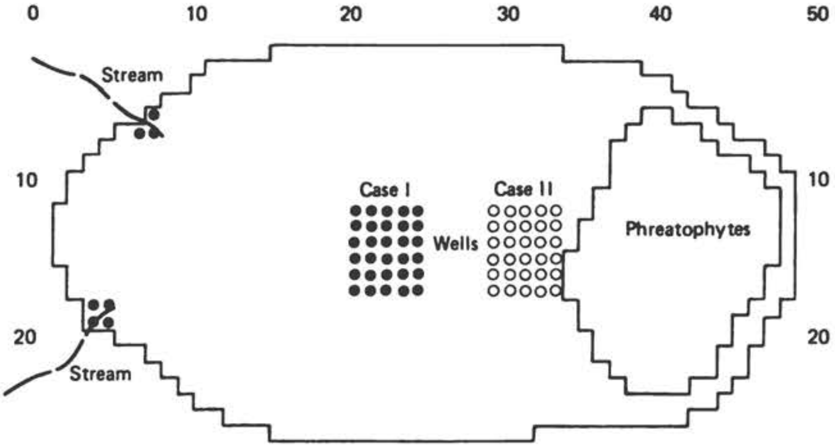


FIGURE 4.6 Schematic map of an intermontane basin showing areas of recharge, discharge, and two hypothetical water-development schemes, Case I and Case II.

how much the rate of recharge and (or) discharge can be changed--how much water can be captured. Although knowledge of the virgin rates of recharge and discharge is interesting, such knowledge is almost irrelevant in determining the sustained yield of a particular groundwater reservoir. We recognize that such a statement is contrary to much common doctrine. Somehow, we have lost or misplaced the ideas Theis stated in 1940 and before.

RESPONSE TIME

Groundwater systems generally respond much slower than other elements of the hydrologic cycle. It can take long periods of time to establish a new steady state. For this reason, groundwater hydrologists are concerned with the time-dependent dynamics of the system.

To illustrate the influence of the dynamics of a groundwater system, we have chosen a rather simple system for analysis. Consider a closed intermontane basin of the sort one might find in the western states. Under virgin conditions the system is in equilibrium: phreatophyte evapotranspiration in the lower part of the basin is equal to recharge from the two streams at the upper end (Figure 4.6).

Pumping begins in the basin, and, for simplicity, we assume the pumpage equals the recharge. The following two assumptions regarding the hydrology are made:

1. Recharge is independent of the pumping in the basin, a typical condition, especially in the arid west.
2. Phreatophyte use decreases in a linear manner (Figure 4.7) as the water levels in the vicinity decline by 1-5 ft. Phreatophyte use of water is assumed to cease when the water level is lowered 5 ft below the land surface.

The geometry and pertinent hydrologic parameters assumed for the system are shown in Table 4.1.

The system was simulated mathematically by a finite-difference approximation to the equations of flow. The equations are nonlinear and of the following form:

$$\nabla \cdot (kh\nabla h) = S\frac{\partial h}{\partial t} + W\ (x \cdot y \cdot t)$$

where k is the hydraulic conductivity, h the hydraulic head (which is equal in our case to the saturated aquifer thickness), S the storage coefficient, and W the source function (time dependent). In essence, this is a two-dimensional water-table formulation of the problem in which the change in saturated thickness within the aquifer is accounted for.

One-thousand years of operation were simulated. Stream recharge, phreatophyte-water use, pumping rate, and change in storage for the entire basin were graphed as functions of time. Two development schemes were examined: Case I, in which the pumping was more or less centered within the valley, and Case II, in which the pumping was adjacent to the phreatophyte area.

The system does not reach equilibrium until the phreatophyte-water use (the natural dis-

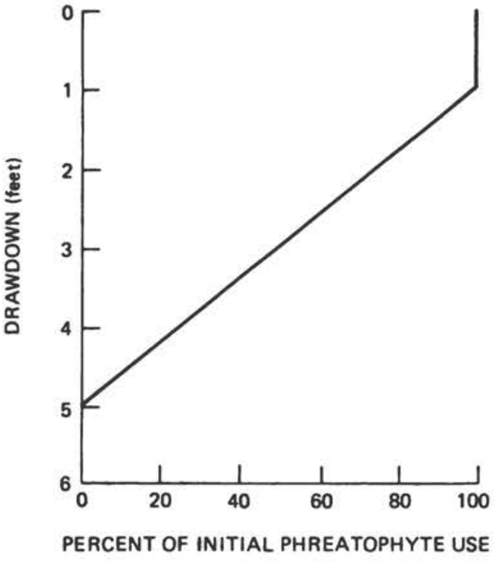


FIGURE 4.7 Assumed linear function relating phreatophyte-water use to drawdown.

charge) is entirely salvaged (captured) by pumping, i.e., phreatophyte water use equals zero (we define equilibrium as $\delta V/\delta t = 0$). In Case I, phreatophyte-water use (Figure 4.8) is still approximately 10 percent of its initial value at year 1000. In Case II it takes 500 yr for the phreatophyte-water use to be completely captured.

We can illustrate the same point by looking at the total volumes pumped from the system, along with the volume taken from storage "mined" (Figure 4.9).

In both cases, for the first 100 yr, nearly all of the water comes from storage. Obviously, as the system approaches equilibrium, the rate of change of the volume of water removed from storage also approaches zero. If the aquifer was thin, it is apparent that wells could go dry long before the system could approach equilibrium.

This example illustrates three important points:

1. The rate at which the hydrologic system can be brought into equilibrium depends on the rate at which the discharge can be captured.
2. The placement of pumping wells in the system significantly changes the dynamic response and the rate at which natural discharge can be captured.
3. Some groundwater must be mined before the system can be brought into equilibrium.

CONCLUSIONS

We have attempted to make several important points:

1. Magnitude of development depends on hydrologic effects that you want to tolerate, ultimately or at any given time (which could be dictated by economics or other factors). To calculate hydrologic effects you need to know the hydraulic properties and boundaries of the aquifer. Natural recharge and discharge at no time enter these calculations. Hence, a water budget is of little use in determining magnitude of development.
2. The magnitude of sustained groundwater pumpage generally depends on how much of the natural discharge can be captured.
3. Steady state is reached only when pumping is balanced by capture ($\Delta R_0 + \Delta D_0$), in most cases the change in recharge, ΔR_0, is small or zero, and balance must be achieved by a change in discharge, ΔD_0. Before any natural discharge can be captured, some water must be removed from storage by pumping. In many circumstances the dynamics of the groundwater system are such that long periods of time are necessary before any kind of an equilibrium condition can develop. In some circumstances

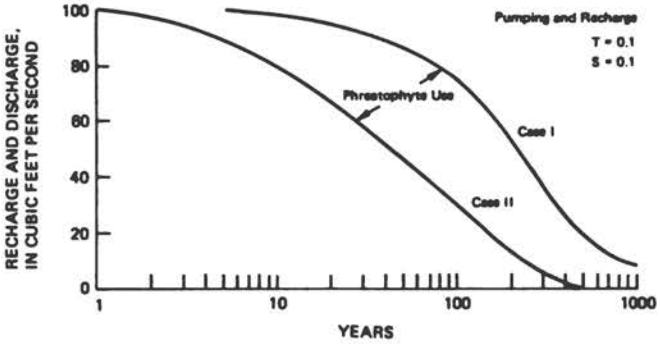


FIGURE 4.8 Plot of the rate of recharge, pumping, and phreatophyte use versus time.

TABLE 4.1 Aquifer Parameters

Basin dimensions	50 x 25 miles
Aquifer	
Hydraulic conductivity (k)	0.5×10^{-3} ft^2/sec
Storage coefficient (S)	0.1
Initial saturated thickness (h)	2000 ft
Phreatophytes	
Area	172 miles^2
Average use (annual)	100 ft^3/sec
Recharge	
Area	7 miles^2
Average recharge rate	100 ft^3/sec
Development	
Area	30 miles^2
Average pumping rate	100 ft^3/sec

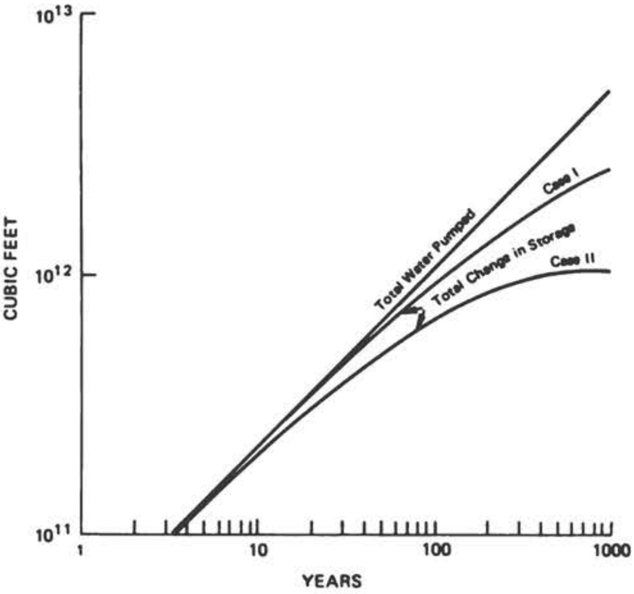


FIGURE 4.9 Total volume pumped and the change in storage versus time.

the system response is so slow that mining will continue well beyond any reasonable planning period.

These concepts must be kept in mind to manage groundwater resources adequately. Unfortunately, many of our present legal institutions do not adequately account for them.

REFERENCES

Bredehoeft, J. D., and R. A. Young (1970). A temporal allocation of ground-water--A simulation approach, *Water Resour. Res. 6*, 3-21.

Brown, R. H. (1963). The cone of depression and the area of diversion around a discharging well in an infinite strip aquifer subject to uniform recharge, *U.S. Geol. Surv. Water-Supply Pap. 1545C*, pp. C69-C85.

Lohman, S. W. (1972a). Ground-water hydraulics, *U.S. Geol. Surv. Prof. Pap. 708*, 70 pp.

Lohman, S. W. (1972b). Definitions of selected ground-water terms--Revisions and conceptual refinements, *U.S. Geol. Surv. Water-Supply Pap. 1988*, 21 pp.

Theis, C. V. (1940). The source of water derived from wells: Essential factors controlling the response of an aquifer to development, *Civil Eng. 10*, 277-280.

U.S. Water Resources Council (1973). Essentials of ground-water hydrology pertinent to water-resources planning, *U.S. Water Resour. Council Bull. No. 16*, 48 pp.

Quality of Water—Surface and Subsurface

5

RAYMOND SIEVER
Harvard University

WERNER STUMM
Swiss Federal Institute of Technology

INTRODUCTION

One of the major concerns of those involved in water management is the maintenance of water quality, the avoidance of degradation that may be hazardous to health. The primary threats to health come from the chemical loading of surface and subsurface water reservoirs derived from a multitude of sources. These chemical loads are in transit to a variety of sinks from which they are removed, long- or short-term, from the surface-subsurface water system. The major public health problem is related to the immediately toxic effects of metals and organic compounds and their long-term carcinogenic or teratogenic effects. (Though of undeniable concern, the question of disagreeable taste and esthetic disturbance, insofar as they are not health hazards, will not be discussed here.)

Our concern is not only with the immediately apprehended effects on the humans who drink the water, but on the foods we eat, related to geologic and agricultural amplification of toxic effects. Grains, fruit, and vegetables grown on the soils and livestock fed on the agricultural products may accumulate and concentrate harmful substances. Degraded water quality may also affect the quantity of agricultural production. Though the problems of poor water quality are better recognized now in the highly industrialized societies of the world, they are equally serious, if not more so, in many of the developing countries.

Unrest in the populations of various countries concerning geologic storage of nuclear-waste materials has focused attention on the general interrelations of waste disposal and water supplies. Where we formerly centered our attention on sewage disposal in relation to surface-water supplies, we must now enlarge our scope to include many more diverse solid and liquid waste disposals of a wide range of chemical substances in various kinds of geologic situations at or beneath the surface.

The scientific problem is to assess human modifications of the natural system, including the many ways in which chemical species enter the aquatic environment, how they are distributed within it, and how they leave it. We propose that the best way to do this is against a background of knowledge of the natural geochemical cycles of the elements, the geochemical baselines that were characteristic of the times well before the industrial revolution. In considering this cycle it is necessary to consider the hydrologic regime--both surface and subsurface--as a major part of a large system, the geochemical system of the outer crust of the earth, the atmosphere, and the oceans.

The engineering problem is, on the one hand, the best way in which to change our present intervention in the natural system so as to maintain high water quality. Our current pattern of water use and disposal is constantly being altered as needs for irrigation, urban water system, industry, and mining compete for available surface- and subsurface-water supplies. So far, our efforts have not been eminently successful in keeping the quality of water supplies uniformly high throughout the industrial countries, although we have done remarkably well in a few places and have preserved potable-water supplies for the overwhelming bulk of the population. As the population grows, as the range of our chemical activities continues to increase, and as hard choices have to be made on energy supplies, we will have to work even harder not to lose ground, much less to improve the quality of already degraded water supplies.

THE GEOCHEMICAL CYCLE AND WATER QUALITY

Our concepts of geochemical cycling at the surface of the earth have evolved to a comprehensive and sophisticated view of the general system by which materials from the interior of the earth are brought by geologic processes into contact with the atmosphere and hydrosphere and chemically react with them (Garrels and Mackenzie, 1971). The nature of the commonest chemical reactions are now fairly well understood in terms of equilibrium concepts, most important in knowing the state toward which the system tends (Garrels and Christ, 1965; Stumm and Morgan, 1978). In addition we have begun to understand the kinetics of the system: rates of influx and efflux into and from various reservoirs and conduits within the system, residence times of elements and compounds, and rate constants of specific chemical reactions.

Weathering

The main input to the hydrologic system for most elements is weathering. Rainwater, including dissolved components such as CO_2, operates on rocks and soils of various compositions to dissolve common and trace elements (see, for example the cycle of Sr in Figure 5.1). These elements are supplied to soil water, partly to seepage into the vadose zone and then below the groundwater table, or run off ultimately to the oceans. Natural weathering may be understood as a process of adjustment of the water-mineral mixture, by which its composition gradually changes by reference to combined equilibrium diagrams (Garrels and Mackenzie, 1967). Though we now have a good appreciation of the chemical nature of weathering, we still have much to learn about the rates of weathering in relation to dissolved CO_2, sulfur species, and physical rates of denudation by erosion.

Artificial Weathering

The introduction of toxic substances to the hydrologic regime can be likened to artificial weathering. As a factory, a city, or a farm operates, it takes materials derived artificially from the earth's crust by mining (e.g., oil,

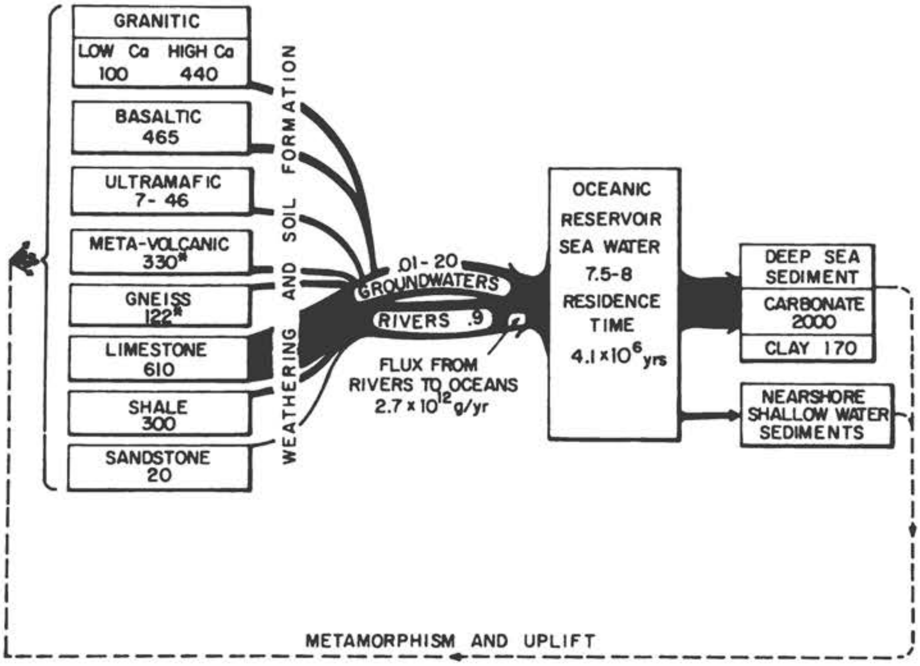


FIGURE 5.1 The geochemical cycle of Sr at the surface of the earth. Strontium concentrations are shown in parts per million (from NRC U.S. National Committee for Geochemistry, 1977).

gas, coal, metal ores, and phosphates), chemically reacts them, and dumps the products into the geochemical system in some form. The acid rains that come from burning high-sulfur fuels can in this way be seen to be the result of artificially enhanced rates of weathering of these fuels. These rates of artificial weathering are enormously greater than the natural process. For newer synthetic organic compounds, such as the chlorinated hydrocarbons, it is weathering of artifact compounds heretofore unknown to the natural world. We need to know how these materials enter the system and how they react or accumulate in various reservoirs--including the food chain.

Hydrologic Networks

Though much of the hydrologic regime is typically seen as a distribution network by which chemical materials are transported from one reaction site to another, we must avoid treating rivers, lakes, and aquifers as passive or inert conduits. In each part of the hydrologic network, there are many kinds and rates of reaction between dissolved and suspended materials in the waters and the walls of the conduit. The "walls" of a river include the algae and other aquatic plants and animals that live there and may extensively modify its composition. The "walls" of a subsurface aquifer are the minerals of the rocks through which the groundwater flows. If that water flows through a solid-waste disposal site, its "walls" will be the solid waste. Human activities in transporting, mining, processing, and fabricating various materials and processes make the networks far more complicated as trucks and trains traverse the country (Figure 5.2).

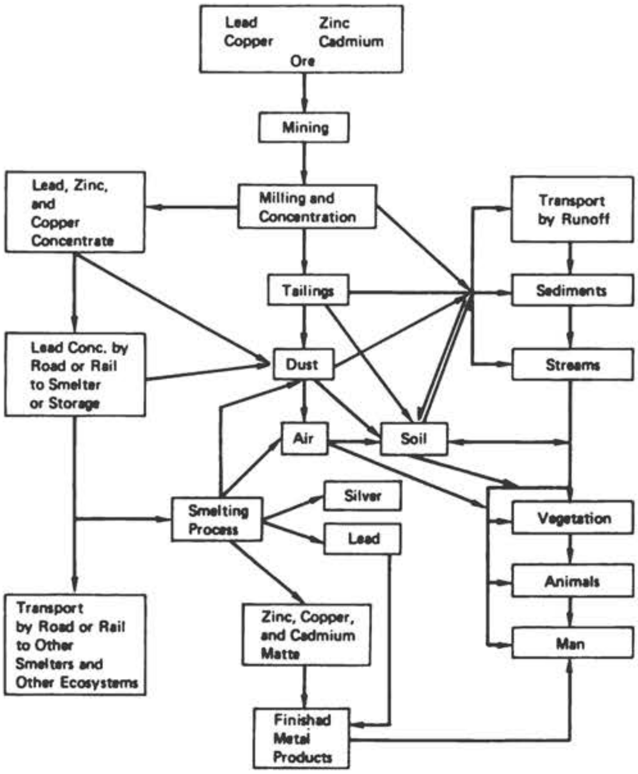


FIGURE 5.2 Sources of trace metals in the environment of the New Lead Belt of southeastern Missouri (from Jennett *et al.*, 1977).

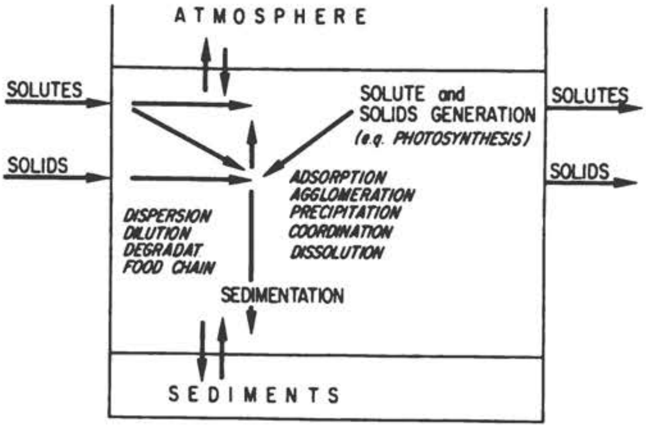


FIGURE 5.3 Factors regulating concentrations of chemical substances in surface waters.

Outputs of the Cycle

The outputs of the geochemical cycle are the sinks in which the elements or compounds are sedimented and removed from the hydrologic system. The relation of these outputs to the inputs determines the concentrations in a water body (Figure 5.3). Outputs range from sediments at the bottoms of lakes and the oceans, where they tend to be removed more or less permanently for our purposes (though only temporarily from the long-term geologic point of view) to river sediments. Deposits of floodplains, point bars, and backwaters may lie undisturbed for decades only to be picked up and further transported by the next large flood. The chemical components of waters are removed as sediment either by direct precipitation or by adsorption on solids, particularly the clays and other colloids. Direct precipitation may play a more important role in the subsurface than in the surface, at least on the continents. In the oceans, direct precipitation takes place primarily by biochemical mechanisms.

Once sedimented, the chemical components cannot be neglected, for there is abundant evidence that chemical components may diffuse back into the overlying water columns of lakes, rivers, and the oceans from the interstitial waters of the sediments.

Anthropogenic Sources of Chemical Substances

The quality of water bodies generally reflects the types and extent of human activities in their catchment areas (Blumer and Sass, 1972; Hutzinger *et al.*, 1978; Suffet, 1977). Major anthropogenic sources of chemicals, in addition to human excreta and gross industrial wastes, include drainage from animal feedlots, commercial fertilizers, and pesticide applications to farm land. During the last decades the gross pollution load has increased and its character has changed, especially in industrialized and densely settled regions; it is now more and more composed of discards of modern industrial society (synthetic chemicals, mining products, phosphates, metals, and by-products of energy production and consumption). The main source of the chemicals found in receiving waters is often not directly from industry, which has made some efforts to limit its onsite emissions, but indirectly through the distribution and use of its products via households (cleaning agents, detergents, solvents), agricultural drainage, and dispersion into the atmosphere.

A comparison of production of direct industrial versus dispersed use is instructive. The world production of industrial chemical substances is 100-200 tons per yr. Nearly half of this production occurs within the United States. Yet there are about 2000 of the 4 million known chemical substances, which are produced at a rate of more than 500 tons per yr for dispersed use.

Additional activities of man, such as deforestation, intensification of agricultural production, urbanization (e.g., the covering of land surface with asphalt), dredging of streams and damming of rivers into lake-like reservoirs, alter the networks; cause the redistribution between surface and subsurface waters; and modify the cycles that couple land, water, and atmosphere. Many of these activities contribute to an acceleration of both nutrient cycles and transport of soil nutrients. This in turn leads to the enrichment of surface waters with nutrients (phosphates, nitrates) followed by increased production of algae and aquatic weeds and by other changes in the aquatic habitats.

The burning of fossil fuels produces a variety of gaseous oxides of carbon, nitrogen, hydrogen, sulfur, compounds of heavy metals, and solid particles that become redistributed among the hydrosphere, lithosphere, and biosphere (Figure 5.4). Of particular concern are the increases in CO_2 in the atmosphere and the production of acid rain and fallout from the oxidation of S and N to sulfuric and nitric acid, respectively. Acid rain may increase the erosion rates of rocks, thereby solubilizing trace-metal ions and phosphate.

Another major identifiable pollution source is road salt in northern climates. Nine million tons of salts are spread on highways annually in the United States (100 tons per km^2 per yr in the region around Boston); eventually it finds its way to surface waters and groundwaters (Ceasar *et al.*, 1976). In the last few decades, sodium and chloride increased in Lake Ontario by 20 percent, sulfate by 6 percent, and calcium by 3 percent (NRC Planning Committee for the International Symposium on Eutrophication, 1969).

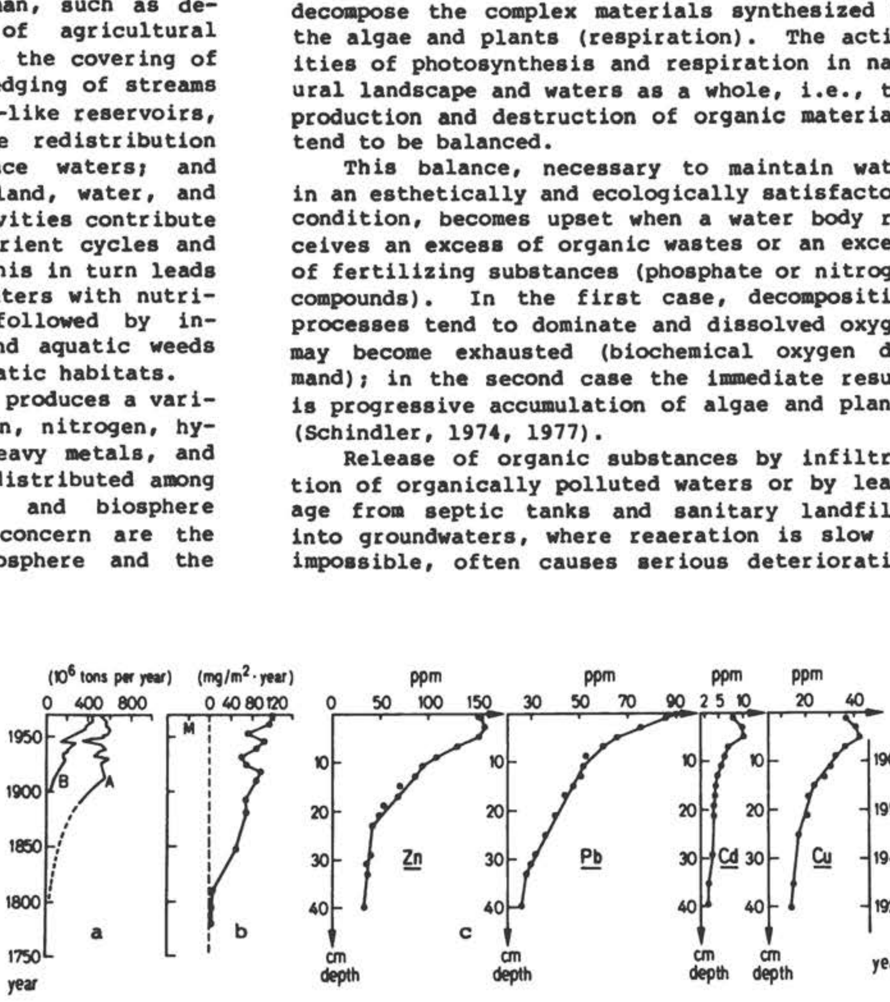
The dispersion of metals to the atmosphere as a consequence of industrial and agricultural activity appears to rival and sometimes to exceed natural mobilizations (Lantzy and Mackenzie, 1979). The increased exposure levels of metals in the environment have potential impact on the structure of ecosystems and perhaps even on human health.

Both terrestrial and aquatic ecosystems have two biotic components: (1) plants and algae able to fix light energy and manufacture biomass (photosynthesis) and (2) organisms, chiefly animals, bacteria, and fungi, that utilize and decompose the complex materials synthesized by the algae and plants (respiration). The activities of photosynthesis and respiration in natural landscape and waters as a whole, i.e., the production and destruction of organic material, tend to be balanced.

This balance, necessary to maintain water in an esthetically and ecologically satisfactory condition, becomes upset when a water body receives an excess of organic wastes or an excess of fertilizing substances (phosphate or nitrogen compounds). In the first case, decomposition processes tend to dominate and dissolved oxygen may become exhausted (biochemical oxygen demand); in the second case the immediate result is progressive accumulation of algae and plants (Schindler, 1974, 1977).

Release of organic substances by infiltration of organically polluted waters or by leakage from septic tanks and sanitary landfills into groundwaters, where reaeration is slow or impossible, often causes serious deterioration

FIGURE 5.4 Sediments as indicators of heavy metal pollution: a, European coal production during the last 170 yr; A, coal, B, lignite; b, the anthropogenic input of heavy metals to the Baltic Sea; c, Zn, Pb, Cd, and Cu in Greifensee.



in water quality. Depletion in dissolved oxygen is followed by reduction of manganese and iron oxides, imparting soluble iron and manganese to the water; nitrate also is reduced to ammonia and sulfate to hydrogen sulfide.

Many synthetic chemicals survive long enough in the environment, usually because they are not readily biodegradable. In drinking water, some of these substances may be harmful to human health; others tend to become concentrated in organisms and may reach, as a result of such biomagnification, toxic levels. Some pollutants, although they may exhibit no toxicity with regard to individual organisms, can nevertheless impair the self-regulation of aquatic ecosystems and damage life-support systems.

Responses to Pollution

The chemical-hydrologic system responds to perturbations in various ways depending on the nature of the chemical load, the dynamics of the water reservoir, the size of the reservoir, and the solid materials suspended or on the walls. Perhaps the most difficult to evaluate is the rapidity with which the system responds to the perturbation and reaches a new dynamic equilibrium. Equally important is the reversibility of a deleterious change. If we pollute a lake, we may discover that it might take a hundred years to reverse. One of the more far-reaching perturbations is that of climatic change. Weathering, and thus the chemical composition of the waters, is highly responsive to temperature and rainfall changes, chiefly because of the changes in the biota and their biochemical activity. Perhaps the more important effect is on the quantity and distribution of surface waters and their consequent influence on subsurface-water distribution. Desertification affects regional groundwater supplies in addition to dessication at the surface. At the same time, because of the change in dilution factors, the chemical composition changes.

We must distinguish between the different domains of analysis: local, regional, and global. The global picture is beginning to emerge reasonably well, partly because the numbers are so large and so few major stream networks dominate the land surface. The 10 or 15 major river systems of the world so overwhelm the discharge of the thousands of smaller rivers that we can get good estimates by studying only a few. It is unlikely that minor adjustments in the inventory of rivers over the globe will have much effect when dealing with numbers of the order of 10^{15} g.

When examining regional and local domains, it is difficult to neatly delimit the system and analyze it in terms of anthropogenic loads. The chemical mass fluxes of given constituents, which can be determined by combining the time series of their concentration with the associated water flows, are the result of many processes operative in the catchment area. Combination of such data with additional information about the geology, rainwater composition, land use, and the density of population and domestic animals may allow model building of the chemical dynamics of a watershed and the estimation of pollutional loads (Rickert *et al.*, 1976). Usually a significant anthropogenic load is best correlated with population density, whereas a significant natural input is often best correlated with water flow (Figure 5.5). A closer analysis must consider that the water carried in streams is made up of a fraction consisting of subsurface water and groundwater that reenters the surface water as well as a fraction arising from surface runoff, which enters the drainage system during, and soon after, precipitation.

Trace Elements

Though we know the broad outlines of the geochemical cycles of the major elements we are less well off with respect to many trace elements and compounds, both organic and inorganic. Even with an element seemingly as well known as lead, a detailed study, such as that of the Missouri lead belt (Jennett *et al.*, 1977), is needed to assess all of the sources and sinks in a relatively small region. There is still much to be learned about the geochemical cycling of mercury and other metals in which organic complexes such as methyl mercury may form as a result of microbiological activity. Our knowl-

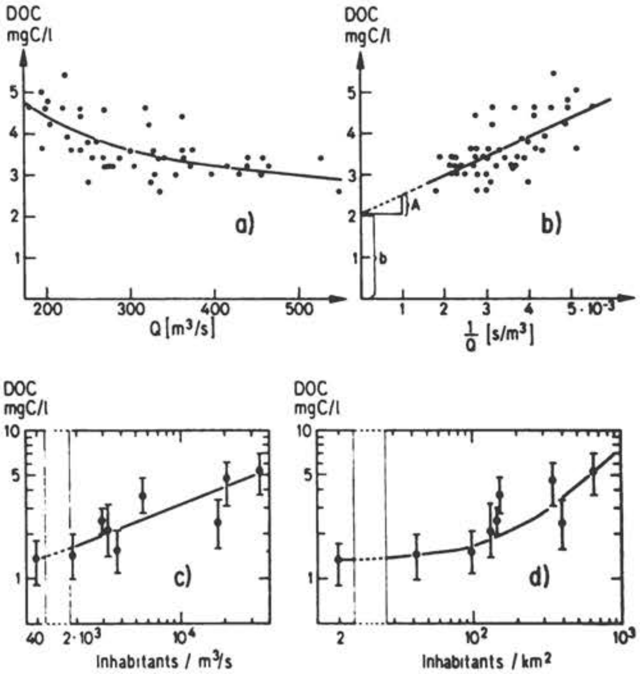


FIGURE 5.5 Dependence of dissolved organic carbon (DOC) on rate of flow, Q (a, b) and on population density (c, d). In (a) and (b), data from the river Aara (Switzerland) are plotted. In (c) and (d), DOC data with mean range between 5 and 95 percent values (i.e., values that are observed in more than 5 and less than 95 percent of all samples) of various rivers are given.

edge of the detailed geochemistry of artificial radionuclides like plutonium or neptunium is small.

Our poor knowledge of the trace elements, some of which may be highly toxic, starts with the paucity of careful regional studies of water compositions with respect to source-rock weathering. Though we know how to approach such studies, careful analysis of the trace-element composition of both source materials and waters is not a trivial matter. Though some work has been done on inorganic trace substances, work on the natural and anthropogenic distribution of organic compounds is still in its infancy. A good evaluation of the role of the biota of rivers and lakes with respect to a large group of inorganics and organics is lacking. We still know too little about the effects of sediment-dwelling worms and other organisms that mine the sediment for their food and may reintroduce some substances back into the water. The microbiological role in water-rock interaction and in precipitation and adsorption of inorganics or organics is likewise poorly known.

This information is needed if we are to specify the results of passage of pollutants into the hydrologic system. A prime example is the current ferment about the possible consequences of invasion of subsurface waters into a subsurface repository of radioactive wastes. To predict the course of events adequately, we need to know the geochemical pathways for a group of elements such as Sr, Cs, Ba, I, Pu, Ra, Am, Cu, and Np. Of these only the first four are moderately well known. A host of other materials, in particular organic compounds resistant to degradation, such as kepone, various plasticizers, surfactants, and others, need study.

The Problem of Prediction

The fundamental bind of predicting the consequences of pollution by any of thousands of conceivable substances is the impossibility of studying the detailed geochemical behavior of all of them. We cannot even think of testing every synthetic substance for possible deleterious effects. There are few unambiguous criteria for the concentrations in natural waters that produce toxic reactions or are mutagenic, teratogenic, or carcinogenic. Even the potential harm of low-level radiation is still being debated after many years of research on the subject. Even if we knew in advance just which compounds were likely to be harmful, we would face enormous difficulties in predicting, before any passage into the hydrologic network, exactly where they would end as precipitates or absorbed into a food or water chain. Though we knew of the potential for damage by polychlorbiphenyls (PCBs), no one predicted the course of transmission of accidentally introduced polybrombiphenyls (PBBs) in Michigan a few years ago. There are too many variables in complex geochemical cycles and too few substances have been mapped in detail to be able to generalize from a knowledge of general chemical characteristics.

From the effects we have observed we can issue a warning: To indiscriminately subject the system to components whose behavior is not known may be hazardous.

Analytical Data

Any study of the geochemical behavior of a polluting substance requires high quality analytical data. Much of the literature is stocked with inaccurate data, particularly for trace elements and many organic compounds, based on the poor methods of the past. The advent of modern analytical methods makes possible much more accurate and precise inventories of nanogram and even picogram quantities of many substances. But for many organics, the problems remain nonnegligible. What we have learned, in particular in the last few years from marine chemical studies, are the stringent requirements for extreme care to avoid contamination at every stage in the analysis, from sample collection to final analytical reagents and instruments.

SURFACE-WATER QUALITY

Surface waters that look blue and clear may nevertheless contain materials that are dangerous to our health. Other waters may look disagreeable and have poor taste but be perfectly safe to drink. Much of the enormously successful classical effort in preserving water quality was devoted to keeping bacterial counts down so that infectious disease transmission was stopped. The task remaining now is the prevention of chemical loading by toxic materials of a non-infectious nature. Here we do not ignore the many countries, particularly in the tropics, where parasites may be carried by surface waters and pose a severe health problem, such as bilharzia in parts of Africa. And microbiological problems remain in abundance. Yet some of the most serious problems of the future are the chemical by-products of modern high-technology living and the production of goods and services. Without surveillance, many of these chemicals enter surface-water supplies in kinds and quantities that are dangerous to our health.

Understanding the Natural System

To monitor the chemistry of a river or lake requires, first of all, a quantitative knowledge of the hydrologic regime of the area involved. The water budget, including rainfall, runoff, evapotranspiration, and infiltration, must be well known before the origin of the chemical composition of a stream can be deduced. This means the necessity for adequate measurements of discharge, rainfall, soil moisture, and the other measurements required to determine the hydrologic budget.

We have established some broad patterns of chemical weathering in relation to stream compositions, yet there is much to be done to determine the specific effects of climate, topography, rock type, soil, and aquatic biology. We are well advanced in our knowledge of rates of weathering of the common rocks exposed in a watershed as derived from the chemical analyses of stream waters and their dissolved and suspended loads. An example of this is the work done by Garrels and Mackenzie (1967) based on hydrologic and chemical analytical work by Feth *et al.* (1964) on the springs and ephemeral waters of the Sierra Nevada Mountains of California. Here the changes in compositions of groundwaters and surface waters were traced back to the reaction of specific minerals in the granitic rocks of the Sierras. This and similar investigations have led to extensive experimentation with modeling of various kinds of natural waters, using a wide range of approaches, as exemplified by the recent American Chemical Society monograph, *Chemical Modeling in Aqueous Systems* (Jenne, 1979). We are now able to do fairly well in computerized chemical modeling of natural waters and can reconstruct the probable nature of the solid phases that reacted with water and gases to form the solutions of the composition analyzed (Nordstrom *et al.*, 1979). These approaches depend on knowledge of the fundamental thermodynamic quantities and on accurate chemical analyses of the waters.

We are beginning to understand the rates and mechanisms of water-rock interactions from laboratory studies of the dissolution of silicates (e.g., Luce *et al.*, 1972; Lagache, 1965; Grandstaff, 1977; Siever and Woodford, 1979). Even with this promising beginning the difficulties of studying heterogeneous chemical reactions with nonstoichiometric substances have prevented us from reaching adequate solutions to what would appear to be the simpler problems, such as the dissolution of pure feldspar (Berner and Holdren, 1979). Carbonates have been studied intensively, partly because of their importance for understanding seawater chemistry, particularly in relation to the CO_2 cycle (Plummer and Wigley, 1976; Berner, 1975). Yet we are far from the ideal of taking a battery of rate constants and calculating the chemical compositions of rivers knowing water-discharge rates and the kinds of rocks that are exposed in the watershed.

One major difficulty with the data that we do have is that few comparisons of water compositions and discharge rates at given monitoring stations are available for many elements and compounds. Systematic changes along water courses as functions of discharge are likewise poorly known. A good example is a study of the Mekong River in southeast Asia, in which the chemical composition was followed in relation to floods and low-water periods (Figure 5.6; Carbonnel and Meybeck, 1975). Although we know that concentrations of most elements are reduced by dilution during high-water stages, the details are not well understood. It may become particularly important to determine the chemical response of the hydrologic regime to extremes of drought and wet seasons, for chemical pollutants will be affected in their own ways, as are the inorganic constituents of common rocks.

Anthropogenic Sources of Stream Loads

The gross load of surface waters with pollutants often can be estimated by considering the density of population, livestock, industrial activities, the area of the watershed, land use (forests, grassland, cropland, urbanization), and the type of waste treatment involved. Municipal wastes carry approximately the following loads per inhabitant per day: 40-50 g of organic carbon, 10-15 g of nitrogen (organic nitrogen compounds and ammonia), and 3-4 g of phosphorus (of which more than half is due to phosphate detergents).

For a stream with a rate of flow Q, the chemical mass flux L is often composed of a Q-independent mass flux A (e.g., from relatively constant waste discharges) and a Q-dependent part, bQ (e.g., from background fluxes including lakes and erosion as a function of equilibration between water and rocks, and elution of soil):

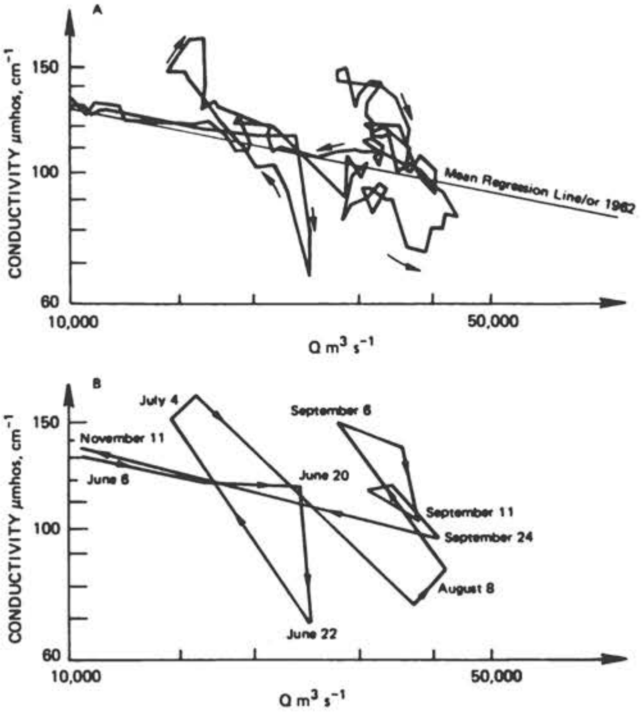


FIGURE 5.6 Variations of chemical composition with discharge of the Mekong River. (A) Daily variation of total ionic content with discharge during the 1962 high-water period. (B) Schematic cyclic variations related to peak discharges caused by Annamitic rivers in 1962 (from Carbonnel and Meybeck, 1975, with permission of the *Journal of Hydrology*, Elsevier Scientific Publishing Co.).

$$L = CQ = A + bQ$$

This equation cannot be applied uncritically; its validity, among other things, depends on the time period over which Q is averaged and the possible temperature dependency of the component being considered.

The scale of environmental changes taking place in industrial society has an especially pronounced impact on many lake systems. One of the most important problems is the progressive enrichment of waters with nutrients (phosphate and nitrogen compounds) concommittant with mass production of algae, increased productivity and other undesirable changes. Such a deterioration of lakes is referred to as eutrophication (NRC Planning Committee for the International Symposium on Eutrophication, 1969). As has been demonstrated in whole-lake experiments (Schindler, 1974, 1977), phosphate is the limiting nutrient that determines the productivity of biomass.

Conceptual models are available on the relationship between phosphate loading and the degree of eutrophication. It is thus possible to estimate the allowable phosphate loading as a function of mean lake depth and hydraulic residence time (Figure 5.7).

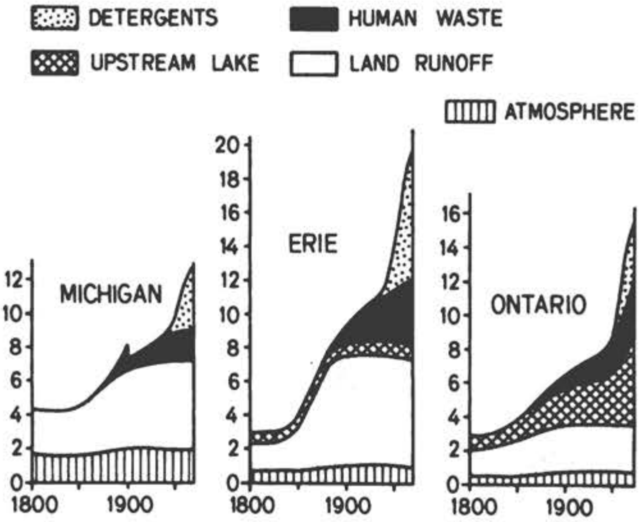


FIGURE 5.7 Historical loadings of total phosphorus to some Great Lakes in thousands of tons per year, as calculated by model (from Chapra, 1977, with permission of the American Society of Civil Engineers).

It is difficult to identify the compounds that are potentially hazardous to man and to ecology, although analytical chemistry has made remarkable progress in improving detection down to extremely small concentrations (10^{-3} g/l) (Figure 5.8). The harmfulness of a substance depends on its specific structure and on the concentration of the substance under consideration, but other factors such as temperature, turbulence, and the presence of other substances are also important.

In evaluating toxicity, we need to distinguish between (1) substances that impair the health of humans and animals or poison aquatic organisms and (2) substances that primarily affect the organization and structure of aquatic

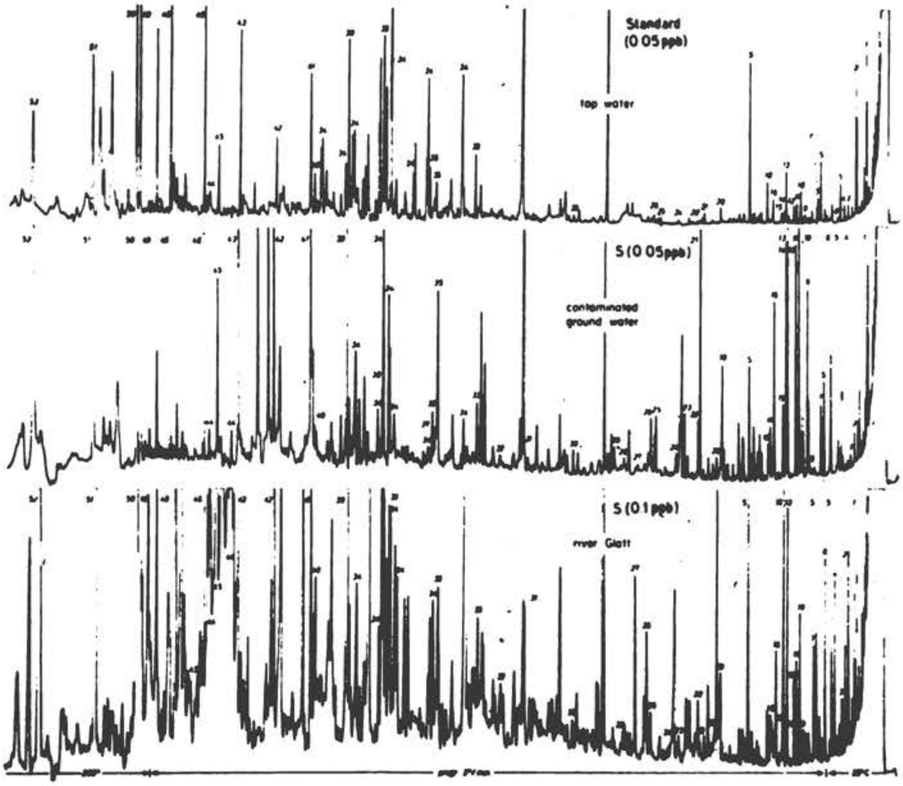


FIGURE 5.8 Each peak represents (at least) one compound. Peaks with a number represent substances identified with mass spectrometry. Some of the more refractory pollutants, especially some hydrocarbons and chlorinated hydrocarbons occur also in groundwater and drinking water (from K. Grob *et al.*, 1975).

ecosystems. In this interaction, contaminants may impair the self-regulatory functions of the system or interfere with food chains.

Of particular concern are human-health hazards from long-term effects by trace concentrations of pollutants in drinking water (Borneff, 1978). Epidemiological studies have shown significant correlations between cancer morbidity and origins of potable-water resources (Cantor and McCabe, 1978). (Although such a correlation does not establish a cause-effect relationship, it appears difficult to explain the correlations by other factors.) Furthermore, substances contained in drinking water (concentrates from inverse osmosis or elutes from adsorbent columns) have been shown to produce mutations in bacteria (Cooper *et al.*, 1978).

Since many carcinogenic substances are also mutagenic (though not necesarily), tests for mutagenic effects in bateria (e.g., Ames tests) are often used as screening tests to recognize potential carcinogens (Ames *et al.*, 1975).

The ultimate objective is to be able to develop sufficient information about a given compound to allow prediction of its environmental behavior and toxicological effects. Dozens of standard tests have been developed to predict such effects. If one considers however that such short- and long-time tests must be carried out for each substance (and its degradation intermediates) with different organisms, that synergistic and antagonistic effects with other substances present may occur, and that each negative test result does not necessarily mean "zero toxicity," one realizes that the predictive value of such standard tests is limited. New chemicals may be produced and may enter the environment faster than they can be tested satisfactorily.

Certain aspects of the pollution potential of individual compounds may be assessed on the basis of a few physical-chemical parameters that--in addition to biodegradability--permit forecasting of the manner in which the compounds are distributed in the environment and thus in turn their fate and approximate residence time as well as their ecological impact (Hutzinger *et al.*, 1977). Information on the biodegradability of a pollutant is of particular importance in estimating its residence time in water. The latter is also affected by the tendency of a substance to escape into the atmosphere, as characterized by its vapor pressure, the distribution equilibrium between water and atmosphere, and the gas-transfer coefficient.

The lipophility--as measured by the *n*-octanol-water distribution coefficient--is a good measure of the tendency of a substance to be accumulated in the food chain and to become biomagnified (Figure 5.9).

The best way to safeguard man and the aquatic ecosystems against unanticipated effects of organic pollutants is to require that all such substances that are mass-produced and have a chance of becoming dispersed in the environment, especially substances used in households and in agriculture (e.g., detergents, cleaning fluids, pesticides), be fully biodegradable.

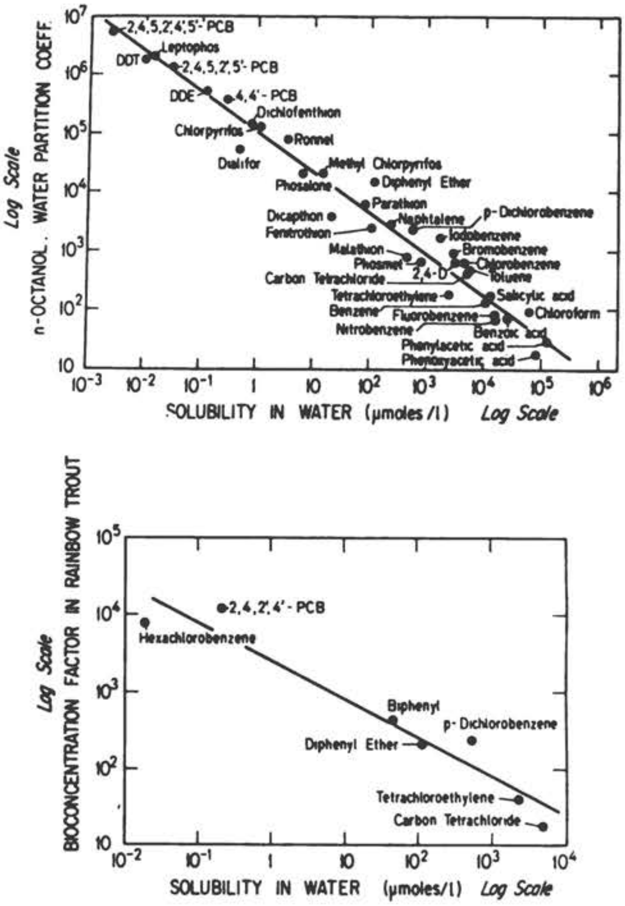


FIGURE 5.9 Both the *n*-octanol-water partition coefficient and the bioconcentration factor in rainbow trout are inversely proportional on aqueous solubility (from Chiou *et al.*, 1977, reprinted with permission from *Environmental Science and Technology*, copyright 1977, American Chemical Society).

Case Histories

Hubbard Brook

About 20 yr ago an ecosystem model of a forested watershed and the conception of a method ("the small-watershed technique") whereby some major parameters could be measured under field conditions were developed by Likens *et al.*, (1977). The method enabled the measurement of input and output of chemicals and the construction of mass balances, especially on nutrients and their cycling. The study shows how the ecosystem moderates and changes inputs and how it affects biogeochemical cycles by its outputs. Meteorological input, strongly influenced by advertent and inadvertent manipulation by man (acidity traced to sulfur pollution and to nitrate resulting from internal combustion engines), is an important source of chemicals in this watershed. Stream-water chemistry of the undisturbed forest ecosystem is highly predictable and reflects environmental conditions (geologic, meteorological, as well as biological features).

Rhine River

Over the last 100 yr the composition of the Rhine River, the catchment area of which has a high population and industrial density (140 inhabitants/km^2; 15,000 inhabitants per m^3/sec flow rate; and a gross national product per unit water flow about three times as high as that for the Ohio River), has changed dramatically. Today in the Rhine River the mass flux of refractory substances (i.e., those--frequently of synthetic origin--that resist biodegradation) exceeds that of biodegradable (mostly excreta and other biogenic wastes) substances (Figure 5.10). Conventional biological waste treatment for municipalities and industries within the catchment area is unable to prevent the progressive accumulation of refractory substances, as biological waste treatment and self-purification are not very effective in eliminating these relatively persistent chemicals.

An attempt has been made by Zobrist and Stumm (in press) for the Rhine catchment area to assess the influence of rock composition and anthropogenic sources on the water chemistry and to extrapolate these findings to estimate the respective dissolved inputs into the ocean from natural and anthropogenic sources. The following data sources were weighed to estimate the composition of a "pristine" Rhine: The mineralogical rock composition, the meteorological inputs, information on relatively unpolluted parts of the alpine Rhine, the interpretation of the concentration-flow relationship, and analytical determinations carried out 125 yr ago. The postulated "pristine" concentrations, in comparison to the mean actual concentrations of today, are given in Figure 5.11. According to these estimates, only the concentrations of bicarbonate and dissolved silica have remained constant. The loads in sodium chloride, in calcium chloride, and in heavy metals have increased more than tenfold, that of calcium sulfate about fourfold.

Swiss Rivers, Seasonal Effects

Both the use of graphical methods, as for example in the extrapolation of the load as a function of *Q*, and of statistical methods have allowed for a rough approximation of the loads arising from the anthropogenic and background sources. Data on the major Swiss rivers, collected as weekly composite samples for several years, have shown that a number of components vary markedly in their behavior throughout the year. Several factors may contribute to these observed fluctuations, among them changes in the amount of water and its sources (rainfall-runoff versus melting of snow and glaciers), fluctuations in contribution of subsurface water to the river, and temperature. Because temperature

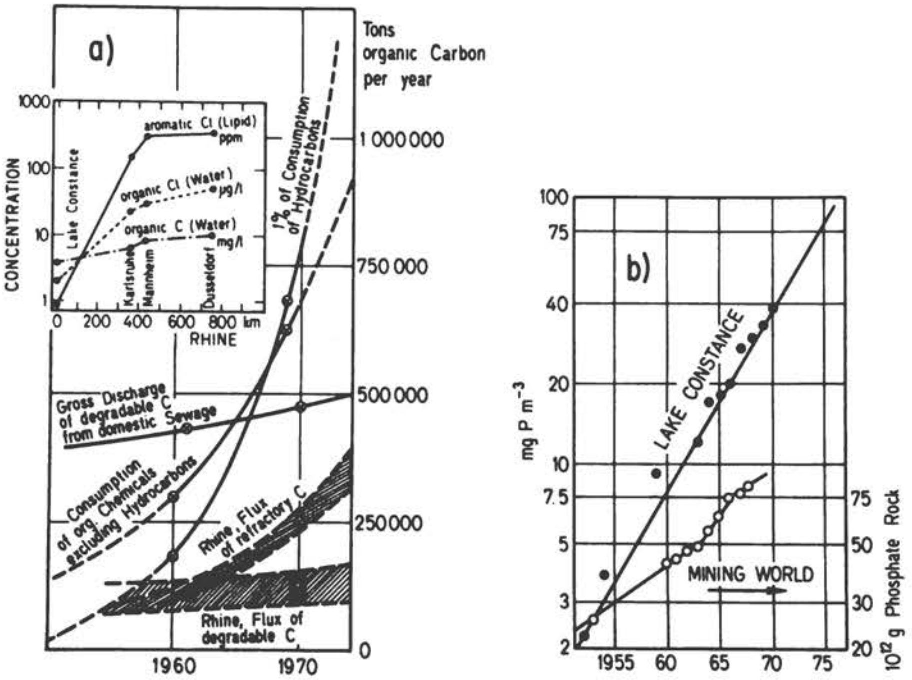


FIGURE 5.10 Change in pollutional load of (a) Rhine River (organic carbon) and (b) Lake Constance (phosphate). a, Increase in consumption of organic chemicals in the catchment area of the Rhine River in comparison to the gross load of the Rhine with degradable organic carbon (excreta). The flux of refractory organic carbon C (calculated and measured for Dusseldorf) reflects the increase in consumption of synthetic chemicals. The flux of biodegradable C remains constant, however (data from Stumm and Roberts, Swiss Federal Institute of Technology). The insert illustrates the increased downstream accumulation of chlorinated organic compounds in the water and in the biota (from data of H. Sontheimer). Reproduced from International Association of Water Supplies of the Rhine River (1973). b, The increase in concentration of P in Lake Constance (measured during overturn) is compared with the quantity of phosphate rock mined globally.

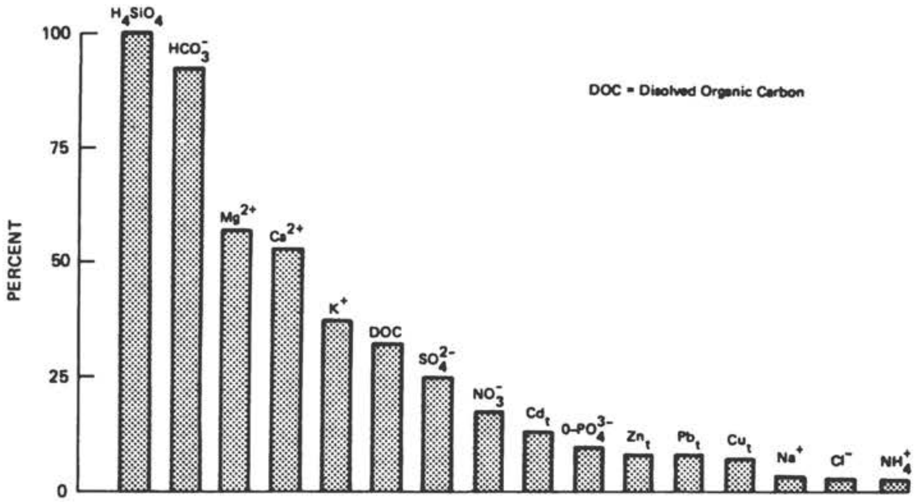


FIGURE 5.11 A comparison of the estimated concentrations in the "pristine" lower Rhine River (percent) with the mean measured values of today (from Zobrist and Stumm, RIOS Workshop, in press).

does not change in the same way as water flow, different situations occur with regard to sources of water, chemical inputs, and temperature-dependent (biological) reactions at different times, for example, in spring and fall. In the case of phosphate (Figure 5.12), for example, the variations in the phosphate load of the rivers may be explained by (1) annual changes occurring in lakes located upstreams (due to photosynthesis and respiration) and (2) drainage from farm land where manure and fertilizers are applied primarily in spring. Correcting for the modulations in the lake allows for better estimates of the phosphate load of the river.

SUBSURFACE-WATER QUALITY

The quality of water drawn from wells is of vital concern to the millions of people who depend on these supplies for their drinking water, for irrigation, and for industry. Because subsurface waters are derived from surface waters at some time in the past, their chemical compositions carry the stamp of their origin at the surface as well as the soils and rocks through which they have passed underground. In analyzing a particular region we need to know the surface hydrology: which streams are effluent and--together with springs--drain the ground-

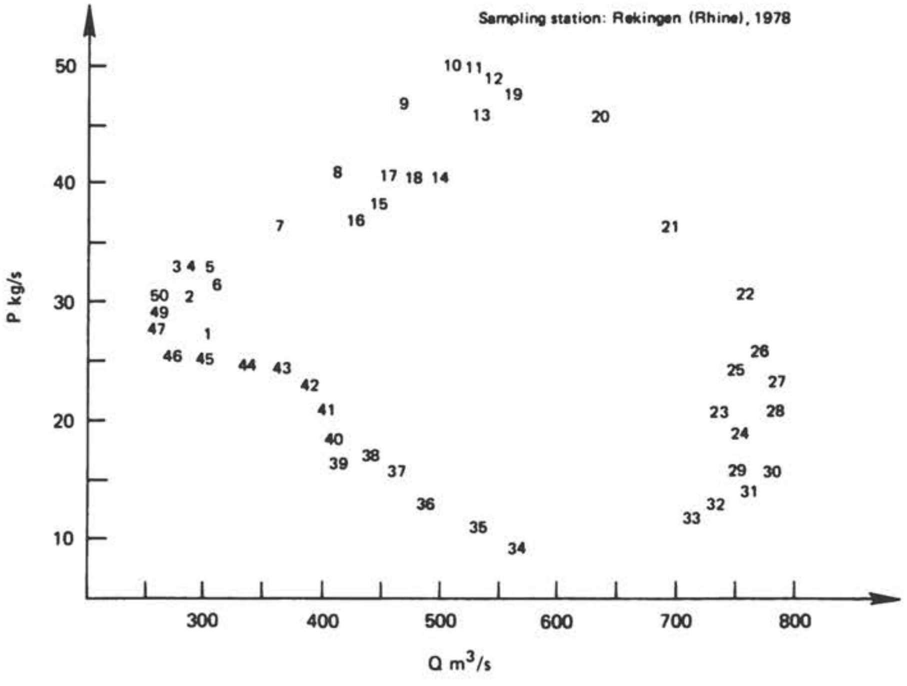


FIGURE 5.12 Phosphate load as a function of water flow. Weekly composite samples. (The points are numbered according to the number of weeks in the year.) Low loads are observed during summer because of removal by algae in upstream lake. The high values in the spring are caused by drainage from farm land. The values at the lowest temperatures (left side of the curve) are representative of the basic river load (from J. Davis, Swiss Federal Institute of Technology, unpublished).

water and which are influent and seep into the ground. We have to know the partition between soil infiltration and runoff. But knowing the partition between surface and subsurface just prepares the way for understanding the chemistry.

Perhaps the most important environment that influences the compositions of groundwaters is the upper part of the vadose zone (unsaturated zone) (see Chapter 3). Soil-water reactions are the primary ones in weathering, in the production of soil-organic compounds and soluble organics that infiltrate downward to the saturated zone. These reactions continually change the content of dissolved gases and concentrations of dissolved solid components as rainwater percolates downward. Many of these reactions are mediated by bacteria and higher soil organisms.

In irrigation regimes we see artificial drainage networks that operate to distribute water from the subsurface and put it back on the top of the soil, there to evaporate and re-infiltrate. The increase in salinity of irrigated soils is a problem that, if not faced, may result in the spoiling of agricultural lands. Yet we have paid too little attention to the fact that, if the salinity increases, the concentrations of many trace substances must be increasing sympathetically with the abundant elements. If there are toxic materials in the irrigation waters, they also will be concentrated and may find their way into the food products being grown. Conventional treatment of salinized soils by flushing with freshwater may not remove some of these substances, which may be sorbed irreversibly onto soil clays but which may, by biological reactions, be imbibed by plant roots. We know little of how these reactions work.

Of particular concern in soil-water systems is the influx of materials from septic tanks in the huge areas of sprawling suburban development surrounding many of our large cities. In some ways they act like giant dispersed primary sewage-treatment systems, cleaning up the major biological wastes but leaving relatively untouched most of the chemical components. If these infiltrate down to aquifers, the region's water resources are endangered. In our aging cities, sewer systems are getting older and leakier. Though major main breaks are repaired, we know from inspection of many breaks that they must have been leaking extensively for extensive periods before the final break. Whereas freshwater supply mains are monitored for pressure, the sanitary drains may not be checked and leakage goes undetected. To the extent that these sewer systems are connected to aquifers supplying freshwater, we may attribute some chemical contaminants to that source. Many parts of the country use sanitary landfill and dumps for waste disposal. In these there may be extensive leakage and infiltration to aquifers below.

In all the reactions of sanitary wastes with infiltrating groundwater we must watch for the trace elements, both organic and inorganic. For example, there are a variety of toxic materials used in newsprint that may become solubilized and leached into groundwater. We have already recognized this threat by forbidding the use of recycled newsprint for paper products for food wrapping. The trace elements in detergents may show up in groundwaters if infiltration routes carry septic-tank effluent to aquifers. The aggregate of all the average households may be a significant source of toxic materials. We simply do not have enough accurate information to tell.

Over many areas of the country, particularly in the high plains, we have been "mining" water for decades (see Chapter 4). As we continue to deplete groundwater supplies in extensive aquifers faster than they are being recharged, we may ask whether there is any effect on the water quality. As artificial recharge experiments are tried, we must watch the quality of the surface water that is used for infiltration. Here rates of groundwater movement are pertinent, for a degraded aquifer can be improved but it will regrade with a time scale similar to that of the original transformation. This is one reason why we have to know the rates of chemical influxes as well as groundwater movement: Some reservoirs that take a long time to become changed for the worse may take equally long to be reversed for the better. The ways by which to study these problems are to determine the steady-state equilibria with respect to flow rates and rates of chemical reaction with the waters. Flow rates have been successfully determined by use of tritium and ^{14}C. As noted above, reaction rate constants have been determined for a number of minerals; from these data we can reconstruct the progressive changes of groundwater as it moves along an aquifer. As we show below, this kind of problem will loom ever more importantly as we bury greater amounts and different kinds of solid and liquid waste at depths where they may interact with aquifers.

Anthropogenic Sources of Subsurface-Water Loads

One of the classic sources of subsurface-water contamination is mining. Typical of underground coal mining is the production of large quantities of iron-rich and sulfateor sulfide-rich waters. Pyrite (FeS_2)--common in coal beds--weathers, oxidizes, and dissolves in the mine or in mine wastes as they are piled in dumps above ground. Below ground they may react more slowly, but they affect the groundwaters adjacent to the mine. We must look to increasing problems in the western United States as a result of the greater exploitation of coal--both from strip and underground mines--in hydrologic regimes where flow rates are slow. Western coals are generally low in pyrite but there are large enough concentrations to lead to significant sources of potential pollution. Mines for other materials, such as metallic ores, may contribute a variety of toxic metals in the same way. Many of the metallic ores are sulfides, which oxidize rapidly in the mine environment. In many of these the trace-element content of the vein being mined may be extraordinarily high--that, in fact, may be why they are being mined--and such metals, as soluble species, may find their way into groundwaters.

Subsurface waste disposal in old mines, in specially constructed caverns, and in permeable aquifers raises new problems. Increasingly, as we become aware of the dangers of waste incineration and of surface disposal, we turn to subsurface burial. Though there has been much attention paid to the radioactive-waste disposal problem in the past few years, the hazards of some highly toxic nonradioactive wastes may be as great. The problem is that it is difficult to find the perfectly sealed underground vault with no possibility of leakage. And, though radioactive wastes eventually decay so that radioactivity is no longer a problem, some chemical wastes may be extraordinarily stable in subsurface environments and can be expected to last for geologic times. Though we can be optimistic about finding good sites in low-permeability rocks, we must evaluate the consequences of invasion by groundwaters and transport to aquifers. The chemical hydrologist needs to be able to predict the results if there is leakage.

Other wastes introduced to groundwater may cause serious problems. As our gasoline service stations age, their buried tanks start to leak and infiltrate the ground. Though the volumes may be small, they may be locally important. Liquid-waste disposal in buried permeable formations is becoming more common. When done without taking care to watch the pressure, the overburden may be lifted and small earthquakes may result, as was found at the Rocky Mountain Arsenal near Denver. The danger here is more than that of small earthquakes; it is that hydraulic fracturing from lifting overburden may open new subsurface pathways so that groundwater-migration patterns may change, leaving open the possibility of migration to potable water aquifers.

Most of the waste-disposal programs that have been responsible have been put well below the interface of near-surface potable water aquifers and the deeper formations that contain waters of varying salinity but generally too high to be potable. But almost daily we hear horror stories of waste materials being stored near the surface, in shallow ditches or leaky tanks, well connected to surface potable waters. What is needed is much more sophistication about routes of migration in the top few hundred meters.

STRATEGIES FOR THE FUTURE

How are we to plan for the future? How can we keep up with the steadily increasing pressures on the quality of water in our surface and subsurface reservoirs? We are dependent on fundamental knowledge of water-rock and water-biological reactions that can come only from careful and detailed studies of specific reactions in specific regions. Further, we are dependent on computer modeling of the larger chemical systems. The models being developed for estimating mineral equilibria from multicomponent solutions are a guide. Yet we should not hold out the goal of understanding in detail every watershed in the world with respect to every possible pollutant. We need to develop a workable and economic approach toward following quantitatively the distribution of a toxic material in the hydrologic regime, once it is known that a clearly deleterious substance has invaded the system. Such approaches will take the form, we suspect, of working backward from high-quality chemical analyses of particular strategically placed monitoring points with a view to discriminating between the natural--the pristine--contribution and the anthropogenic load. Once the load is defined, the hunt begins for the sources and sinks. We must not only learn how to identify the sources as to location and kind but to discover the sinks as well. For if we want to intervene intelligently, we may want to work to enhance the efficiency of sinks so that we can engineer the removal of undesirable substances by taking advantage of the natural cycling of all compounds on the surface of the earth.

REFERENCES

Ames, B. N., J. McCann, and E. Yamasaki (1975). Methods for detecting carcinogens and mutagens with the Salmonella/Mammalian-microsome mutagenicity test, *Mutat. Res. 31*, 347 pp.

Berner, R. A. (1975). The role of magnesium in the crystal growth of calcite and aragonite from sea water, *Geochim. Cosmochim. Acta 39*, 489-504.

Berner, R. A., and G. R. Holdren, Jr. (1979). Mechanism of feldspar weathering I. Experimental studies; II. Observations of feldspars from soils, *Geochim. Cosmochim. Acta 43*, 1161-1186.

Blumer, M., and J. Sass (1972). Oil pollution: Persistance and degradation of spilled fuel oil, *Science 167*, 1120-1122.

Borneff, J. (1978). Occurrence of carcinogens in surface water and drinking water, in *Aquatic Pollutants*, O. Hutzinger *et al.*, eds., Pergammon Press, Oxford.

Cantor, K. B., and L. J. McCabe (1978). The epidemiological approach to the evaluation of organics in drinking water? in *Water Chlorination, Environmental Impacts and Health Effects*, R. L. Jolley *et al.*, eds., Ann Arbor-Humphrey Science Publishers Inc., Ann Arbor, Mich., pp. 379-394.

Carbonnel, J. P., and M. Meybeck (1975). Quality variations of the Mekong River at Pnomh Penh, Cambodia, and chemical transport in the Mekong Basin, *J. Hydrol. 27*, 249-265.

Ceasar, J., H. Collier, J. Edmond, F. Frey, G. Matisoff, A. Ng, and R. Stallard (1976). Chemical dynamics of a polluted watershed, the watershed, the Merrimack River in northern New England, *Environ. Sci. Technol. 10*, 697-704.

Chapra, St. C. (1977). *J. Environ. Eng. Div. 103, Am. Soc. Civil Engineers*, 150.

Chiou, C. T., V. H. Freed, D. W. Schmedding, and R. L. Kohnert (1977). *Environ. Sci. Technol. 11*, 475.

Cooper, J. C., D. P. Lang, and C. C. Smith (1978). Mutagenicity of complex mixtures from drinking water, in *Water Chlorination, Environmental Impact and Health Effects*, R. L. Jolley *et al.*, eds., Ann Arbor-Humphrey Science Publishers, Inc., Ann Arbor, Mich.

Dagley, S. (1975). Microbial degradation of organics compounds in the biosphere, *Am. Sci. 63*, 681-689.

Feth, J. H., C. E. Roberson, and W. L. Polzer (1964). Sources of mineral constituents in water from granitic rocks, Sierra Nevada, California and Nevada, *U.S. Geol. Surv. Water-Supply Pap. 1535-I*, 70 pp.

Garrels, R. M., and C. L. Christ (1965). *Solutions, Minerals, and Equilibria*, Harper & Row, New York, 450 pp.

Garrels, R. M., and F. T. Mackenzie (1967). Origin of the chemical compositions of some springs and lakes, in *Equilibrium Concepts in Natural Water Systems*, W. Stumm, ed., Advan. Chem. Ser., pp. 222-242.

Garrels, R. M., and F. T. Mackenzie (1971). *Evolution of Sedimentary Rocks*, W. W. Norton & Company, Inc., New York, 397 pp.

Grandstaff, D. E. (1977). Some kinetics of bronzite orthopyroxene dissolution, *Geochem. Cosmochim. Acta 41*, 1097-1104.

Grob, K., et *al.* (1975). *J. Chromatogr. 106*, 275.

Hutzinger, O., I. H. van Lelyveld, and B. C. J. Zoeteman, eds. (1978). *Aquatic Pollutants: Transformation and Biological Effects*, Pergamon Press, Oxford, 519 pp.

Hutzinger, O., M. Th. M. Tulp, and V. Zitko (1977). Chemicals with pollution potential, in *Aquatic Pollutants: Transformation and Biological Effects*, O. Hutzinger, I. H. van Lelyveld, and B. C. J. Zoeteman, eds., Pergamon Press, Oxford.

International Association Water Supplies of the Rhine River (1973). *IAWR 3*, 69.

Jenne, E. A., ed. (1979). *Chemical Modeling in Aqueous Systems: Speciation, Sorption, Solubility, and Kinetics*, American Chemical Society Symposium Ser. 93.

Jennett, J. C., B. G. Wilson, I. H. Lowsley, K. Purushothaman, E. Bolter, D. D. Hemphill, N. L. Gale, and W. H. Trauter (1977). Transport and distribution from mining, milling, and smelting operations in a forest ecosystem, in *Lead in the Environment*, NSF/RS-770214, U.S. Supt. Doc., pp 135-178.

Lagache, M. (1965). Contribution a l'etude de l'alteration des feldspaths dans l'eau, entre 100 et 200°C., sous diverses pressions de CO_2 et application a la synthese des mineraus agrileaus, *Bull. Soc. Fr. Min. Crystallogr. 88*, 223-253.

Lantzy, R. J., and F. T. Mackenzie (1979). Atmospheric trace metals; Global cycles and assessments of man's impact, *Geochim. Cosmochim. Acta 43*, 511-525.

Likens, G. E., F. H. Borman, R. S. Pierce, J. S. Eaton, and N. M. Johnson (1977). *Biogeochemistry of a Forested Ecosystem*, Springer-Verlag, New York, 156 pp.

Luce, R. L., R. A. Bartlett, and G. A. Parks (1972). Dissolution kinetics of magnesium silicates, *Geochim. Cosmochim. Acta 36*, 35-50.

NRC Planning Committee for the International Symposium on Eutrophication (1969). *Eutrophication: Causes, Consequences, Correctives*, National Academy of Sciences, Washington, D.C., 661 pp.

NRC U.S. National Committee for Geochemistry (1977). *Geochemistry and the Environment, Vol. 2*, National Academy of Sciences, Washington, D.C., 163 pp.

Nordstrom, D. K., L. N. Plummer, T. M. L. Wigley, T. J. Wolery, J. W. Ball, E. A. Jenne, R. L. Bassett, D. A. Crerar, T. M. Florence, B. Fritz, M. Hoffman, G. R. Holdren Jr., G. M. Lafon, S. V. Mattigod, R. E. McDuff, F. Morel, M. M. Reddy, C. Sposito, and J. Thrailkill (1979). A comparison of computerized chemical models for equilibrium calculations in aqueous systems, in *Chemical Modeling in Aqueous Systems*, E. A. Jenne, ed., American Chemical Society Symposium Ser. 93, pp. 857-892.

Plummer, L. N., and T. M. L. Wigley (1976). The dissolution of calcite in CO_2 saturated solutions at 25°C and 1 atmosphere total pressure, *Geochim. Cosmochim. Acta 40*, 191-202.

Rickert, D. A., W. G. Hines, and S. W. McKenzie (1976). Methodology for river-quality assessment with application to the Willamette River Basin, Oregon, *U.S. Geol. Surv. Circ. 715-M*.

Schindler, D. W. (1974). Eutrophication and recovery in environmental lakes; Implications for lake management, *Science 184*, 897-899.

Schindler, D. W. (1977). Evolution of phosphorus limitation in lakes, *Science 195*, 260-262.

Siever, R., and N. Woodford (1979). Dissolution kinetics and the weathering of mafic minerals, *Geochim. Cosmochim. Acta 43*, 717-724.

Stumm, W., and J. J. Morgan (1978). *Aquatic Chemistry, 2nd ed.*, Wiley-Interscience, New York.

Suffet, I. H., ed. (1977). *Fate of Pollutants in the Air and Water Environments*, Wiley-Interscience, New York.

Zobrist, J., and W. Stumm (in press). Chemical dynamics of the Rhine catchment area in Switzerland; Extrapolation to the "pristine" Rhine River input into the ocean, RIOS-Workshop.

Predictive and Reactive Systems for Aquatic Ecosystem Quality Control

6

JOHN CAIRNS, JR.

Virginia Polytechnic Institute and State University
and
The University of Michigan Biological Station

TYPES OF IMPACTS ON AQUATIC ECOSYSTEMS

The sources of pollutants that impact freshwaters in the United States are thought to be nearly equally divided between point and nonpoint sources. The Federal Water Pollution Control Act Amendments of 1972 classify nonpoint sources in four categories: agricultural activities, forest-related activities, mining activities, and construction activities. Urban runoff is a nonpoint source that does not reach the aquatic ecosystem through a pipe. An Environ-Control, Inc. (1974) report on a study of the Delaware River estuary revealed that 40-80 percent of the total annual chemical oxygen demand and biological oxygen demand entering receiving waters from a city originated from sources other than treatment-plant discharge. The point sources primarily are toxicants (potentially hazardous chemicals), heated wastewaters, suspended solids, and degradable organics (Cairns, 1978). In addition, deleterious impacts to stream biota may result from velocity changes from impoundments or once-through cooling systems for steam-electric power plants (e.g., Kerr, 1953; Jensen, 1974) as well as channelization (e.g., Barton *et al.*, 1972). This chapter will not discuss these effects but rather a strategy for mitigating them.

It is now abundantly clear that serious damage may result when the intrusion of toxic chemicals, heated wastewater, degradable organics, and particular material exceed the ability of natural ecosystems to cope with them. This means managing the intrusion of these potentially harmful materials at an ecosystem or regional level so that their collective impact will not be deleterious. A more extensive discussion of this need is in Cairns (1975, 1978).

PREDICTIVE CONTROL

A hazard assessment designed to estimate risk to an aquatic ecosystem requires evidence to make a scientific judgment on (1) toxicity--the inherent property of the chemical that will produce harmful effects to an organism (or community) after exposure of a particular duration at a specific concentration and (2) environmental concentrations--those actual or predicted concentrations resulting from all point and nonpoint sources as modified by the biological, chemical, and physical processes acting on the chemical or its by-products in the environment (Cairns, 1978). The strategy for estimating the hazards on aquatic ecosystems from activities is similar, although the information required will differ. The range of concentrations causing no adverse biological effects may be called the assimilative capacity (Cairns, 1978).

The basic hazard-evaluation procedure is designed around two distinct types of information: (1) concentrations of the chemical (or other stressor) that do not produce adverse biological effects and (2) the environmental concentrations that will result from production and use of the chemical. Figure 6.1 depicts the hazard-evaluation process. In Figure 6.1a the "no adverse biological effect" level and the concentration that will result from introducing the chemical into the environment (e.g., Mill *et al.*, 1977) are well apart. In fact, only estimates of these concentrations are known and are indicated by dotted lines that envelope the solid concentration lines. This is because Tier I testing (see Table 6.1 for definitions and criteria for Tiers) consists of comparatively crude short-term dose-response assays with a graded series of static or nonrenewed concentrations of a test material using lethality as an endpoint. Tier II testing is more sophisticated and expensive (e.g., continuous flow instead of batch, longer time, and more subtle endpoints) and Tier III even more so. Frequently, Tier I testing will sufficiently improve the estimates so that one will, at decision point P, be able to determine that the concentrations are indeed different. Thus there may be justification for terminating testing in Tier I and concluding, at a certain risk level, that introduction of the chemical will not cause an environmental hazard. One should remember that risk cannot be reduced to zero.

In Figure 6.1b, the two concentrations are closer together. Now, testing must be carried through Tier III before the same statement can be made at a comparable risk level. Figure 6.1c depicts the case where testing up to the same point leads one to conclude that the chemical would cause an environmental hazard because the environmental concentration would be greater than the "no adverse biological effects" concentration.

TABLE 6.1 Tiers Are Groups of Toxicity Tests That Furnish Information Necessary to Make a Decision[a]

Tier 0
  Chemical-physical data
  Analytical technique

Tier I
  Acute toxicity screens
  Environmental transformation and degradation
  10-30 day feeding studies
  3 *in vitro* mutagenicity/carcinogenicity tests
  Analytical to 1-10 mg/liter

Tier II
  90-day subacute tests
  Teratogenic-mutagenic-carcinogenic tests
  Environmental fate
  Bioconcentration tests

Tier III
  Chronic tests
  Multigeneration mutagenicity
  Metabolic studies
  Terrestrial plants and animals

Criteria for Number of Tiers
  Production volume
  Use and exposure levels
  Nature of potential hazard
  Physical-chemical properties
  Structure-activity relations

All tested through Tiers 0 and I.

[a] Reproduced with permission of the American Fisheries Society (Maki, 1979).

REACTIVE CONTROL

Living organisms, unlike any other analytical tool available, will respond to every possible substance or mixture of substances at some level, no matter what their chemical or physical characteristics may be. Even water as pure as distilled water is toxic to aquatic organisms such as fish and aquatic insects, because some dissolved ions are lacking or others are too abundant. The universality of the detection capability of the living organism is unequaled among instruments devised by man (Smith *et al.*, 1974). There is no nonliving sensor that can detect toxicity, although some can detect toxic concentrations or conditions. Even these are inadequate--(1) some substances have adverse biological effects at concentrations below present analytical capability; (2) some substances interact synergistically with others so that the combined toxicity is greater than their individual additive toxicities; and (3) in aquatic ecosystems, water quality is a major determinant in the expression of toxicity. However, one must have chemical-physical data to properly evaluate the cause of the toxicity, so the two sets of data are as essential as they were in hazard evaluation.

The relationship between predictive control and reactive control is generally poorly understood. For example, many industrial and regulatory wastewater quality control personnel believe that only predictive controls (i.e., toxicity tests) are necessary to prevent environmental damage. As a consequence, publications such as *Water Quality Criteria of 1972* are used unaccompanied by error-control systems. If this much credence were given to unverified predictions in other aspects of industrial management, those companies would soon be in deep financial trouble.

The schematic of the relationship of predictive and error control is given in Figure 6.2. Thus, the discharge of an effluent causes some alteration (not necessarily adverse) in one or more environmental parameters. If the receiving system is being monitored properly, these changes should be detected and, if adverse, some corrective measures taken. For the predictive control, a change in a parameter causes a change in algorithm output. Properly handled, this should result in corrective action that will protect the receiving system. This brief discussion indicates that only when reactive and predictive systems are combined will effective environmental quality control be possible. Reactive control (monitoring or error control) is only effective when the predictive system has failed. One hopes that a prompt response will reduce damage, but it is unrealistic to assume that a monitoring system is adequate as a sole defense against harm. Most predictive systems, especially where toxic chemicals are concerned, are based on the response of a few species and a comparatively small number of individuals exposed under a range of conditions (almost certainly less than the total environmental repertoire). One is predicting system effects from a relatively small data base. It would be astonishing if these predictions were always correct--hence the need for monitoring.

A variety of methods suitable for reactive control exists. These generally can be divided

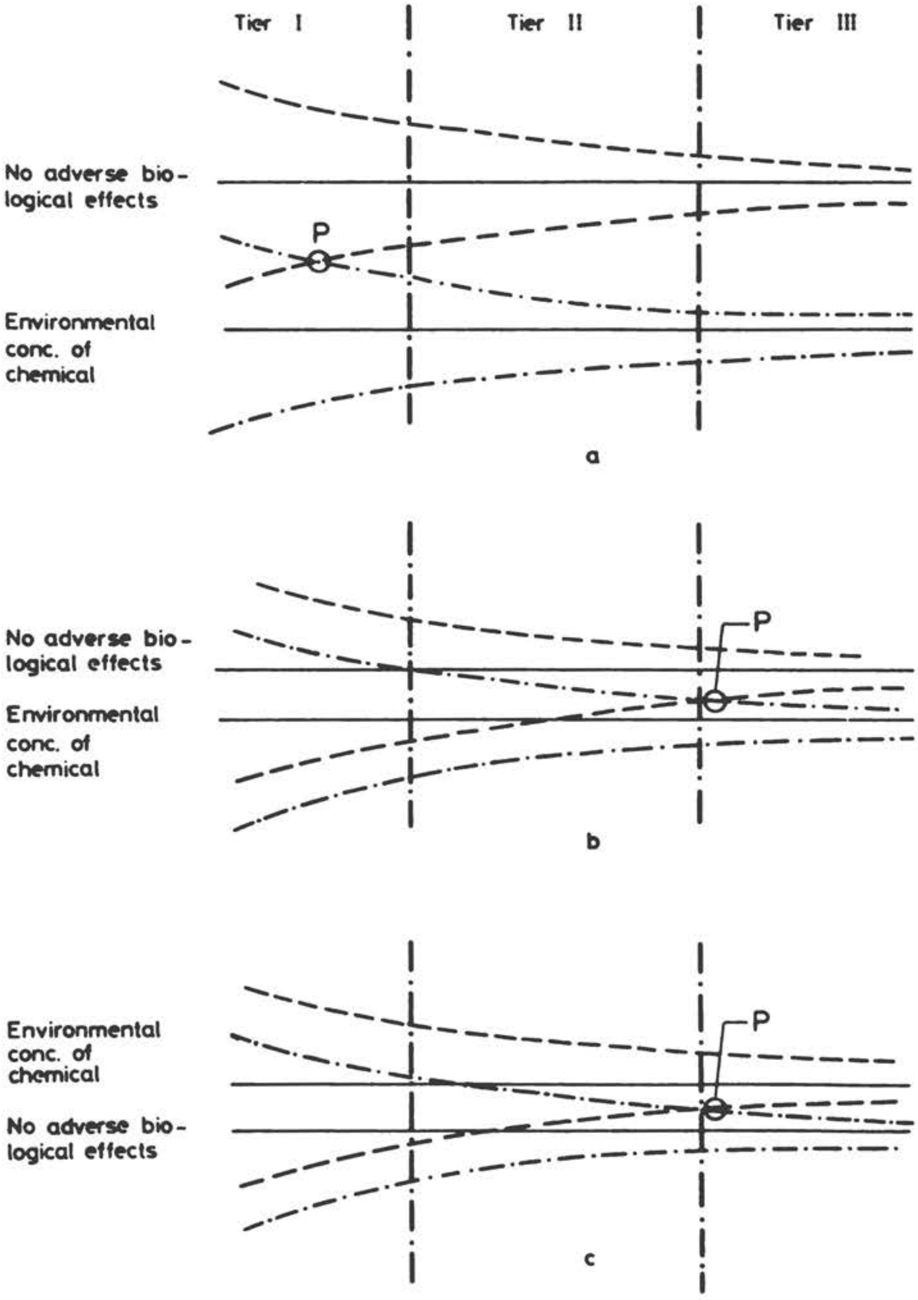


FIGURE 6.1 The relationship of a chemical concentration that produces no adverse biological effects with the actual environmental concentration of the chemical [from Cairns and Dickson (1978), copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, Pa., adapted with permission].

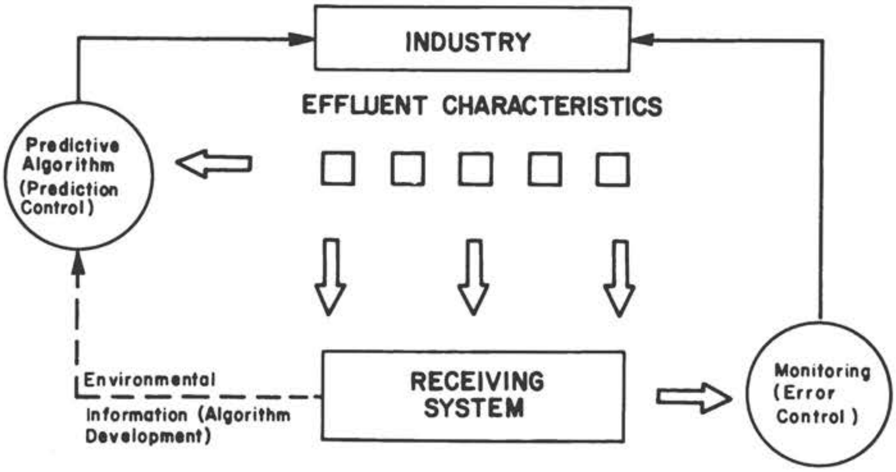


FIGURE 6.2 Information flow in environmental quality control processes (from Herricks and Cairns, 1979).

into two basic groups--*structural* and *functional* characteristics of the biological community. Structural characteristics usually involve counts of both kinds and numbers of species present. In Europe the saprobian system (e.g., Ferginstad, 1965) or the Trent Index (Woodwiss, 1964) are examples of commonly used structural-assessment methods. In the United States, diversity indices (e.g., Wilhm and Dorris, 1968) are regularly used, although a variety of other methods (e.g., Beak *et al.*, 1959) are also used. Functional indices are used to determine whether the biological system is performing properly. These are frequently based on rate processes such as carbon fixation, detritus processing (e.g., Paul *et al.*, 1978), colonization (e.g., Cairns *et al.*, 1979), and the like. Stations (sampling areas) in the exposed portion of the aquatic ecosystem are compared with reference (or control) stations in unexposed areas. A major change in one of the exposed stations relative to the reference

		Water Solubility Testing																															
		+																-															
		Partition Coefficient (Octanol/H_2O) Testing																															
		+								-								+								-							
		Adsorption Testing																															
		+				-				+				-				+				-				+				-			
		Desorption (Leachability) Testing																															
		+		-		+		-		+		-		+		-		+		-		+		-		+		-		+		-	
		Volatility Testing [Hi-(+), Lo-(-)]																															
		+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	-
Probable Sites of Distribution	Air	X		X		X		X		X		X		X		X		X		X		X		X		X		X		X		X	
	H_2O - Aqueous Phase	X	X	O	O	X	X	X	X	X	X	O	O	X	X	O	O																
	H O-Sediment	O	O	X	X			X	X	O	O	X	X	O	O	X	X	O	O	X	X	O	O	O	O	O	O	O	O	O	O	O	O
	Soil	X	X	X	X			X	X			X	X			X	X	X	X	X	X												
	Animals-(Bioaccumulation)	X	X	X	X	X	X	X	X									X	X	X	X	X	X	X	X								

X - Primary Sites of Distribution
O - Distribution Sites of Secondary Importance

FIGURE 6.3 Testing for environment mobility [from Stern and Walker (1978), copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, Pa., adapted with permission].

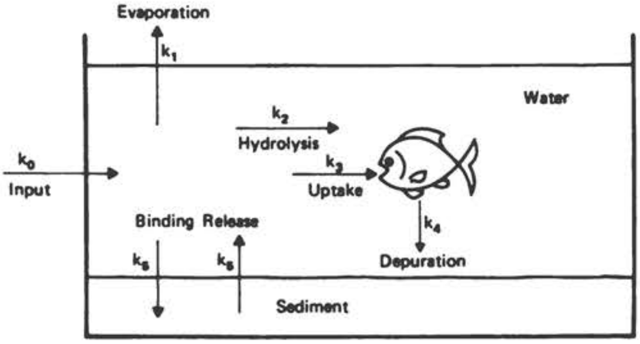
station is a signal for corrective action (reactive control). A recent summary of the literature on reactive control methods can be found in Hellawell (1978).

HAZARD EVALUATION

The basic strategy for developing laboratory protocols for hazard evaluation consists of two parts: (1) *determining the environmental pathways of potentially hazardous substances* and (2) *selecting a group of test organisms and methods that will best define the boundary conditions within which damage may result.*

Environmental Pathways

The determination of the fate of a potentially hazardous substance introduced into the environment is a necessary precursor of any meaningful attempt to establish effective standards for regulatory purposes. For example, a chemical that becomes associated primarily with sediments requires a different assessment strategy than one associated with the water column. Toxicity tests for the former should involve benthic organisms, whereas planktonic organisms would be most appropriate for the latter. Additionally, a rapidly degrading substance would require far less chronic tests on aquatic organisms than a persistent one. Figure 6.3 provides an example of an approach designed to identify the principal environmental component into which a chemical may be distributed after release into the environment. In this example the following series of tests were used: water solubility, partition coefficient (octanol-water), adsorption by natural solids, desorption (leaching), and volatility. For example, if a chemical is soluble in water, does not transfer to octanol,



MATERIAL BALANCE EQUATION

$$V\frac{dCw}{dt} = \underbrace{k_0}_{\text{Input}} - \underbrace{k_1 ACw}_{\text{Evaporation}} - \underbrace{k_2 VCw}_{\text{Degradation (Hydrolysis)}} - \underbrace{k_3 FCw}_{\text{Fish Uptake}} + \underbrace{k_4 FC_f}_{\text{Fish Depuration}} + \underbrace{k_5 SCw}_{\text{Sediment Binding}} + \underbrace{k_5 SC_s}_{\text{Sediment Release}}$$

Where V = Volume of water, ml
A = Surface area, cm^2
F = Fish mass, gm
S = Sediment mass, gm
Cw = Concentration of chemical in water
k = Rate constant
C_f = Concentrations of chemical in fish
C_s = Concentration of chemical in sediment

FIGURE 6.4 Pond model [from Branson (1978), copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, Pa., adapted with permission].

TABLE 6.2 Predicting the Fate of Chlorpyrifos in a Pond[a]

	7-day[b] Concentrations of Chlorpyrifos	
	Water[c], g/liter	Fish, g/g
Predicted	1.0	0.8
Found	1.0	1.0
	Percentage of Total	
Compartment	2 days	25 days
Water	48.0	0.8
Soil	25.0	0.5
Fish	0.8	0.1
Air	3.8	11.4
Metabolized	2.9	11.0
Hydrolyzed	25.0	76.0

[a] From Branson (1978), Copyright, American Society for Testing and materials, 1916 Race Street, Philadelphia, Pa. Adapted with permission.

[b] Similar agreement at days 2, 4, and 28.

[c] 5.7 g/liter at t = 0.

TABLE 6.3 Properties of a Chemical Useful in Predicting the Concentration of the Chemical in Aquatic Environments[a]

Property
- Molecular structure
- Water solubility
- Vapor pressure
- Adsorption spectra (ultraviolet, visible)
- Particle size (if substance is particulate)

Rate Constants
- Photodegradation (ultraviolet, visible)
- Biological degradation
- Chemical degradation
- Evaporation
- Sediment binding
- Uptake by organisms
- Depuration by organisms

Partition Coefficients
- octanol-water
- air-water
- sediment-water

[a] Methods for deriving these properties are being prepared by the Task Groups of ASTM Subcommittee E35.21 on Safety to Man and Environment [from Johnson *et al.* (1978), copyright, American Society of Testing and Materials, 1916 Race Street, Philadelphia, Pa. Adapted with permission].

does not readily adsorb to soils, readily leaches from areas in which it is deposited, and has a low degree of volatility, testing of persistence and ecological effects could be limited to these conditions and biological targets associated with the liquid phase of water bodies and, to a lesser degree, their sediments. The integration of data developed in tests of environmental mobility with the information required by Section 8 of the Toxic Substances Control Act (TSCA; Public Law 94-469) will provide a useful indication of possible "target" organisms.

A new and relatively untested approach that shows promise is called the *environmental rates approach* (Branson, 1978). This requires that properties be measured as time-concentration rates. These rates are incorporated into a suitable model for predicting environmental concentrations. Figure 6.4 shows a pond model, the key properties, and the materials-balance equation for predicting the fate of chlorpyrifos in the pond. The predicted and experimentally found concentrations in the fish and water (Table 6.2) show rather good agreement.

It is important to recognize that the environmental concentration of a chemical is governed by (1) properties of the chemical, (2) rate of introduction into the environment, and (3) characteristics of the specific environment(s) into which it is introduced. Many of the properties useful in this regard are so fundamental that they are likely to be available from other data banks before biological program design begins. Table 6.3 lists some of these characteristics. The environmental characteristics or properties are equally fundamental (Table 6.4).

The determination or estimation of persistence is an important factor in designing biological testing, particularly where the length of the test is concerned. Figure 6.5 depicts a simple schematic for persistence testing. Aquatic sediments and terrestrial soils in some areas contain large amounts of chemical contaminants that, if available, could be significant hazards. The properties of the sediments and the characteristics of the associated microbiota control availability, although these factors are not well defined. Studies are needed to relate a chemical's properties and the properties and characteristics just mentioned to bioavailability in order to predict the environmental behavior of both new chemicals and chemicals already in the environment (e.g., PCBs, kepone).

TABLE 6.4 Properties of Aquatic Environments Useful in Predicting the Fate and Concentration of a Chemical in Those Environments[a,b]

- Surface area
- Depth
- pH
- Flow/turbulence
- Carbon in sediment
- Temperature
- Salinity
- Suspended sediment concentration
- Trophic status
- Adsorption spectra (ultraviolet, visible)

[a] From Johnson *et al.* (1978). Copyright, American Society of Testing and Materials, 1916 Race Street, Philadelphia, Pa. Adapted with permission.

[b] One might also add to the list (1) dilution and (2) inflow-outflow of replacement time for water.

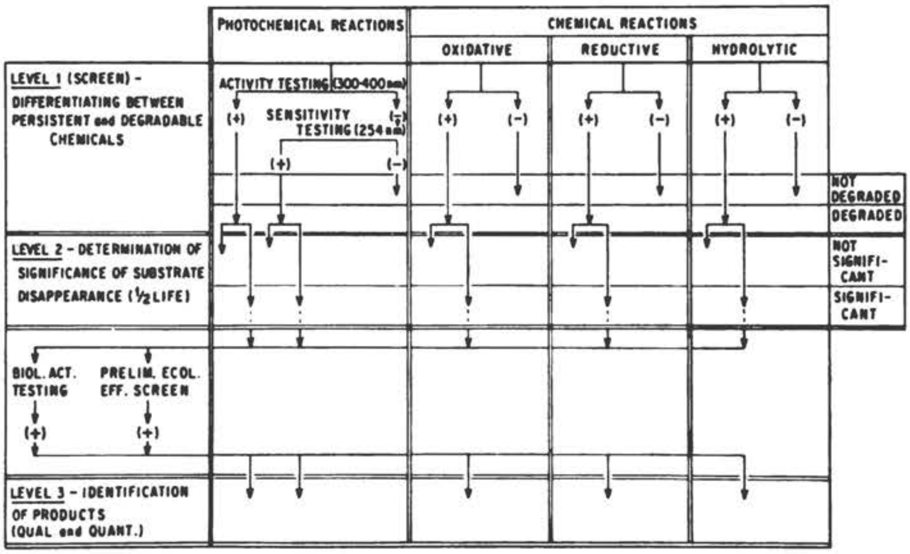


FIGURE 6.5 Persistence testing [from Stern and Walker (1978), copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, Pa., adapted with permission].

Some specific information needs (Maki *et al.*, in press) that will improve predictive capabilities are as follows:

1. Determining the effects of microbial transformation processes on the environmental fate of chemicals.
2. Developing predictive models for fate determination of toxic chemicals in the environment (i.e., where do they go and how long do they persist?)
3. Determining the type of biotransformation data needed to predict the environmental concentrations of chemicals.
4. Determining the effects of chemical-physical variables on biological transformation of chemicals.
5. Quantitatively evaluating the efficacy of extrapolating laboratory data to the "real world."
6. Evaluation of the methods used to describe the kinetics of microbial transformations of chemicals.

An important element is the preparation of a hazard estimate for criterion documents (paragraph 11 of the *Settlement Agreement in Natural Resources Defense Council et al. v. Train*, 8 ERC 2120-D.D.C., 1976--required the U.S. Environmental Protection Agency (EPA) to publish water quality data for 65 specified pollutants by June 30, 1978). An overview diagram of the integration process is given in Figure 6.5. The screening sequence shown was designed for testing only photochemical and chemical degradation. As a consequence, other forms of degradation should be added to the sequence where it is important. Figure 6.5 utilizes three levels of screening. Level 1 (similar to the previously discussed Tier system) was designed to differentiate between persistent and degradable chemicals. Those that prove to be persistent would then have detailed appropriate tests performed for ecological effects. Degradable chemicals would be shunted to Level 2 testing, which consists of a determination of the rate of substrate disappearance usually involving chemical half-life determinations. If Level 2 testing shows that degradation is insignificant, the compound will receive the same ecological testing as the persistent chemicals identified at Level 1. Because Level 3 testing (identification and quantification of degradation products and their rates/conditions of formation) is so expensive (costs in excess of $100,000 would be common), it should only be carried out if toxicity testing reveals significant adverse effects on human health or other organisms.

Test Organisms and Boundary Conditions

There are some fundamental differences in the strategies to protect humans from deleterious environmental conditions and those designed to protect other organisms: (1) humans are a single species--other species are so numerous a complete count has not yet been made; (2) direct tests upon humans that might result in harm are illegal in most societies, and such tests must be carried out on surrogate species; (3) direct tests on "other" species are, with some exceptions, permitted but they are too numerous to carry out tests for all, even if they could all be maintained in the laboratory; (4) the amount of risk to which other species may be exposed is considerably greater than that tolerated for humans; and (5) it is possible that community response may be more than the aggregate of individual species response to stress.

The basic problem may be summarized as one of extrapolating from a small number of individuals of a few species to a comparatively vast number of species. The difficulty is exacerbated by the need to extrapolate from a small range of test conditions to a comparatively vast array of conditions with many parameters that vary independently from each other. How can one be reasonably confident that the variability in response of the natural system has been fully exposed? One is, of course, most concerned that the most vunerable organisms have been identified so that a no-adverse-biological-effects threshold that will ensure the protection of all species has been identified. There are two tactics that enhance the probability of achieving this goal: (1) use an array of species from a number of trophic levels in the screening tests and (2) carry out thorough studies of the most sensitive species to ensure that the most sensitive stage in the life cycle has been tested (Figure 6.5).

Figure 6.6 shows a laboratory protocol prepared in 1973 for the U.S. Army Medical Research and Development Command and subsequently published (Cairns and Dickson, 1978). The delay in publishing was to provide an opportunity for the protocol to be tested at a variety of munition plants. It worked quite well, although some of the more recently developed protocols summarized in Dickson *et al.* (1980) will undoubtedly prove to be superior. The initial steps for both laboratory and field protocols are identical:

1. *Characterization of the Compound (or Stress)* Generally most of the information is already available for nonbiological reasons. Examples in this category are water solubility, molecular structure, vapor pressure, particle size, photodegradation, and partition coefficients (octanol-water; air-water; sediment-water). Further details can be found in Cairns *et al.* (1978).
2. *Characterization of the Receiving System* Water hardness, temperature, dissolved oxygen concentration, flow variability, stream order, and other important limnological characteristics.
3. *Determination of Testing Priority* The large number of chemicals being produced each year (thousands in this country alone) and the scarcity of competent personnel, facilities, and funds makes establishment of a testing priority mandatory. Table 6.5 shows a useful way of determining priority for testing.
4. *Toxicity Evaluation Protocol* The laboratory protocol used as an illustration has two basic strategies: (1) An array of species from different trophic levels is used for the inexpensive screening tests (Tier I) and from these the most sensitive is selected for the more detailed and expensive tests (Tiers II and III). The same strategy is used for selecting the most important nonbiological parameters. (2) Data gathering alternates with decision "boxes" so that one need not generate more information without scientific justification for doing so. This has proved to be a useful approach for the U.S. Army Medical Research and Development Command and should, appropriately modified, prove to be useful in other situations as well. The strategy for the field protocol (Figure 6.7) is essentially similar, but fish are eliminated from the screening studies for financial and other reasons.

Probably the most worthwhile modification would be to expand the range of trophic activities so that as many as 10 species representing different components of the food web are represented in the screening tests. Thus, one might have a detritus feeder, a primary producer, a primary consumer, a secondary consumer, and other trophic levels in the test series.

BIOLOGICAL MONITORING

Biological monitoring is merely a systematic surveillance of the effects of a potential stressor (e.g., a potentially toxic substance) for the purpose of maintaining quality control in the biota of the system into which the substance is being introduced. Without the directive to take immediate corrective action when deleterious effects are detected, biological

TABLE 6.5 Determination of Priority for Testing[a]

Predictability of Toxicological Characteristics of Chemical Substance	High (1)	Medium (2)	Low (3)[b]
Widespread distribution (4), substantial amounts released	4	8	12
Widespread distribution (3), small amounts released	3	6	9
Localized distribution (2), substantial amounts released	2	4	6
Localized distribution (1), small amounts released	1	2	3

[a] To determine final priority number, multiply the appropriate number above by the following vulnerability factor for the receiving system: high = 3, medium = 2, low = 1 (from an approach determining priority testing, NRC Committee on Toxicology, 1975).

[b] High number indicates high priority.

TOXICITY EVALUATION PROTOCOL – LABORATORY STUDIES

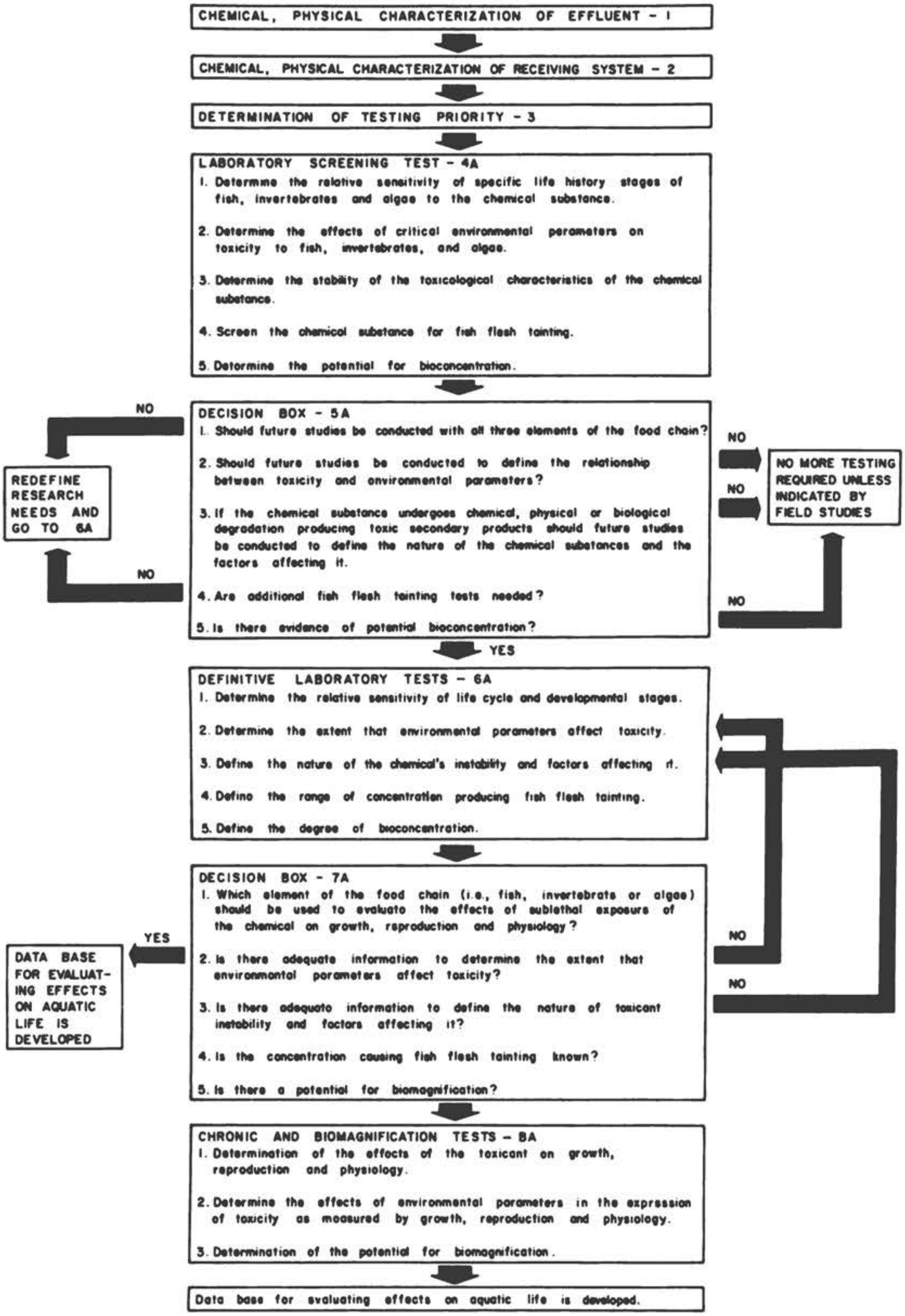


FIGURE 6.6 Toxicity evaluation protocol-laboratory studies [from Cairns and Dickson (1978), copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, Pa., adapted with permission].

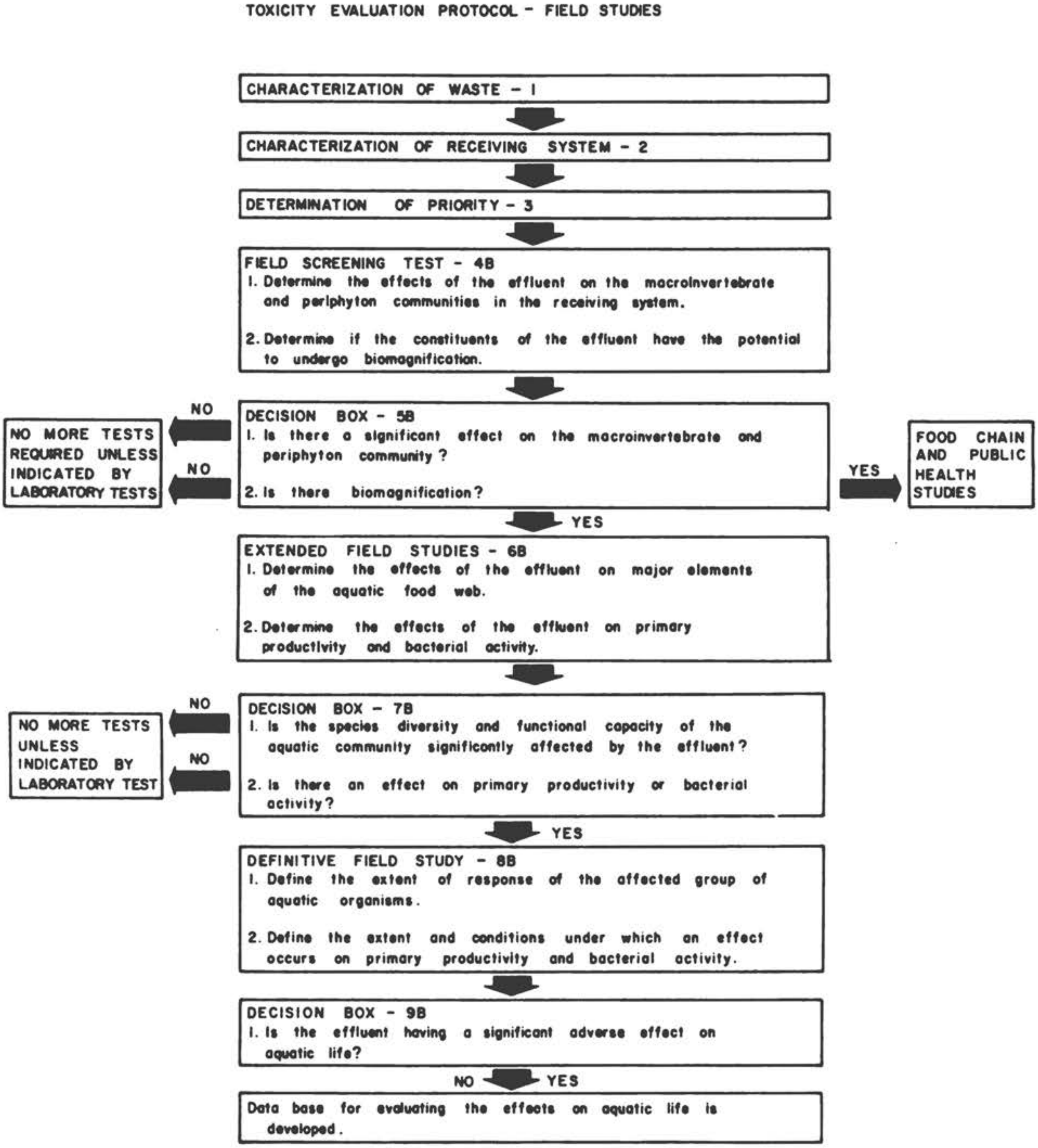


FIGURE 6.7 Toxicity evaluation protocol-field studies [from Cairns and Dickson (1978), copyright, American Society for Testing and Materials, 1916 Race Street, Philadelphia, Pa., adapted with permission].

monitoring is merely giving the illusion of protection without the reality. Biological monitoring, as just defined, is relatively rare at this time. However, civilization is now in a transitional period in some ways analogous to the agricultural revolution. In that period the unmanaged environment was not producing sufficient food to meet society's needs. As a consequence, humans found themselves in a new management role. Similarly, the unmanaged environment is now not capable of assimilating all of society's wastes, *as now delivered*, without being harmed. The rate at which potentially harmful materials enter the environment must be synchronized with the environment's ability to assimilate these wastes. Because industrial wastes and other potentially deleterious materials entering the environment vary in both quality and quantity at the entry site and the environmental assimilative capacity also varies but not in synchrony with the former, management is essential if the benefits of both systems are to be optimized and the debits minimized. The information feedback loop that makes effective quality control possible is called biological monitoring.

There are two basically different types of biological monitoring systems: (1) biological early warning systems that alert management personnel to the appearance of deleterious substances before they enter the receiving system and (2) receiving-system methods that detect biological changes in the system itself. A recent summary of the literature on early-warning systems may be found in Cairns and van der Schalie (in press). Annual reviews of literature useful in receiving-system monitoring (but not necessarily so labeled) may be found in the *Journal of the Water Pollution Control Federation*.

Almost any species, community, or parameter can be used for biological monitoring. Appropriate selection will depend on the conditions at each site. In addition to the scientific justification of the ecological importance of the species and/or parameter used, the sensor chosen should be reliable in terms of information produced and not particularly difficult to work with. If monitoring methods are to be widely employed, they must not be too difficult for a technician to use routinely.

Single Species Versus Community Criteria

Most of the tests upon which decisions are made regarding the concentrations needed to protect natural communities from hazardous chemicals or other stressful situations (e.g., heated wastewater or suspended solids) are carried out with single species under carefully controlled laboratory conditions. Only a comparatively few species are now used for test purposes. Fortunately, microcosm tests using simple communities are being developed by R. L. Metcalf (University of Illinois), and others. These are not now commonly used, but interest in them is high. There are several underlying assumptions to our present regulatory policy. First, protecting the most sensitive significant species in a system can be accomplished without testing that species directly. Second, community response will not vary so much that site-specific tests in all areas are necessary. These assumptions are not explicitly stated but are implicit in criterion and standard regulatory documents that rely on a few tests or a few species for widespread determination of "safe" concentrations. There are three key questions that should be addressed as soon as possible.

1. Are the commonly used bioassay species good surrogates for predicting responses of entire communities, and does an application factor derived from tests carried out on a single species actually protect an entire community?
2. Do communities in different areas made up of different aggregations of species respond in a similar way to an identical concentration of a toxicant? (In this instance one might have to allow for differences in water quality that affect toxicity.)
3. Do communities of different maturity respond similarly to identical concentrations of a toxicant?

Structural Versus Functional Parameters

Most of the biological measurements for assessing pollutional impacts on receiving systems have been based on species inventories or "critter counts." Comparatively few tests on function (e.g., ATP activity, carbon fixation, and autotrophic-heterotrophic ratios) are used. One can hypothesize three relationships between structure and function but, unfortunately, evidence to select the most probable is sparse. These possibilities are as follows:

1. That aquatic community structure and function are so intimately associated that one cannot be altered without altering the other. If this hypothesis is correct, only one type of assessment needs to be used in reactive or error control to ensure protection of the other.
2. That there is so much functional redundancy in most natural communities (i.e., many species feeding on detritus in a particular way) that loss of a species (which would affect structure) would not impair function. If this assumption is correct, the best receiving system parameters for error control would be based on species inventories, because structure would be more sensitive than function.
3. That it is possible to impair function without killing or eliminating species (e.g., inhibition may occur at much lower concentrations than lethality). In this case the best receiving system parameters for error control would be based on function.

Evidence is not sufficient to make a reliable choice among the assumptions just stated, however, it would be prudent to have a mix of both structural and functional receiving system parameters for error control.

USE OF BIOLOGICAL INFORMATION

There are some heartening indications that biological information based on receiving system conditions can be used as evidence for less stringent effluent limitation. The most notable example is the three types of demonstrations needed to obtain a variance to closed-cycle cooling [described in regulations published by the EPA in the *Federal Register*, 1974]. Public Law 92-500, Section 316(a) calls for the following demonstrations:

Type 1 The use of onsite data to demonstrate absence of prior appreciable harm to the biological community.

Type 2 A demonstration using EPA water-temperature criteria that conditions at the site would protect important representative species.

Type 3 The use of a combination of biological and engineering data to demonstrate no adverse effects.

Although it has not happened yet, there is no scientifically justifiable reason why these demonstrations could not be carried out for certain classes of chemicals (e.g., those that are neither persistent nor likely to undergo biological magnification). This would replace arbitrary standards based on little site-specific information with site-specific error control based on biological information based on receiving system condition.

The Toxic Substances Control Act (TSCA) signed into law (Public Law 94-469) on October 11, 1976, is an interesting example of prediction control based on a mixture of biological and chemical-physical information. This law provides that no person may manufacture or process a new chemical substance or manufacture or process a chemical substance for a new use without obtaining clearance from the EPA. One of the main objectives of the TSCA is to establish a procedure for evaluating hazards to human health and the environment before widespread use of a new chemical. In short, this is a form of *predictive error control* using biological information. After examining the data and evaluation, the administrator of the EPA must judge the degree of risk associated with the extraction, manufacturing, distribution in commerce, processing, use, or disposal of the chemical substance. If it presents an unreasonable risk to health or the environment, the administrator of the EPA may restrict its use or ban it.

ECOSYSTEM REHABILITATION

If both predictive and error-control systems fail or if an accident or catastrophe (e.g., a major oil spill) occurs, ecosystem rehabilitation or restoration will be necessary. This must be carried out with a full understanding of both the structural and functional characteristics of the ecosystem state to be achieved. Rehabilitation must also be carried out with a knowledge of the most effective and efficient technology available for the process. These technologies include the effective manipulation of the physical and chemical qualities of habitats, as well as the technologies of effective species stocking and restoration programs. The emerging field of ecosystem restoration and pollution mitigation should become a basic component of an ecosystem management capability. Those engaged in this new field must develop a detailed knowledge of the fates and effects of each type of system perturbation. Practitioners will need to know both the rates of recovery for different systems and the limits of tolerance beyond which perturbed systems cannot recover to desired states. Additionally, we must determine the degree of reversibility of the states of the systems that have been subjected to disturbance (e.g., Cairns, 1980). The ecosystem damage suffered and the subsequent recovery potential of perturbed ecosystems depend primarily on three factors: (1) the ecological type of system perturbed, (2) the nature of the disturbance, and (3) the environmental pathway of the disturbance agent.

SUMMARY

The development of predictive and reactive (error) control systems for aquatic ecosystem quality control will not occur unless biologists feel a professional responsibility to contribute time and energy. Additionally, changes in legislation and regulatory agency policy to permit ecosystem quality control based on the methods discussed is mandatory. Finally, without concurrent development of organizational capabilities for ecosystem management, the other efforts will not produce the desired results.

ACKNOWLEDGMENTS

The author acknowledges the assistance of Darla Donald in the editing and preparation of this manuscript and Betty Higginbotham for the final typed copy.

REFERENCES

Barton, J. R., E. J. Peters, D. A. White, and P. V. Winger (1972). Bibliography on the physical alteration of the aquatic habitat (channelization) and stream improvement, Brigham Young University Publication, Provo, Utah, 184 pp.

Beak, T. W., C. deCourval, and N. E. Cooke (1959). Pollution monitoring and prevention by use of bivariate control charts, *Sew. Ind. Wastes 31*, 1383.

Branson, D. R. (1978). Predicting the fate of chemicals in the environment from laboratory data, in *Estimating the Hazard of Chemical Substances to Aquatic Life, STP 657*, J. Cairns, Jr., K. L. Dickson, and A. W. Maki, eds., American Society for Testing and Materials, Philadelphia, Pa., pp. 55-70.

Cairns, J., Jr. (1975). Critical species, including man, within the biosphere, *Naturwissenschaften 62*, 193-199.

Cairns, J., Jr. (1978). The modification of inland waters, in *Wildlife and America*, H. P. Borkaw, ed., Council on Environmental Quality, U.S. Government Printing Office, Washington, D.C., pp. 146-162.

Cairns, J., Jr., ed. (1980). *Recovery Process in Damaged Ecosystems*, Ann Arbor-Humphrey Science Publishers, Inc., Ann Arbor, Mich., 167 pp.

Cairns, J., Jr., and K. L. Dickson (1978). Field and laboratory protocols for evaluating the effects of chemical substances on aquatic life, *J. Test. Eval. 6(2)*, 81-90.

Cairns, J., Jr., K. L. Dickson, and A. W. Maki, eds. (1978). *Estimating the Hazard of Chemical Substances to Aquatic Life, STP 657*, American Society for Testing and Materials, Philadelphia, Pa., 278 pp.

Cairns, J. Jr., D. L. Duhn, and J. L. Plafkin (1979). Protozoan colonization of artificial substrates, in *Methods and Measurements of Periphyton Communities: A Review, STP 690*, R. L. Weitzel, ed., American Society for Testing and Materials, Philadelphia, Pa., pp. 34-57.

Cairns, J., Jr., and W. H. van der Schalie (in press). Biological monitoring, Part I: Early warning systems, *Water Res.*

Dickson, K. L., A. W. Maki, and J. Cairns, Jr., eds. (1980). *Philosophy and Implementation of Hazard Assessment Procedures for Chemical Substances in the Aquatic Environment*, American Fisheries Society, Beltsville, Md., 159 pp.

Environ-Control, Inc. (1974). *Total Urban Water Pollution Loads: The Impact of Stormwater*, Council on Environmental Quality, Environmental Protection Agency, and the U.S. Department of Housing and Urban Development, Washington, D.C., 236 pp.

Ferginstad, E. F. (1965). Some remarks on a new saprobic system, in *Biological Indicators of Pollution*, Environmental Health Series, U.S. Public Health, No. 999-WP-25, pp. 232-235.

Federal Register (1974). Vol. 39, October 8.

Hellawell, J. M. (1978). *Biological Surveillance of Rivers*, Water Research Centre, Stevenage, England, 331 pp.

Herricks, E. E., and J. Cairns, Jr. (1979). Monitoring and mitigation of aquatic hazards, in *Control of Specific (Toxic) Pollutants*, E. R. Frederick, ed., Air Pollution Control Association and Water Pollution Control Federation, Washington, D.C., pp. 220-231.

Jensen, L. D. (1974). *Second Workshop on Entrainment and Intake Screening*, The Johns Hopkins University, Department of Geography and Environmental Engineering Rep. No. 15, The Johns Hopkins Press, Baltimore, Md., 192 pp.

Johnson, H. E., R. Parrish, G. F. Lee, C. E. Johnson, J. Sanborn, D. Hansen, and J. Hamelink (1978). Discussion session synopsis, in *Estimating the Hazard of Chemical Substances to Aquatic Life, STP 657*, J. Cairns, Jr., K. L. Dickson, and A. W. Maki, eds., American Society for Testing and Materials, Philadelphia, Pa., pp. 71-77.

Kerr, J. E. (1953). Studies on fish preservation at the Contra Costa steam plant of the Pacific Gas and Electric Company, California Department of Fish and Game, *Fish. Bull. 92*, 87 pp.

Macek, K. J., D. F. Walsh, J. W. Hogan, and D. D. Holz (1972). Toxicity of the insecticide Dursban to fish and aquatic invertebrates in ponds, *Trans. Am. Fish. Soc. 101*, 420-427.

Maki, A. W. (1979). An analysis of decision criteria in environmental hazard evaluation programs, in *Analyzing the Hazard Evaluation Process*, K. L. Dickson, A. W. Maki, and J. Cairns, Jr., eds., American Society for Testing and Materials, Philadelphia, Pa., pp. 83-100.

Maki, A. W., K. L. Dickson, and J. Cairns, Jr., eds. (in press). *Biotransformation and the Fate of Chemicals in the Environment*, American Society of Microbiology, Washington, D.C.

Mill, T., J. H. Smith, W. Mabey, B. Holt, N. Bohonos, S. S. Lee, S. Bomberger, and P. W. Chou (1977). Environmental exposure assessment using laboratory measurements of environmental processes, in *Proceedings of the Ecology Symposium*, Corvallis, Oreg., pp. 1-21.

NRC Committee on Toxicology (1975). *Principles for Evaluating Chemicals in the Environment*, National Academy of Sciences, Washington, D.C., 454 pp.

Neely, W. B., and G. E. Blau (1977). In *Pesticides in Aquatic Environments*, Klan, ed., Plenum Press, New York, pp. 145-163.

Paul, R. W., Jr., E. F. Benfield, and J. Cairns, Jr. (1978). Effects of thermal discharge on leaf decomposition in a river ecosystem, *Vehr. Int. Verein. Limnol. 20*, 1759-1766.

Smith, L. L., Jr., S. I. Auerbach, J. Cairns, Jr., D. I. Mount, G. A. Rohlich, J. B. Sprague, and W. L. Klein (1974). *ORSANCO 24-Hour Bioassay*, Biological Water Quality Committee, Ohio River Valley Water Sanitation Commission, Cincinnati, Ohio, 21 pp.

Stern, A. M., and C. R. Walker (1978). Hazard assessment of toxic substances: Environmental fate testing of organic chemicals and ecological effects testing, in *Estimating the Hazard of Chemical Substances to Aquatic Life, STP 657*, J. Cairns, Jr., K. L. Dickson, and A. W. Maki, eds., American Society for Testing and Materials, Philadelphia, Pa., pp. 81-131.

Wilhm, J. L., and T. C. Dorris (1968). Biological parameters for water quality criteria, *BioScience 18*, 477-481.

Woodwiss, F. S. (1964). The biological system of stream classification used by the Trent River Board, *Chem. Ind. (NY) 11*, 443-447.

Robust Estimators in Hydrology

7

MYRON B. FIERING *and* GEORGE KUCZERA
Harvard University

INTRODUCTION

Hydrology serves at least two disjoint clienteles--the scientific community and the engineering community. This rift is exacerbated by recent emphasis on modeling techniques. The scientific community is concerned primarily with causality and proposes increasingly complex and sophisticated models of natural phenomena, whereas the engineering community is concerned with water management and is closely identified with models involving operations analysis and decision theory. As the two groups become more highly specialized, interaction and collaboration diminish. Robust statistics may be particularly important to help bridge many of the boundaries between these two disciplines.

Robustness is difficult to define (Fiering, 1982a, offers 12 alternative definitions). The concept of stability of parameters under a moderate perturbation among the data is a necessary condition for predicting system changes under new conditions or management policies. Robustness accommodates features of controlled dynamic response--measured rather than precipitous changes in parameters and system performance.

We maintain that statistics and economics are the eyes and ears of formal and informal decision making, whereupon models of hydrologic processes are useful principally to help decide among alternatives. The bases for these decisions are the predictions derived from the models. Hence it is naive to argue whether these models are *descriptive* or *prescriptive*; in fact, they are both. We fit them to observations, attempt to decide if they fit "well enough," and if so learn and describe something of the hydrology of the basin. We then typically use the models to predict performance under a new set of circumstances.

ROBUSTNESS

Conventional water-resource analysis focuses on the time trace of water available in a basin, on the storage and control works, on the societal tensions derived from competing uses, and on the decision calculus selected for design and management. A water balance governs the entire analysis; attempts to balance projected supply and demand display a tendency to self-fulfillment--all of the available resource is utilized! Long-term planning deals with supply and demand of water as it pertains to economic growth. In addition, such planning should predict important alterations in existing regimes, display significant regional trends, and identify areas subject to current or potential stress.

To do so using models of hydrologic processes requires (1) substantial improvement in our ability to model mechanistic aspects of hydrology and to assess the parameters of these models; (2) improved mathematical, computational, and statistical techniques; (3) enhancement of data collection and retrieval systems; and (4) a decision-making calculus that we propose be based on concepts of resilience and robustness.

Robustness and Resilience

The concept of robustness of design has analogs in other scientific and social settings (Fiering and Holling, 1974). Traditional methodologies for designing against probabilistic phenomena seek to reduce to some prescribed level the probability of occurrence of damaging consequences. Engineers historically have built dams and levees for flood control but have only recently begun to change the free body--the mechanics and economics of the system--by weather modification, early-warning systems, systematic flood proofing, zoning, and flood insurance. These alternatives delineate and stretch the boundaries of acceptable outcomes rather than control the probability of an unacceptable event.

Dams and levees reduce hydrologic threat. In so doing, the community develops along patterns that make it more susceptible to catastrophic damage when, inevitably, the erected restraints prove to be inadequate. This form of behavior is common in situations involving exposure to risk. For example, we drive faster on roads equipped with guardrails. This adjustment tends to drive the "system" toward stability; we lose fewer cars over cliffs protected by guardrails. But stability does not imply robustness, which is the ability of the system to accommodate surprise and to survive under anticipated perturbation. Therefore, a blowout along a road with a guardrail could be more dangerous than on a road without one because speeds tend to be higher with a guardrail present. Stability is not necessarily a desirable characteristic of system performance.

Systems cannot be designed to be *fail-safe*, at least not over reasonable time periods and at reasonable costs; system design should reflect consideration of the rate and cost of repair or restoration. Flood-control structures tend to make the system more brittle, more vulnerable to surprise. In ecological terms this is equivalent to drawing tighter the boundaries of acceptable behavior; it is analogous to what engineers do when they impose rigid standards for ecosystem management without proper regard for those wide swings in state variables that characterize many healthy, recovering systems.

We can attempt to alter the nature and frequency of the flood threat by weather modification. There are important uncertainties inherent in this policy, so we can further secure our position by a series of staged *safe-fail* alternatives that protect against the inevitable failure or flood event. In ecological terms we define a failure as a precipitous "flip" of the ecosystem into an unacceptable mode or state. We try to prevent failure by avoiding the boundaries of acceptable ecosystem behavior but also try to maintain and extend these boundaries to keep them elastic or flexible, by modifying system dynamics (such as introducing new species or cutting firebreaks). Thus, we attempt to avoid failure by avoiding the boundaries that separate acceptable from unacceptable positions and, at the same time, by pushing the boundary away from current and predicted positions. Prediction over a time horizon consistent with the boundary's elasticity is imperative if adaptive, resilient response is sought.

Engineers have long attempted to prevent pollution episodes by building *safe-fail* environmental systems, i.e., large waste-treatment plants or double-hulled oil tankers. We must explicitly acknowledge that repair of inevitable failure is an element of resilient design. Global optimality might be calculated readily enough but its utility rests on the precision of a few parameter estimates, such as on the numerical stability of solutions to large optimization routines or on the suitability of the objective function to the several conflicting interests impacted by the decision. We must ask if the optimal solution can be implemented or if a near-optimal solution stands a significantly better chance. We should not be so intent on locating optimality that we compromise resilience for a myopic optimum.

Resilience--An Analogy

Implicit in the notion of resilience is the dimension of time. The sensitivity of system response (Fiering 1982b) to a decision variable x_i is given by the derivative $\partial f/\partial x_i$, which does not reflect dynamic behavior. The total derivative df/dx_i contains terms dx_j/dx_i, so that if x_i is forced through a small change dx_i, the other available decisions x_j can be manipulated to optimize system response. Adjustments in some of these decision variables can be made more rapidly than in others. Some operating decisions--lowering or raising gates on a reservoir and switching fuel in a generating plant--can be made rapidly. Political and institutional adjustments--changing standards or regulations--might require months. Structural changes might require years, and social changes--shifts in land use or demographic patterns--could take decades. Thus, minimizing the total derivative df/dx_i for a prescribed shift away from a solution vector x, or minimizing the differential df subject to constraints on the changes dx_i (e.g., that the new solution lie on or within a unit hypersphere whose center is at the original solution x) suggests a resilient system trajectory from the state given by x to a reduced response level.

In many applications of systems analysis it is customary to assess a penalty function for failure of the system to perform within prescribed ranges; these ranges are called *environmental standards*. The penalty functions are typically increasing with respect to deviations from the standard, although they are not necessarily symmetric. A suggested measure of performance is associated with the rate of decline along the trajectory between initial and terminal (upper and lower) positions. If a trajectory is slowly completed there is often time to adjust to anticipated policy shifts and thus to sustain less damage.

PROBABILITIES OF EXTREMES

Hydrologists live at the tails of density functions. Our advice is requested when things go well, i.e., when precipitation and flow levels are near the centers of their density functions. We are, however, consulted frantically when major droughts and floods are experienced.

Unfortunately, our predictive powers are least impressive at these tails. Long records are typically not available, so that estimates of the higher moments required to define the tails are highly unstable. For example, a transcription error in keypunching the smallest "annual flood" in a record of N annual events may materially affect the estimate of a major flood associated with a long-recurrence interval of T yr? If moments are used to define the density function, this could easily occur.

Should goodness of fit be tested by conventional measures of deviation, when should our concern more realistically reside in the cumulative function? Historically these problems have not been critical in design calculations. They were scaled on the basis of mean and median flows, near the center of their distribution functions, and were consequently not overly sensitive to small errors in hydrologic analysis. In recent years, with more marginal water-resource projects being considered, with management for water quality an important objective, and with irrigation failure in the face of rising population a critical issue in project evaluation, assessment of hydrologic extrema becomes more important for project planning, design, and operation. Good parameter estimation alone does not provide stable estimates of periodicities for major events.

The objective should be to develop distribution functions that are robust in terms of estimating both the probability density (or ordinate) assigned to a particular argument and the cumulative density or integral

$$\int_{-\infty}^{x} f(x)\,dx = F(x).$$

It is not intrinsically useful to develop techniques for robust estimation of parameters if small perturbations in the parameters produce large deviations in the density or cumulative-density functions. It is more important to ask what we are going to do with the estimates of probability and cumulative-density functions, and hence with return periods and expected losses, than to worry excessively about the stability of parameter estimates. These parameters have no inherent value, and their close estimation implies no operational payoff even though the scientific purpose of "better models" is served. Many density functions contain exponential terms, so that even small errors in parameter estimates tend to produce wide fluctuations at the tails. Thus, stability or robustness of the parameters themselves is not inherently useful; it is more important to utilize density functions that yield stable or robust estimates of the recurrence probabilities, even though the parameters of those functions may vary significantly from one estimate (or sample) to another. It is suggested that parameters other than moments be utilized in analyzing hydrologic extremes because it is more acceptable that such parameters vary widely, particularly if there is no obvious physical interpretation readily attached to them.

The Wakeby Distribution

One such function is derived from the Tukey family of λ-densities. It contains five parameters and can most conveniently be written in inverse form, which is in itself robust because the argument is an integral or cumulated area that is inherently more stable than a point or functional value:

$$x = e + c(1 - F)^{-d} - a(1 - F)^{b},$$

where F is the cumulative density $F(x)$, constants a through e are parameters of the function $f(x)$, and x is explicitly given. Thomas and Houghton (as reported in Houghton, 1977) fit this function to a large number of annual flood sequences. There is continuing need for research in fitting the five parameters of this Wakeby density.

Fiering (1978, 1979) fitted regionalized Wakeby functions to low-flow values throughout the state of New Jersey and in the Miami River (Ohio). Greenwood *et al.* (1978) proposed the use of probability moments for estimating the five parameters of the Wakeby density and demonstrated by Monte Carlo analysis the stability of their probability assignments, which result from Wakeby densities whose parameters are highly unstable from sample to sample.

The five parameters define the location and the shape of the two tails of the Wakeby distribution. The heuristic justification for using five parameters inheres in the fact that two parameters are required to define each of the tails and the fifth is a location parameter. By judiciously setting the parameters the Wakeby distribution can be made to mimic the normal, log-normal, log-Pearson III, or virtually any of the commonly used density functions. The flexibility associated with being able to independently fit the two tails provides an important advantage; the cumulative density function is stable (i.e., robust) even though the five parameters may vary widely from sample to sample in the same region.

Old Problems, New Criteria

Neither conventional numerical techniques for measuring goodness-of-fit nor graphical techniques for comparing alternative functions meets the needs of hydrologic analysis. The U.S. Water Resources Council (1977) stated that except for special cases the standard flood density for federal-project analysis would be the log-Pearson III and its associated parameter-estimation techniques. One interesting feature of these requirements is the use of regional statistics to smooth the anticipated sampling fluctuations of the higher moments required to fit the log-Pearson III. This recognizes that the skew coefficient is notoriously unstable for the small sample sizes commonly encountered in hydrologic practice. The use of regional moments represents an important step in the adoption of robust statistics or statistics that are not extremely sensitive to small perturbations of the data base. The James-Stein (Effron and Morris, 1977) technique for estimating moments (the mean, standard deviation, and skew coefficient of flood records) represents familiar quantities. Major errors in their assessment would be detected because the numbers would not "look right." Much of traditional analysis is directed at obtaining stable estimates of these moments. But a small estimating error can lead to egregious misspecification of the return period of major floods even though the density function appears to be acceptably close to the empirical observations. This is because a small error in the fat part of the density, where the mean is typically located, is levered into a much larger change in the recurrence interval for an extreme event. What is needed is a new way of looking at distributions of hydrologic events to incorporate as much robustness as site-specific and national observations can allow. Such a procedure might encourage the inclusion of information from region-wide flood sites, thereby de-emphasizing the importance of individual events at a particular site.

The combination of regional and site-specific flood information was investigated by Kuczera (1980) and is summarized below in the context of a two-parameter, log-normal flood distribution. Empirical Bayes (EB) procedures, conceptually similar to James-Stein procedures, provide a way of augmenting site-specific with regional flood information. The EB procedure is Bayesian with the exception that regional flood information is used to estimate empirically the parameters of the prior distributions of the mean and standard deviation of floods. These prior distributions are then updated by site-specific flood data to yield posterior distributions. Given a loss function, an asymptotically optimal estimator of the T-year flood can be derived. Simulation studies demonstrate that whenever exploitable regional information exists, the EB estimator of the 100-yr flood significantly outperforms the maximum likelihood estimator, particularly for short records.

Analyses of several regions of the United States show a substantial reduction in regional information carried by the prior distribution of the standard deviation when spatial correlations are taken into account. The prior distribution can be further sharpened by augmenting the regional flood records with physiographic data using a regionalization procedure. Analysis of New England basins suggests that the information content carried by the prior distribution of the standard deviation is the equivalent of about 20 "effective years" of site record.

Outliers are the source of much embarrassment to hydrologists; we do not know how to deal with them through conventional forms of analysis. Including them in estimates of moments does not change the mean very much and might not change the standard deviation by a great deal; the skew coefficient is profoundly affected, as is the recurrence interval associated with design events.

To illustrate some of the difficulties that outliers can confer on estimation of extreme events, we undertook a simple Monte Carlo experiment using Houghton's four Wakeby distributions (which purport to represent the range of floods in the United States) as the parent distribution in the Monte Carlo simulation; we evaluated the mean-squared-error performance of three estimators of the 100-yr flood for record lengths of 15 and 30 yr. The estimators considered were the log-normal maximum likelihood (ML) estimator, a robust log-normal method-of-moments estimator utilizing Tukey's biweight estimator of location and Lax's estimator of scale (Mosteller and Tukey, 1977), and the log-Pearson III, method-of-moments estimator. The efficiency of an estimator relative to the ML estimator is defined as the ratio of the ML mean-squared-error to the mean-squared-error of the estimator. The Monte Carlo results are summarized in Table 7.1. Note that in the presence of straggly tails (Type IV parent in the table), the robust log-normal estimator exhibits a significantly better performance than the ML estimator. However, in the case of the Type III parent, the performance of the robust estimator deteriorates markedly. This follows from the fact that these robust estimators of location and scale are designed to stabilize moment estimators when the underlying distribution exhibits a fatter tail than that of the normal distribution; this is not the case for the Type III parent in log space. An adaptive robust estimator of location and scale is needed. Second, the log-Pearson III estimator performs poorly in comparison with the ML estimator except in the case of the Type

TABLE 7.1 Efficiency of Robust Log-Normal and Log-Pearson Estimators of the 100-Year Flood for Different Distributional Assumptions and Record Lengths

Wakeby Distribution Parent Type	Mean-Squared-Error Efficiency of Estimator Relative to Log-Normal Maximum Likelihood Estimator[a]			
	Robust Log-Normal		Log-Pearson III	
	15 Yr	30 Yr	15 Yr	30 Yr
I	1.042 (0.037)	0.966 (0.013)	0.125	0.307 (0.058)
II	0.966 (0.020)	0.922 (0.013)	0.397 (0.042)	0.502 (0.038)
III	0.591 (0.015)	0.674 (0.006)	0.816 (0.089)	1.178 (0.091)
IV	1.697 (0.307)	1.235 (0.064)	0.037	0.109 (0.035)

[a] Where possible, standard errors are reported; plus and minus ranges are in parentheses.

III parent. Because both estimators use almost identical estimates of location and scale, it seems plausible to attribute the poor performance of the log-Pearson III estimator to the instability of the sample skew coefficient in the presence of outliers. This is not to suggest that the log-Pearson approach should be abandoned simply because the ML estimator and its robust version exhibit in some instances uncomfortably high negative biases. Rather the development of robust estimates of skewness needs to be pursued further, the use of regional skew to stabilize site-specific skew estimates being a first step.

ROBUST HYDROLOGIC MODELS

The Rainfall-Runoff Relationship

The objective of hydrologic science is prediction of the time trace of relevant hydrologic variables. Suppose a basin, a convenient hydrologic quantum, is "described" by a set of parameters θ; if we are concerned with a rainfall-runoff relationship then one set of elements θ would be appropriate, whereas another set θ^* would be appropriate for studying migration of pollutants within the vadose zone. There is a subset of elements that would appear in most arrays of parameters. These represent the minimal set of information required to characterize the basin for almost any analysis. We assume that θ is independent of the state variables of the system.

A general rainfall-runoff model is written

$$q + f(x, S|\theta),$$

where the scalar q is the basin runoff, x is a vector-valued description of the driving function (including, for example, precipitation intensity, duration, coverage, and direction of storm movement), S is a vector of state variables, and the notation $f(x|\theta)$ represents a vector function for a basin characterized by θ (the time trace of the runoff q associated with x). Specification of the functional form f and characterization by the vector comprise an unresolved problem in hydrology. Among the unsettling but answered questions are the following:

1. What is the appropriate function f?
2. Which variables correctly characterize the driving function x and the state variables S?
3. What parameter assignments θ should be made? (This encompasses two different questions, the first of which deals with identifying the parameters that comprise the vector θ and the second with efficient techniques for estimating these parameters.)
4. How do the vectors x and θ shift with time, and how do they interact? That is, could slow-moving changes in the local climate induce significant changes in the parameterization θ or in the functional representation f? Do the $x(t)$ and $\theta(t)$ cycle?
5. How do the vectors x, S, and θ, and the function f, respond to the works of man? The stimuli are more abrupt than natural stresses; hence, the responses are likely to be qualitatively different.

Many efforts have been made to resolve these questions; the literature is replete with rainfall-runoff models and with comparisons among alternative models. Important recent works are those edited by Chapman and Dunin (1975) and Biswas (1976). Both studies review the role of prediction in watershed management, with special emphasis on catchment models. Both point up the intellectual conflicts that arise from alternative modes of describing and explaining hydrologic phenomena and thus are strongly parallel to this essay. Our work is a further attempt to bridge some of the gaps identified in these earlier efforts. The results of this modeling have been rather embarrassing, not so much because we have been unable to develop good fits but, in a perverse way, precisely because *we have been successful* in fitting functions to data. The "best" set of estimates can be derived for any given application. For example, by imposing some criterion such as minimal residual squared error, parameter estimates $\hat{\theta}$ can be deduced for many reasonable functional forms. If the fit is good the resulting relationship might become widely accepted, but its indiscriminate use can lead to serious errors in prediction.

At the simplest statistical level, errors can be laid to the paucity of historical records. Errors are also introduced by improper specification of the list of driving variables x. Further, the tacit assumption of stationarity of parameters pre-empts dynamic interactions in many rainfall-runoff models. Thus, as we contemplate increasingly important decisions occasioned by growing demands for water use to serve an expanding population aspiring to a higher standard of living, we exceed the conditions under which conventional rainfall-runoff models are calibrated. Man's interventions threaten to alter the basic mechanisms and hence those relationships that are founded in hard science. As developmental lead times increase--i.e., as we contemplate hydrologic consequences that might not occur for several decades or millenia--even modest inaccuracies or imprecisions in prediction can be levered into important errors that must not be overlooked. We must demand from our models better performance than we now get.

Virtually all the basin models currently in use concentrate on the relationship between rainfall and runoff. But runoff accounts for only about one third of the throughput or flux in a catchment; the remainder is treated as a residual. It is asking too much of a basin model to predict reliably beyond the range of its observation set, particularly when two thirds of the throughput is not included among the outputs; the largest fluxes are ignored. A small percentage change in this large residual, caused by a natural or man-induced change in one or more of the other fluxes, is levered by deterministic continuity constraints into a large (and unexplained) change in the runoff. This results in weak predictive ability of the relation when it is applied to those new combinations of x and $\hat{\theta}$ that might occur as development proceeds. The bulk of the throughput is not accounted for by the model and consequently does not contribute to the fitting or validating procedures.

It is tempting to call for direct assessment of evaporation, hoping thereby to reduce the errors in that important flux and hence in the rest of the model. But techniques for direct assessment are currently so noisy that it might not be advantageous to build into the model that extra bit of structure required to accommodate available evaporation data. It is not surprising that models fail when generalized to new storm events, new locales, or future events.

Efforts to handle this shortcoming have produced two kinds of rainfall-runoff relationships, neither totally satisfactory for the predictive needs of hydrologists. The first is a detailed relationship involving many state variables and connections to express continuity among storages or boxes. It is a tractable mathematical exercise to examine inputs (rainfall) and outputs (runoff or discharge) and to select the best set of parameters, where "best" implies some criterion of fit between observed and calculated outflow hydrographs. A major shortcoming of the analysis is that the derived set is not necessarily unique. It is virtually impossible on the basis of the observations and hydrologic experience to promote one parameter set over another. The dilemma is that the use of simple causal models with lumped parameters likewise does not guarantee a unique set of basin parameters. In either case the reliability of derived outflow hydrographs is subject to serious question when driven by new precipitation values.

Another way of dealing with hydrologic prediction is to rely on statistical curve-fitting techniques, including error terms that bound the anticipated sampling fluctuation. But again, when basin descriptors are not systematically incorporated, it is impossible for the model to reliably *predict* the runoff consequences of natural or man-made changes in the basin; the black box model simply does not allow for this level of manipulation.

Both extremes could be supplanted by a new class of rainfall-runoff relationships that utilize the best features of each. Recalling that runoff accounts for only about one third of the total throughput, it is reasonable to seek a model of the following form:

$$y = f(x, S|\theta),$$

where the output y represents a stochastic process or vector of time series, and f denotes a functional form. The symbol f incorporates several simultaneous relationships for the elements of y in terms of x and S for a given θ. The exact form of the relationship(s) is not so important as the concept that a better, more stable, more robust relationship can be developed if just the right amount of "structure" is used or if *several* of the "flows" in a rainfall-runoff formulation are modeled *simultaneously*. We traditionally focus on rainfall and runoff alone; modern theory and measurement capability together dictate that more comprehensive basin budgets should be developed. It is essential to collect measurements on selected fluxes and state variables--perhaps three or four would suffice--to write the resulting continuity and transfer equation, and then simultaneously to solve for the parameter set $\hat{\theta}$. These robust estimates of system behavior define a model that is no longer properly called a rainfall-runoff model because the parameters are jointly tuned to rainfall, runoff, and *other* system fluxes or state variables.

Details of a Model

A model currently being developed represents our perception of how to utilize the large amounts of hydrologic data soon to become available. We propose such a model as a framework for new research in the next decade.

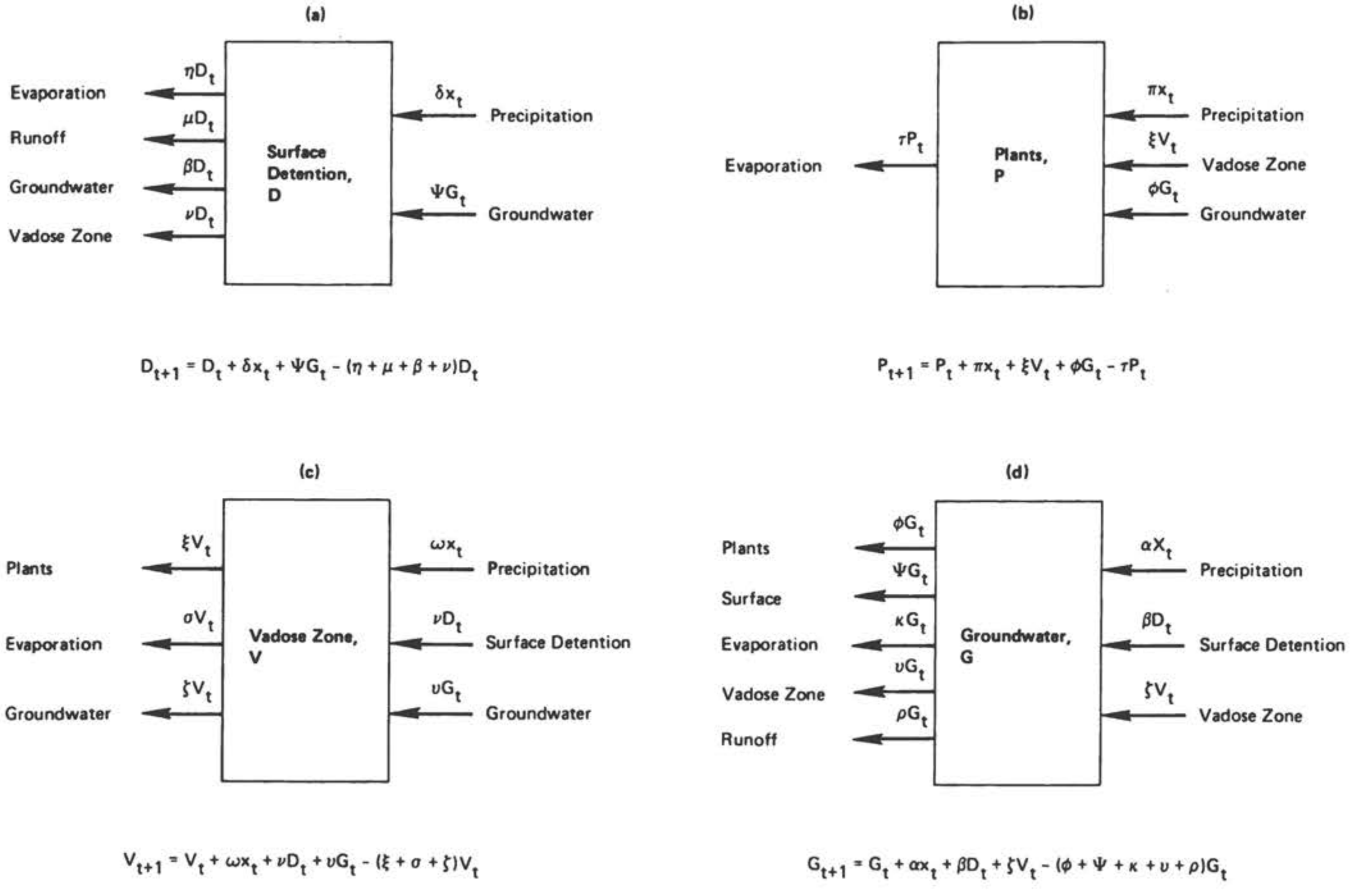
The model has four storage compartments representing the surface detention, plant storage, vadose zone, and groundwater aquifer. Figures 7.1a through 7.1d show the inputs and outputs associated with each of the four compartments. The principal driving quantity is x_t, the precipitation during the time period t. In this version of the model we ignore regional groundwater throughput, the application of irrigation, inputs for snowmelt, and other sources that might be elements of the driving vector x. Thus, this model is not so much a water budget for the region as a formalism that describes the distribution of the water delivered in a precipitation event.

Figure 7.2 shows the connections among the four compartments that comprise the model. The distribution of input is shown on the right; the precipitation value in the time period t is divided into direct evaporation, direct loss to surface channels, direct overland flow, and an input component to each of the four compartments. Evaporation output, shown on the left side of the figure, is derived from precipitation and directly from each of the compartments. The runoff component of output consists of direct channel and overland flow, both of which are linearly related to the precipitation x_t and to the surface detention and groundwater compartments. Flows within the system are shown as line segments directed from a source to a sink; many potential connections are ignored because they are presumed to be small or to have no hydrologic meaning. An important feature of the model is the assumption that all directed line segments or fluxes are linear in the driving quantity (e.g., the input precipitation x_t or the current storage at the start of the time period t in any of the four compartments). It is also assumed that the storage capacities of these compartments are never exceeded and that they never become negative.

Continuity equations for each compartment are summarized in Figure 7.1. Along with these equations there are constraints that bound the parameters. The surface runoff is given by the deterministic continuity equation

$$q_t = x_t\ (\gamma + \lambda) + \mu D_t + \rho G_t .$$

Analysis of the equation is based on the premise that the runoff q_t; precipitation s_t;



$$D_{t+1} = D_t + \delta x_t + \Psi G_t - (\eta + \mu + \beta + \nu) D_t$$

$$P_{t+1} = P_t + \pi x_t + \xi V_t + \phi G_t - \tau P_t$$

$$V_{t+1} = V_t + \omega x_t + \nu D_t + \upsilon G_t - (\xi + \sigma + \zeta) V_t$$

$$G_{t+1} = G_t + \alpha x_t + \beta D_t + \zeta V_t - (\phi + \Psi + \kappa + \upsilon + \rho) G_t$$

FIGURE 7.1 Components of a basin model.

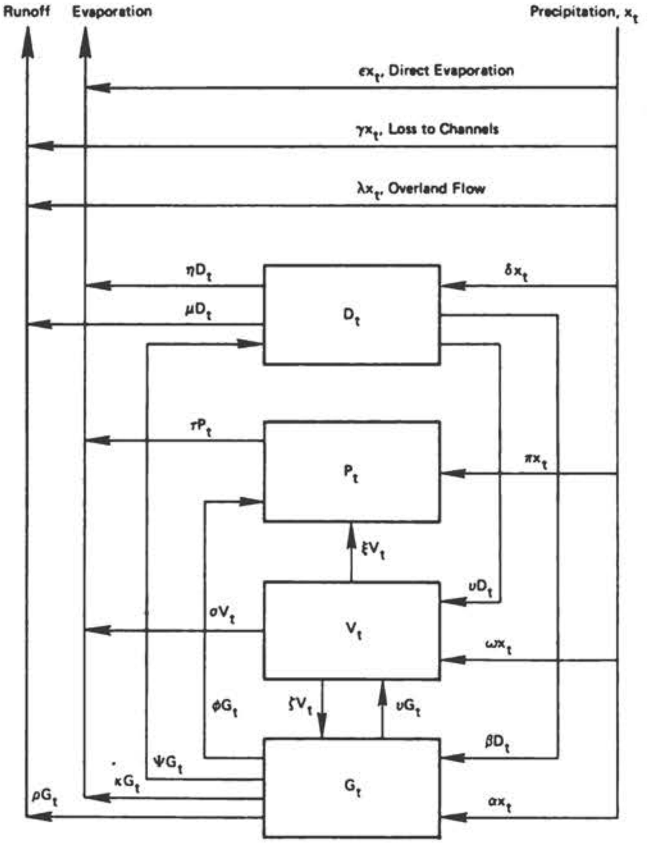


FIGURE 7.2 Connections among the compartments.

and the state variables of each compartment D_t, P_t, V_t, and G_t can be measured, directly or remotely. The continuity equations as written are deterministic, but in reality perfect closure will not result because of random perturbations. The system thus requires the addition of a random disturbance term that reflects spatial heterogeneity, noise in measurement processes, and misspecification of state variables and fluxes.

The parameters can be estimated by using several techniques, among them constrained least squares. In addition there is a conservation constraint,

$$\varepsilon + (\gamma + \lambda) + \delta + \pi + \omega + \kappa = 1,$$

in which the sum ($\gamma + \lambda$) and all subsequent terms are known, imputing a value for ε. Of all 20 parameters, 15 are known directly from fitting to the observations; 3 are calculated by continuity; and the single indeterminacy $(\gamma + \lambda)$ remains. This felicitous result rests on the assumption of linear system rules that dictates an operating policy different from that of most real systems. The set of parameters may have some persistent lumpiness in that modifications of the basic model might produce combinations of parameters for which a completely unique assignment cannot be made, resulting in some aggregated parameter values. Even if these are tightly estimated it is impossible to know *a priori* if equally good estimates can be ascribed to the constituent values; this will have to await empirical validation. This and other shortcomings aside, it is useful to speculate on some inferences.

The principal advantage of this approach is that we fit the parameters to, and constrain them by, an array of time series that is larger than usual; this reduces the nonuniqueness in parameter estimation. We have a rainfall-runoff relationship whose form is not unlike conventional models, but whose parameters are tuned to measurements of more series of concurrent observations than is commonly the case. The set of parameter values can be fitted by independent or coupled applications of least squares, but in either case we can improve the robustness of the model by imposing tighter constraints on those parameters whose values are known to lie within specified ranges. Some of the stocks may be known to be approximately zero in some basins. We then solve for a set of consistent parameter estimates that can represent the impact of the works of man. It should be possible to find a suitably large number of sites at which observations could be used to formulate redistribution rules conditioned on basin characteristics. For example, a particular class of basins might be characterized by redistribution primarily at the expense of evaporation and another at the expense of direct infiltration. This type of basin toxonomy might be adequate to yield appropriate precision for the estimated runoff. The consequences of such parameter modification--whether reflective of natural or man-made influences--are perceived *throughout* the system and can be used to assess the capability of the basin to meet a variety of demands imposed over time.

System response is defined by the five concurrent traces (runoff, surface detention, plant storage, vadose storage, and groundwater storage) that result from passing sequences of precipitation events through the model.

Sources of Error

There are many sources of uncertainty and imprecision in the proposed formulation. First, the model might contain the wrong combination of compartments and vectors. Second, even if the correct combination were included, some of the formulations for flux rates are certain to be *nonlinear*; for example, the natural operating policy for the basin might be better represented by logistic functions. Third, ordinary *statistical error* dictates that parameter estimates are subject to anticipated sampling fluctuations (temporal error). Fourth, the data are taken at several places within the basin and are manipulated to construct a single value representative of the basin. *Network techniques* have been suggested for taking basinwide averages. Fifth, the data are subject to *measurement* errors.

Assessment of Water Resources

It is difficult to append to a conventional model those auxiliary computations in the proposed model that quantitatively assess the water resources of a basin. For example, ecosystem health, agricultural productivity, and the suitability of the region for various levels of development depend critically on the projected traces of the stocks or storages. Conventional runoff models do not provide an index of basin assessment. The proposed model makes explicit the states of several stocks within the catchment and thereby allows a variety of impacts to be assessed by applying different transformations to each of the storages; we can develop an index that incorporates dependent outputs in imaginative and innovative ways. For example, Thomas and Smith (1977) suggest a geographical diversity index unambiguously to assess the ecological impact of a major hydraulic project. Generalization of this methodology to the systematic assessment of similar basin impacts and outflows could lead to an unambiguous ranking or index of basin status that would then be incorporated in a national assessment program.

PARAMETER ESTIMATION IN WATERSHED MODELS

Philip (1975) pointed out that although modern attempts to understand hydrologic processes based on the established and proven methods of natural science have been successful, these studies have been restricted to "small scale local processes." Synthesis of such local processes to model large and spatially heterogeneous basins presents formidable tasks of integration, system characterization, and data collection. Because many hydrologic observations represent integrated outputs from a spatially heterogeneous system, Philip notes that "the process of inferring the spatial (or sometimes temporal) variation of the parameters of the system from the (integrated) output runs into the problem that any one of a large (even an infinite) number of assumed modes of variation may yield approximately the same output."

For this reason and also on account of the enormity of the data-collection program necessary for a successful synthesis, the concept of lumped or effective parameters naturally arises. Because these parameters are no longer physically measurable, use of statistical (or curve-fitting) parameter estimation techniques is inescapable. Consequently, the existence of a well-posed parameter estimation problem is a necessity for the watershed model to be robust.

Recent work has identified reasons why the parameter estimation problem may be poorly based. In such cases the high variance and arbitrary bias in parameter estimates lead to questionable relationships between model parameters and measured watershed characteristics and also undermines the validity of representing land-use changes by altering model parameters. Despite poor parameter precision, the calibrated watershed model can furnish an embarrassingly good fit to observed runoff.

Mein and Brown (1978) studied the covariance structure of the optimized parameters of a particular watershed model and suggested that the high variance associated with parameter estimates arises in part from a model structure burdened by redundant parameters. This concept of redundance is analogous to near-multicollinearity in linear regression that arises when near-linear dependence exists between one or more pairs of explanatory variables. Put another way, a least-squares estimation procedure that minimizes the sum of squared differences between predicted and observed runoff cannot be expected precisely when different configurations of parameters produce nearly identical predicted runoffs; it is asking too much of the mathematics.

Kuczera (1980) demonstrated that the use of least squares can lead to arbitrarily biased parameter estimates in lumped watershed models that explicitly contain *unobserved* storage state variables such as groundwater volume or soil moisture. This is because watershed models are formulated in terms of continuity equations (a runoff and several storage mass balances) that must be closed by random disturbances on account of misspecification of state variables, fluxes, and measurement errors. In using least squares to estimate parameters without observations on storage state variables, it is implicitly assumed that random disturbances exist only in the runoff continuity equation. Such limited data parameter estimates are biased in the sense that, if observations on storage state variables were available and employed in a consistent parameter estimation procedure, the difference between the limited data and full data parameter estimates is most likely to be statistically significant.

Model structure is obviously desirable as it enhances our confidence in the model's predictive ability. However, excess structure may demand data that is unavailable to us and so bias parameter redundancy from the viewpoint of the parameter estimation technique and so induce unstable parameter estimates. The analyst undertaking construction of a lumped watershed model should be aware of the limitations that available data and parameter estimation techniques place on model structure; otherwise, model robustness may be jeopardized.

CONCLUSIONS

The problem of how much structure to add to simplistic models (whether causal or associative) is central. In a simple case, with two inputs x_i and two outputs y_j, we can write a set of equations

$$y_1 = a_{10} + a_{11}x_1 + a_{12}x_2,$$

$$y_2 + a_{20} + a_{21}x_1 + a_{22}x_2,$$

and estimate the parameters, hoping to predict y_j closely for a new set of x_i. But if model specification is erroneous, and only some of the driving variables are included, we might have,

$$y_1 = b_{10} + b_{11}x_1,$$

$$y_2 = b_{20} + b_{22}x_2,$$

and attempt to "control" y_j over the long run by controlling x_i according to the coefficients b_{ij}. This is an attractive strategy if the observed correlation between (x_1, y_1) and x_2, y_2) is high, but might not be cost effective or even reasonable because of the correlation between x_1 and x_2. Thus, controlling y_2 by x_2 alone might introduce counterproductive results in terms of y_1 if x_2 is not included in the model for y_1.

This can be overcome, of course, by knowing the causal model--recourse to first principles that are presumed to be inviolate and stationary over long periods of time. But the operational question must be predicted on the notion that full knowledge of the relationships, full axiomatic understanding, is out of the question for the near future; how much structure, how many variables, how high a correlation should we demand in order to make predictions?

In consideration of these questions, we conclude the following:

1. Constrained and joint least squares and related estimating techniques should be utilized on multivariate data traces to account for systems stores as well as the more customary fluxes. This builds more structure into models, but does not necessitate full understanding of their underlying principles.
2. James-Stein and other techniques should be adopted for regionalization of extreme-event analysis.
3. Special attention should be directed at hydrologic models as the linkage between local phenomena (impacts) and global processes (circulation).
4. Monitoring of long-term environmental consequences will enhance the predictive ability of our models and allow for timely societal adaptation to unanticipated events.
5. It is not always necessary to collect new data to formulate and calibrate hydrologic models. New modeling techniques and consequent innovative uses of existing data bases might suffice to develop rational assessments and policies.
6. It is essential to have a clear notion of the free-body cut--of where, how, and when the system is impacted by exogeneous events. This is a meta-question, whereas the issue of parameter estimation is at a more local level. Nonetheless, they are in the same spirit.

REFERENCES

Biswas, A., ed. (1976). *Systems Approach to Watershed Management*, McGraw-Hill, New York.

Chapman, T. G., and F. X. Dunin, eds. (1975). *Prediction in Catchment Hydrology: A National Symposium on Hydrology*, Australian Academy of Science, Canberra.

Effron, B., and C. Morris (1977). Stein's paradox in statistics, *Sci. Am., May 1977*, 119-127.

Fiering, M. B. (1978). Consulting report prepared for the State of New Jersey, unpublished.

Fiering, M. B. (1979). Consulting report prepared for the Miami Conservancy District, unpublished.

Fiering, M. B. (1982a). Alternatives of resilience, *Water Resour. Res. 18*.

Fiering, M. B. (1982b). A screening model to quantify resilience, *Water Resour. Res. 18*.

Fiering, M. B., and C. S. Holling, (1974). Management and standards for perturbed ecosystems, *Agro-Ecosystems 1*, 301-321.

Greenwood, J. A., *et al.* (1978). Probability Weighted Moments: Definition and Application, *IBM Research Report RC 7108 (30439)*.

Houghton, J. (1977). Robust Estimation of the Frequency of Extreme Events in a Flood Frequency Context, Doctoral Thesis, Harvard University.

Kuczera, G. (1980). Robust Hydrologic Models at the Watershed Scale, Doctoral Thesis, Harvard University.

Mein, R. G., and B. M. Brown (1978). Sensitivity of optimized parameters in watershed models, *Water Resour. Res. 14*.

Mosteller, F., and J. W. Tukey, (1977). *Data Analysis and Regression*, Addison-Wesley Publishing Co., Inc., Reading, Mass.

Philip, J. R. (1975). Some remarks on science and catchment prediction, in *Prediction in Catchment Hydrology: A National Symposium on Hydrology*, T. G. Chapman and F. X. Dunin, eds., Australian Academy of Sciences, Canberra.

Thomas, H. A., Jr., and F. E. Smith (1977). The Cross-Florida Barge Canal: Over-all Assessment, in a Report from Meta Systems, Inc., to the U.S. Army, Corps of Engineers.

U.S. Water Resources Council (1977). *Guidelines to Determining Flood Flow Frequency*, Bulletin 17A of the Hydrology Committee, Washington, D.C.

Empirical and Causal Models in Hydrology

8

V. KLEMES
National Hydrology Research Institute,
Environment Canada

INTRODUCTION

Hydrology is the basic science involved with water-resource management and enters into it through the use of models of hydrologic phenomena in water-resource planning, design, and operation. It is therefore important to examine the nature of hydrologic models and the relevance of their merits and drawbacks to water-management decisions. One of the controversial issues in hydrologic modeling has been the merits and demerits of empirical models as compared to those of causal models (Kartvelishvili, 1967, 1975; Scheidegger, 1970; Mandelbrot, 1970; Klemes, 1970, 1971, 1974, 1978; Yevjevich, 1974; Jackson, 1975; Pilgrim, 1975).

To put this issue into proper perspective one must begin with the fact that while the basis of all our knowledge is, in the last analysis, empirical, the knowledge itself is not merely a sum of empirical facts but emerges from the ability of the human mind to discover relationships among these facts.

"All science is the search for unity in hidden likenesses . . . The progress of science is the discovery at each step of a new order which gives unity to what had long seemed unlike . . . For order does not display itself of itself; if it can be said to be there at all, it is not there for the mere looking . . . order must be discovered and, in a deep sense, it must be created. What we see, as we see it, is mere disorder" (Bronowski, 1972).

The relationships initially discovered are of necessity simple and lead to only limited knowledge and understanding. They generally tell us what change in one observed quantity corresponds to a change in another. Such relationships are commonly labeled "empirical"--they tell us what happens but do not derive the outcome from the dynamic mechanisms governing the process, i.e., from the "necessary relationships between objects, events, conditions, or other things at a given time and those at later times" (Bohm, 1957). Those relationships that are based on the dynamics of the process are commonly labeled "causal"; the road to their discovery is basic research. There is a definite hierarchy in terms of the explicative power of these two kinds of relationships in the sense that a causal relationship gives us more information than an empirical one. But there is no conflict; as long as an empirical relationship is valid, a causal relationship cannot negate it--it can only supplement by indicating, for instance, some limits for the validity of the empirical relation or by pointing to external (and possibly not yet empirically established) factors that may modify it.

Empirical relationships are also used as convenient summaries, or reductions, of results of complex causal chains (e.g., Darcy's law). If such summaries were not possible, every causal model would have to be developed from the absolute "first principles" (i.e., elementary empirical facts and/or axioms) known at the time. Thus, in the physical sciences a causal model would have to have been formulated in terms of molecular interactions a century ago, atomic interactions half a century ago, and interactions among subatomic particles today. Whereas this approach may be seen as a theoretical ideal of causal modeling from the point of view of science in general, it would be of little use to any particular branch of science. In fact, science has become specialized into the innumerable disciplines because of the infeasibility of such an approach. The essence of specialization is to split the (possibly) infinite causal chain into segments, each containing only a few chain links; the scope of one discipline is thus intentionally limited to seeking causal relationships among phenomena within only a relatively small range, whose lower boundary represents the discipline's "first principles" or "scientific basis," coinciding with the discipline's objective.

A discipline seldom considers its own first principles to be an integral part of the discipline itself and tends to view them as being on the other side of the "free-body cut." Accordingly, problems encountered in the first principles are regarded as a hindrance rather than a challenge and lead to an impatient desire to get rid of them rapidly without much involvement. This is not only convenient but to a great extent necessary: It is the only way things can get done, it is the root of efficiency and productivity--unless we draw the line somewhere we are bound to drift along the causal chain of things indefinitely without being able to strengthen any of its links. On the other hand, to draw the line and take an action always implies a decision based on incomplete knowledge and understanding and therefore is prone to errors and unforeseen consequences. The relationship between the disciplines of water-resource management and hydrology is a typical example of this situation.

HYDROLOGY AS A SCIENCE

The common definition of hydrology as a "science that deals with the processses governing the depletion and replenishment of the water resources of the land areas of the earth, and treats the various phases of the hydrologic cycle" (WMO and UNESCO, 1974) does not convey the true perception of hydrology by the discipline of water-resource management. Here hydrology is perceived as more like a collection of techniques that enables one to make inferences from hydrologic data about the future distribution of water resources in space and time. In other words, the emphasis is not on the study of hydrologic processes and on the understanding of the mechanisms (physical, chemical, and biological) underlying these processes, i.e., on hydrology as a science, but rather on the *prediction of states of hydrologic processes in space and time*. Hydrology is of interest, here, only insofar as it can help in the determination of these future values of hydrologic state variables. If they could be foretold from a crystal ball, hydrology would be of little use to water management. As a matter of fact, much of the hydrologic effort originating in the domain of water-resource management bears a strong resemblance to a search for such a "hydrologic crystal ball." This observation is not meant to have a pejorative connotation; the search for a crystal ball is the implicit ideal of every empirical approach--to find something simple that works. In hydrology the empirical approach has many different labels, such as operational, prescriptive, analytical, and statistical (Klemes, 1978), which all roughly correspond to what in science is more generally known as "reductionism," whose objective is "finding a wonderful new calculus that will break through the barrier of the unknown" (Ziman, 1978). The essentially reductionist approach of water-resource management to hydrology is understandable and, to a great extent, inevitable. As Ziman (1978) observed, "whatever one's philosophical attitude towards reductionism, there is an inescapable scientific necessity of trying to 'understand' and 'explain' the behaviour of any system in terms of a relatively few comprehensible elements without recourse to an elaborate extra-cerebral computation." The purpose here is to make some observations regarding the origin of reductionist pressures in hydrology and their effect on hydrology as a science.

Scientific disciplines usually evolve from the construction of empirical models to the development of causal models. This transition can occur only after the science in question has reached a fairly advanced stage of development (Bohm, 1957). Often, this transition coincides with (and, perhaps, leads to) a break of the developing science away from its parent discipline, with the establishment of this breakaway branch as a new discipline. In this regard, hydrology is a young science compared with its sister sciences, such as meteorology, climatology, geology, and mineralogy. Whereas the latter sciences separated from their parent disciplines (agriculture and mining) a long time ago and have been recognized as sciences in their own right for many decades, hydrology still remains under a strong spell of "hydraulic engineering" for which the term "water-resource management" often is only a more recent (and more ambitious) equivalent. This dependence is reflected in the status of hydrology in universities and in its main sources of research funds. In most universities throughout the world, hydrology is attached to departments of civil engineering where it is usually taught as a sideline by professors of hydraulics or fluid mechanics. Most hydrologic research has traditionally been financed as a part of the planning, design, or operation of specific engineering projects such as dams, flood protection, and navigation schemes.

The fact that hydrologists typically have engineering backgrounds, combined with the usual applied context of hydrologic analyses, tends to reinforce the reductionist bias in hydrology, the trend to find that "wonderful new calculus" that will break through the barrier of the unknown separating raw hydrologic data from information on future values of hydrologic variables. Among the best-known examples of this trend are the rational formula; the unit hydrograph; the search for various correlations, periodicities, and symmetries in hydrologic data; flood-frequency formulas; stochastic operational models; and most recently the transfer-function models.

The strong reductionist bias in hydrology can be seen not only in the proliferation of the empirical models but also in the approach to causal modeling. This perhaps is best evident in the development of the so-called conceptual hydrologic models aimed at incorporating the general pattern of physical mechanisms governing hydrologic processes. More effort is spent on trying to determine the properties of the individual links in the causal chain (the "conceptual boxes") by optimizing the fit of the model output to an observed output than is spent on the study of the physical phenomena involved. The rather low popularity of the latter line of work may seem surprising because most hydrologists would agree with the definition of hydrology cited at the beginning of this section, which explicitly points in this direction. However, when viewed in the historic perspective discussed above the situation is understandable.

A third factor has contributed to this state of affairs in recent years: the computer. It has made the pursuit of finding the "wonderful new calculus" (or, more specifically, the "perfect transfer function") much less demanding, more publication productive, and therefore more attractive and even more prestigious than basic research in hydrologic processes. It now seems clear that the computer has done a disservice to many branches of science by diverting some of the best talent into pursuing purely computational problems of little relevance to the given science. As Fiering (1976) puts it, "Fascination with automatic computation has encouraged a new set of mathematical formalisms simply because they now can be computed; we have not often enough asked ourselves whether they ought to be computed or whether they make any difference. . . ."

Having reminded the reader that contemporary hydrology has a strong reductionist bias and having stated the main reasons, one can now ask how this affects the science of hydrology and its applications to water-resource management. Before trying to indicate some answers, the two main sources of the difficulties arising in providing information on future states of hydrologic processes--the service expected from hydrology by water-resource management--are addressed below.

The principal source of difficulties is the extreme variety, variability, and complexity of processes that affect hydrologic phenomena (Chapter 2). The different temporal and spatial scales of these processes, their direct and indirect interactions, and inherent instabilities lead to great irregularities in the fluctuation of hydrologic processes in time. These irregularities manifest themselves as noise and make prediction of future states of a process a difficult problem. The situation is aggravated by the fact that reliable hydrologic records are usually relatively short and thus grossly inadequate for making inferences about long-term future behavior of hydrologic processes. Examples of temporal (and spatial) variability of streamflow are shown in Figure 8.1.

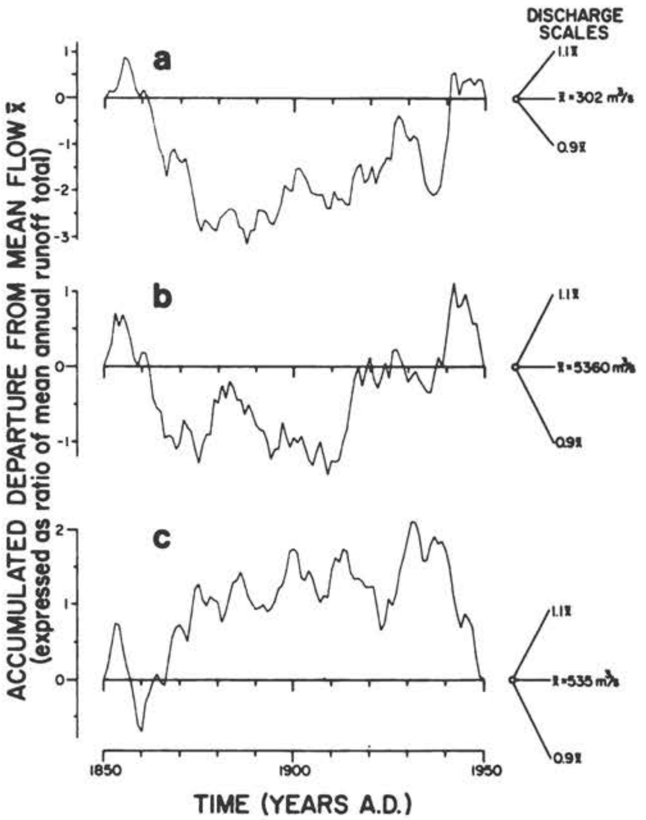


FIGURE 8.1 Accumulated departures from 100-yr mean flow *x* (1851-1950) for three European rivers (mean flow for any interviewing period is given by the slope of a straight line connecting its endpoints projected on the curve; it can be measured by the discharge scale shown). a, Elbe, Decin, Czechoslovakia; b, Danube, Orshava, Romania; c, Gota, Sjotorp Vanersbury, Sweden (Data sources: a, Novotny, 1963; b and c, Yevjevich, 1963).

The second major source of difficulties is that most hydrologic data pertain to variables at a point, whereas the hydrologic information desired for water-resource management usually involves variables pertaining either to a large area or to a different point for which data are not available. The difficulty in making this type of inference is proportional to the process spatial variability and irregularity that in turn increase with spatial heterogeneity of the environment in which the process evolves. Unfortunately, this spatial heterogeneity of physical variables affecting hydrologic processes is extremely high. Examples of spatial (and temporal) variability of point precipitation are shown in Figure 8.2.

MERITS AND DEMERITS OF EMPIRICAL MODELING

The main merits of empirical models are (1) the possibility of developing them without much understanding of the modeled phenomenon, (2) their simplicity achieved by short-circuiting complex causal chains, and, as a result of these two features, (3) their potential for making the collected data useable without much delay and hence their promise of high cost effectiveness of the modeling exercise.

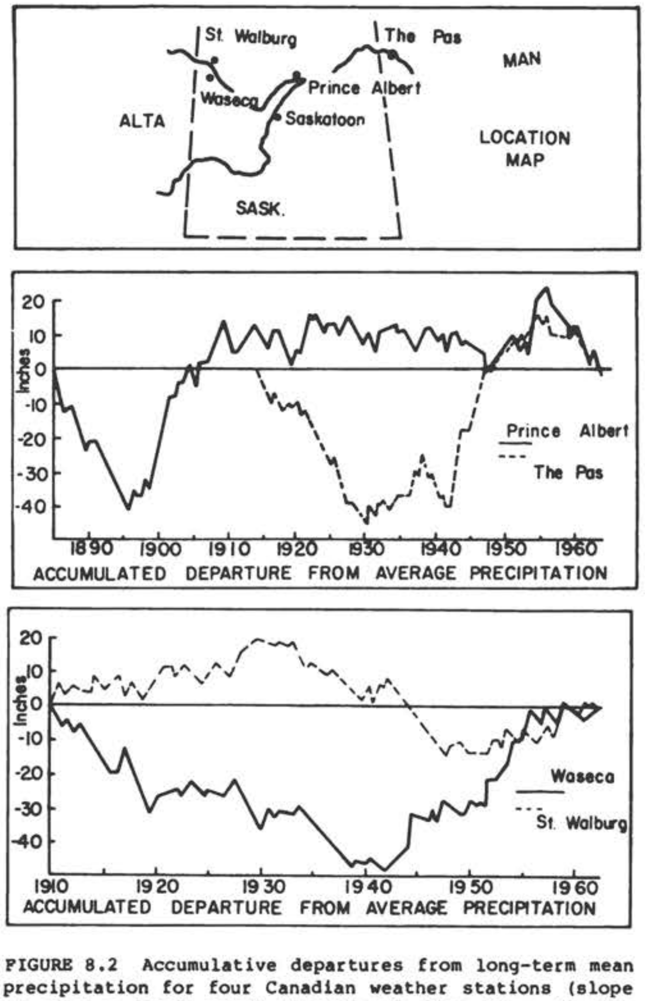


FIGURE 8.2 Accumulative departures from long-term mean precipitation for four Canadian weather stations (slope of plot = local mean) (reproduced from Water Studies Institute Report No. 2, Oct. 1965, Saskatoon, Saskatchewan, Canada, as cited in NRC Geophysics Study Committee, 1977).

The drawbacks follow naturally from the above merits but need more elaboration.

1. Because of the extreme complexity of the environment in which hydrologic processes evolve and the fragmentary (both time and space-wise) information that hydrologic data normally contain, empirical relationships based on these data usually can give only approximate results. However natural this may be, the uncertainty in the result causes displeasure to the user, who then exerts pressure on the modeler to "improve" the model. If the door to more information (in terms of more data and better understanding of the process) is closed, the only choice available to the modeler is to try to extract more information from the data at hand using some "better calculus." While sound in principle, this course of action faces several dangers.

One of them is overfitting, which amounts to regarding part of the noise in the data as information. This error is easy to commit because the demarcation line between noise and information is often blurred in the data. A well-known example from the not-so-distant past is the various harmonic analyses of hydrologic series. Their claims of discoveries in most historic records of periods with wavelengths other than those corresponding to the astronomic cycles of the earth and ranging from a few months to many decades have never been substantiated by more recent data.

Another is the danger of sidetracking into polishing some aspects of the modeling methodology that, while perhaps of value for the methodology itself, are unimportant in the context of the particular application. An outstanding example is the history of flood-frequency analysis, where much progress has been made in computation of plotting positions, in (efficient, unbiased, and consistent) estimation of population parameters, in treatment of outliers, and in goodness-of-fit testing. On the other hand, the distribution type is chosen arbitrarily, the calendar boundaries of the year are chosen arbitrarily, the typical nonhomogeneity of the flood-causing factors (some floods are caused by snowmelt runoff, some by convective storms, some by frontal storms, and some perhaps by hurricanes) and the influence of the changing character of the basin are ignored, the presumed mutual independence of flood events is usually not checked, the difference in the hydraulic behavior of the river under relatively small and extremely large floods is not considered, different hydraulic conditions along the river channel are dismissed, and the high uncertainty in the reported values of peak discharges (which are seldom directly measured) is not taken into account. When examining the mainstream of the flood-frequency literature from this point of view, one wonders what there is to be learned about the frequency of floods as opposed to the art of distribution fitting to samples of exact numbers drawn by random mechanisms from homogeneous infinite populations.

Another danger originating in the desire to improve a model without additional information on the process is to elaborate formally some aspect of the model known to be wrong and present merely for operational reasons with the understanding that the implied error "does not matter" in the given context. The case in point is, for example, the "baseflow separation" problem in the unit hydrograph technique. The concept of separate direct runoff and baseflow hydrographs is hydrologically doubtful but is convenient for estimating the water balance of an isolated flood event as long as the estimate does not have to be too accurate. The separation line is an operational device that has little, if any, hydrologic significance, and the great efforts that have been spent on the refinement of its shape have not improved unit hydrograph modeling at all.

In summary, the striving for "improvement" of an empirical model in the absence of additional information tends to be scientifically sterile and to have an extremely low benefit/cost ratio from the point of view of applications.

2. Empirical models must be regarded essentially as interpolation formulas. They have no justification outside the range of the underlying data sets and their use for extrapolation involves risk of large errors. This aspect is of prime importance in water-resource management because the hydrologic information required for most of its applications involves extensive extrapolations in space, in time, and in the range of the variable concerned (extreme values). An example of a possible consequence of such extrapolation (in time, in this case) for water-resource management can be drawn from Figure 8.1. Extrapolation of mean flow computed from one period of the historic record could lead to a serious error when extrapolated into the following period.

3. A great handicap of empirical modeling is the uncertainty about the adopted model structure. In the absence of theoretical (physically based) reasons for a specific structure, auxiliary (and, on the whole, subjective) criteria for model selection must be adopted. Mathematical convenience is a popular refuge, current fashion running a close second. The inherent danger is that the high degree of arbitrariness in the model form is either not realized or the awareness of it gradually fades away and the assumptions are treated as facts of life and later used as unalterable building blocks in other models.

A typical example is the use of long-term means in hydrology. The arithmetic mean is one measure of central tendency of a process attribute whose central tendency *can be expected* to remain constant as the process develops in time. Central to the usefulness of the long-term mean concept is the *justifiability of the expectation of the constancy* of this central tendency. This justifiability is by no means universal and depends on the nature of the process. If we are concerned, for example, with the central tendency of the amount of beer in bottles coming from a specific, well-tested bottling machine, then the expectation of constancy is justified. If, however, we deal with the amount of precipitation or runoff in individual years this may not be the case as Figures 8.1 and 8.2 indicate. We have become greatly concerned about the apparent lack of this constancy (as our increasing interest in climatic change indicates) but, nevertheless, we continue to regard long-term means of precipitation and runoff as valid concepts and use them as the basis of many hydrologic models. It does not seem to matter much that we then arrive at model structures that are hydrologically inexplicable; our hydrologic (and water-management) conscience apparently can accept this situation more easily than it can the idea of re-examining the arbitrary and hydrologically irrelevant concept of constant central tendency that has been deeply ingrained into our minds by the never ending exposure to superficial interpretations of statistics. We do not seem to realize that, from the empirical modeling viewpoint, the mean of a time series is merely one of many fits by a horizontal straight line. Why a geophysical series should be fitted by a straight line, in particular by a horizontal one that minimizes the sum of squares of vertical deviations, is difficult to see (Klemes, 1974).

4. The essential arbitrariness in the selection of the form of an empirical model exposes the modeler to the danger of adopting a basically wrong modeling approach as a result of its success in some other situation and because of some superficial similarities between the modeling problem at hand and a problem where the model proved successful. This seems to be the case with much of the applications of classical statistics in hydrology and consequently with much of the applications of statistical hydrology in water-resource management. Statistical methods in current use have been designed for the analysis of large masses of data from repeatable and controlled experiments. In empirical hydrologic modeling they are applied to hopelessly small samples generated by "unique and uncontrolled experiments," as hydrologic and other geophysical processes can be characterized. With regard to the mathematical concept of stochastic process that we routinely invoke, a given historical series is a sample of size one. It also is a sample of size one for many planning and design purposes, such as the estimation of reservoir sizes or mean flows for the economic life of a project. Yet we are arrogant enough to use this single measurement to construct distributions, "estimate probabilities," and do many other things that must make Richard von Mises turn over in his grave ("First the collective then the probability," von Mises, 1957).

5. It is generally acknowledged in the

philosophy of science as well as in science itself that the fundamental reason for empirical modeling is the availability of data combined with a lack of understanding of the relationships among the phenomena they describe. This combination is conducive to the adoption of a "let-the-data-speak-for-themselves" philosophy that has many supporters in hydrology and, which, through the development of black-box (transfer-function) modeling, has become a kind of ideology for many hydrologists. This lack of understanding of the process mechanisms, originally regarded as regretable, has been transformed into a virtue--questions about the internal workings of the system being modeled are excluded by design. This philosophy, most prominent in stochastic hydrologic analysis, has two distinct aspects.

The first is its failure to recognize that what makes the data speak is the context; without it their numerical values have little to tell and what they do tell is often misleading. For instance, it would be naive to automatically identify the mean of a series of numbers with a reasonable estimate of the mean of the variable whose states they represent. And it would be even more naive to hope that the estimate can be improved by polishing the formula for the mean. An intelligent inference about the mean of the *physical variable* concerned, e.g., a distance, can only be made by investigating *physical problems* of the following kind: Do the numbers represent measurements of the same distance or of different distances? Were they obtained by the same instrument, same person, under the same conditions? Do they represent distances between stationary or moving objects (moving with uniform or nonuniform motion)? Does it at all make sense to talk about a 'mean distance' in the given case? It is important to realize that these problems are not reducible to problems of blind-mathematical manipulation of numbers.

The second aspect is more subtle and has a Machiavellian flavor. The untenability of the "speak-for-yourself" attitude to data has often been exposed not only in the hydrologic context (Fiering, 1967; Klemes, 1971; Kartvelishvili, 1975) but also in statistical literature by such eminent scientists as Norbert Wiener, M. G. Kendall, A. Stuart, J. Neyman, and M. S. Bartlett (Klemes, 1978). It is therefore difficult to believe that the belief in this philosophy, professed by many hydrologists, is sincere. A more likely explanation is that it serves merely as a convenient and, it is hoped, dignified cover for the hydrologist's reluctance to admit his despair face to face with the enormous complexity of hydrologic processes (Karteveli-shvili, 1975, suggested that the development of an adequate causal theory of hydrologic processes may be much more demanding than was the development of the theory of relativity or the quantum theory), his lack of *hydrologic* ideas, and his overabundance of *computing* ideas combined with his good formula-manipulative skills (a natural consequence of the systems-analytical bias prevalent in graduate hydrologic programs in many, if not most, leading universities in recent years), his compliance with the merciless rules of the publish-or-perish modus operandi of contemporary science, his preference for the cozy atmosphere of his office over the inconveniences of field research, and, last but not least, his repeated failures to secure resources for long-term research of "merely academic" interest with little promise of immediate applicability, and his surrender to pressures for fast results. Many a bird can be killed with the stone of black-box modeling, and "occasionally words must serve to veil the facts. But this must happen in such a way that no one becomes aware of it; if it should be noticed, excuses must be at hand, to be produced immediately" (Machiavelli in "Instructions to Raffaelo Girolami"). Herein may well be the greatest disbenefit of empirical modeling to hydrology.

To summarize, empirical models are useful (and, indeed, indispensible as starting points in cases of insufficient data and/or understanding) as long as they are not mathematically strained beyond their intrinsically limited "carrying capacity."

WHY DO HYDROLOGIC MODELS WORK?

It can be argued that if hydrologic modeling has a strong reductionist bias and if this bias has so many dangerous consequences as claimed in the preceding section, then the performance of hydrologic models could not be as good as it appears to be. For, as many would testify, hydrologic models have been used successfully in innumerable instances in both design and operation of water-resource engineering projects.

There are a number of reasons why hydrologic models work or seem to work. The most common are listed below, approximately in decreasing order of occurrence:

1. *Model Is Empirical and Works Well As an Interpolation Formula* An empirical model is a formal statement of observed facts and thus has to give good results, i.e., results within the observational accuracy and compatible with the level of observed noise, within the range of the observations. Inside this domain its good performance is conditional on the invariance with time of the conditions reflected by the variables involved. All regression models fall into this category. A typical example is the stage-discharge relationship for a river channel cross section. While true in the above sense, its good performance hinges on the stability of river morphology in time and is limited to the range of river stages and surface slopes for which the measurements of flow velocity and cross sectional area were actually carried out. The reliability of the relationship beyond the range of the measurements cannot be deduced from the quality of its fit within that range.

2. *Model Works Well Because It Portrays Only a Small and Relatively Well-Understood Segment*

(or Component) of the Hydrologic Cycle The most typical cases are those in which the hydrologic process is locally (spatially and temporally) dominated by (i.e., can be approximately reduced to) processes of hydraulics for which relatively good causal models exist. Thus it is a hydraulic rather than hydrologic model that works. Examples include flood-routing models, flow on a hillslope, the latest addition being the various "urban-hydrology" models whose success is proportional to the degree to which the urban catchment consists of impervious surfaces, sewer pipes, and well-defined prismatic channels (hydraulic elements).

3. *Model Is Essentially Untestable and Its Good Performance Is a Matter of Faith* Into this category fall most models intended to portray aspects of the long-term behavior of hydrologic processes, especially if their characteristics have been conservatively estimated. For example, few hydrologic records will ever be long enough to make possible a conclusive refutation of the correctness of the magnitude of a 50-, 100-, or 1000-yr flood if this magnitude is inflated; no single reservoir will operate long enough under the design-release policy to enable the analyst to prove that the actual risk of failure was less than the design value of, say, 2 percent; if the value of a "probable maximum" precipitation or flood is set by an order of magnitude higher than any observed event, then its correctness is beyond reproach because of the vagueness of the definition--if a higher flood occurs it can always be classified as an "improbable event" (everybody knows that improbable events do occur--people do win millions in lotteries).

4. *Results Obtained by a Good Economic Decision Model Reflect Favorably on the Quality of a Hydrologic Model Embedded in It* For example, Slack *et al.* (1975) state ". . . the use of the normal distribution to represent the distribution of floods is generally better than either the Gumbel, lognormal, or Weibull distributions. Nothing is gained in terms of reducing expected opportunity design losses if the underlying distribution . . . is identified over and above simply using the normal as the assumed distribution." Here the first sentence may convey a wrong impression that the normal distribution is a *better hydrologic model* for flood peaks (in the sense that it provides better estimates of flood frequencies in specific cases), whereas what it really means is that it is *better in the given decision context* as is obvious from the second sentence of the quotation.

5. *Model Is Largely Irrelevant to Results Obtained with Its Aid* Claims of good performance of a specific hydrologic model are sometimes based on results that, while obtained with the aid of the model, are irrelevant to its structure and some parameters and depend on circumstances external to the model. For example, it can be claimed that a given stochastic model represents an historical flow record well because it leads to an optimum reservoir-operating policy that is essentially the same as the policy that actually would have been optimal during the period of record. However, because the optimal policy for a typical economic-loss function depends mainly on the mean inflow and is largely invariant with regard to the inflow model structure (Klemes, 1977), the claim of the good performance of the model is misleading--most models would do as long as they have the same mean and loss function (Jettmar and Young, 1975).

6. *Empirical Model Has a Form That, Without Conscious Effort of the Modeler, Happens to Describe Some Essential Aspect of the Physical Mechanism of the System* Such models do not occur frequently, but when they do they tend to acquire great popularity because their success is higher than their empirical nature would suggest. A famous example is the unit-hydrograph model. Originally formulated as a purely empirical concept, it has since been shown to represent outflow from a system of linear storages fed by a pulse inflow--a crude but physically sound model for many small catchments. Similarly, the empirically chosen autoregressive model for runoff series was later shown to represent outflow from a system of linear storages fed by an autoregressive input of a lower order. Yet another example is the empirical concept of partial runoff-contributing area that was later found to follow from a physically based model of flow on a hillslope (Chapter 1).

7. *Model's Good Reputation Is Based on Superficial Appearances* Whenever a reasonable result is obtained by a model that accommodates in a logical way a number of factors believed to be relevant to the problem on hand, the model tends to acquire a good reputation. However, it may well be that many of the factors chiefly responsible for this reputation are in fact redundant as far as the results are concerned. Conceptual hydrologic models are prone to acquiring this "false dignity." For example, Mein and Brown (1978) demonstrated that a particular conceptual-deterministic model, designed to simulate monthly flows on the basis of 13 optimized parameters and a daily time step, performed only marginally better than the same model with only 3 parameters optimized and the rest set ". . . arbitrarily high or to zero and . . . others fixed to values which can be physically justified" (in the first case the model accounted for 95.3 percent of the sum of squares of deviations, in the second for 93.6 percent). Similarly, an undeserved good reputation of a regression model may derive from high correlations that may be the result of spurious correlation as pointed out by Benson (1965) and Pilgrim (1975).

8. *Model Works for the Wrong Reasons* A prominent example from astronomy is that of the Ptolemaic geocentric planetary model that, although physically unsound, predicted the common astronomical events with a reasonable accuracy. Klemes (1974) pointed out that the same may be true of some hydrologic models, e.g., of the fractional noise model for hydrologic time series. Based on the hydrologically implausible assumption of infinite memory and a constant

mean, the model produces time series statistically similar to those exhibited by hydrologic (and other geophysical) variables that are more likely to possess finite memory and time-varying means (see Figures 8.1 and 8.2). Examples of this situation are also common in models with large numbers of optimized parameters where wrong assumptions may be compensated by physically unrealistic values of some parameters so that the model gives good results.

9. *Model Is Deemed Good by Default* Models often acquire a good reputation because they did not have a chance to fail. They may not have been used long enough for their limitations to emerge. Model building being a dynamic activity in present-day hydrology, models are continually modified, recalibrated, and otherwise improved, and "good" models are superseded by "better" models before they reveal themselves as bad models.

10. *Model That Does Not Work Is Not Publicized* Last but not least, the impression that hydrologic models work generally well arises from a tautology. We learn only about the models that seem to work well and thus can only conclude that the models reported in the literature seem to work well. Moreover, because model performance tends to be interpreted as a reflection of the modelers' (model users') own competence and skill, their reports tend to have an optimistic bias. Critical evaluations of hydrologic models are rare. A notable contribution to this area is that of Pilgrim (1975) who discussed the many weaknesses of hydrologic modeling and indicates that most of them can be traced to our limited ability to properly incorporate into the models the causal relationships governing the hydrologic cycle.

DRAWBACKS, DANGERS, AND POTENTIAL BENEFITS OF CAUSAL MODELING

The most serious drawback of causal modeling in hydrology is its well appreciated difficulty, the clear awareness of the vastness of the void separating our data and understanding of the processes involved from our goals, (Chapters 1 and 2) and thus the necessity of an extensive program of basic research. The perspective of long years of painstaking observations and hard thinking without a guarantee of significant result appeals neither to many researchers nor research managers and perhaps least of all to graduate students who should be the main source of talent. The difficulty appears perhaps even greater than it is in reality because the prospective researchers view it from the perspective of their own academic background that, as already stated, typically is slanted to hydraulic engineering and water-resource systems analysis and is inadequate in climatology, geology, biology, chemistry, and physics, whose indispensability for the task soon becomes evident to them (see Chapters 2, 3, and 6).

The attendant danger of this difficulty is the temptation of reductionism: to make shortcuts and to fill the void between the data and the goals with logically plausible assumptions that are sometimes correct but often wrong and, more often than not, individually untestable. The result is a "conceptual" model that purports to be causal but sometimes is only a disguised and somewhat structured empirical construct whose elements are regression coefficients with physically sounding names. These models have a greater potential than "blind regression" and other statistical models (even though they do not always perform better, as is shown in WMO, 1975, and Garrick *et al.*, 1978) provided that their structure correctly reflects the basic aspects of the process; however, the danger resides in their tendency to elevate the preconceived hypothetical structure on the pedestal of truth and to divert attention from the investigation of the behavior of hydrologic process into the dead-end street of parameter optimization (see point 8 in the foregoing section).

The potential of causal-hydrologic models lies in their ability to derive the behavior of a hydrologic process for a given set of states of nature (physical variables) from the dynamic mechanisms of the process, without recourse to model calibration by empirical fitting. Consequently, causal models offer a possibility to predict the behavior of a hydrologic process under conditions that did not exist during the process-recorded history, i.e., under conditions for which an empirical model cannot be constructed. Herein lies the argument against frequent claims that, because simple statistical and black-box models often outperform complex causal models, there is no need for engaging in causal modeling. It is not a question of prediction accuracy for known conditions but one of model credibility in unknown conditions. It is the difference between blind extrapolation and sound judgment. The following examples illustrate this point.

Garrick *et al.* (1978) found that simple seasonal averages derived from a historical flow record give, in one particular case, better predictions than a highly sophisticated conceptual model (SSARR). In another instance, Todini and Wallis (1977) found that a simple transfer function model (CLS) performed as well or better than the SSARR and similar models. The crux of the matter is that in these and other similar studies the physical conditions in the basin were essentially the same during periods from which data were used for model development and during those used for its testing. If, for example, flows were to be estimated for a future situation when a large inundation area within the basin would be eliminated, or a large rural area urbanized, neither of the two empirical approaches cited above could accommodate the new situation, whereas the relatively inferior SSARR model could because of its distributed nature and its built-in facility for flow routing. The fact that the "accommodation" presently achievable may be unsatisfactory demonstrates only the inadequacy of the present state of causal modeling, not an infeasibility of causal modeling in general. However, it also demonstrates one additional point: progress in causal modeling can result only from more hydrologic knowledge and not from more "causally inspired" manipulation of the little knowledge we have.

Another example is offered by Eagleson's (1972) causal model for flood-frequency distribution. In a particular instance, the model may give a fit to observed flood peaks that is inferior to a fit obtained by statistical flood-frequency fitting techniques. However, the latter techniques are useless for estimating changes in flood regime due to changed land management, while Eagleson's model can offer approximate guidance because, for a given rainfall distribution (which represents one of the "first principles" of the model and remains outside of it), it relates flood-peak frequencies dynamically to such factors as the runoff-contributing area and the conditions of flood propagation within the basin. For example, he shows that if the runoff-contributing area in one of the catchments under study increased (e.g., due to urbanization) from one third to one half of the total area, a 100-yr flood could become approximately a 10-yr flood.

As another example, approaches have been suggested in which the distribution of annual (or seasonal) runoff total is causally related to the amount of perennial snow and ice and the amount of energy available for melting (Klemes, 1971) or to a number of other climatic and physical variables (Eagleson, 1978). Such approaches could, for instance, provide guidance for estimating water resources affected by future climatic changes predicted in the literature (Budyko *et al.*, 1979).

The inherent ability of causal models to assess the effects of environmental changes, both natural and man-made, on water resources is by far the most important aspect of their potential in water-resource planning and management, especially in the present epoch when the rate of environmental changes is higher and is increasing more rapidly than in any other epoch of the recorded history. On a more general level, causal approach seems to represent the only means capable of increasing the low credibility of essentially untestable models. Another important aspect of causal models is their potential to point out ways to efficient shortcuts and thereby to better empirical models. For example, the kinematic wave approximation of the equations of motion suggests that the simple concept of a nonlinear-storage reservoir (or a system thereof) may have a relatively great potential in basin-runoff modeling (Laurenson, 1964; Klemes, 1973).

On the scientific level the greatest potential of causal hydrologic models is in their intrinsic ability to meet the following challenge. "There is need for a systematization of research into mathematical models of hydrologic systems in order to provide the framework for rational methodology for the use of hydrologic models. This is necessary both for the organization of research results into a body of coherent knowledge and for the ready application of research work to field problems" (Dooge, 1972). This need has become urgent during the last decade when the proliferation of models has revealed many inconsistencies and incompatibilities in the assumptions underlying different models used to portray specific aspects of one and the same phenomenon. The only framework that can reduce these inconsistencies and incompatibilities is the causal, i.e., physically based, modeling. The following example will illustrate this point.

In the analysis of rainfall-runoff relationships it is a common practice to fit independently the system's response of the catchment, the flow-routing model (models) for its river channels, the probability distributions of precipitation and runoff, the probability distributions of precipitation and runoff extremes, and the stochastic structures of the precipitation and runoff series. As a rule, the results from the individual models are incompatible with each other in the sense that they could not be produced by one and the same physical system. For example, the unit hydrograph model of a basin is incompatible with a nonlinear cascade model of the basin's river channels; a random precipitation series combined with the unit hydrograph model cannot produce a fractional-noise-type series of runoff; and a nonlinear catchment model is incompatible with a runoff-series model based on the concept of autocorrelation. These problems have been analyzed by Klemes (1978) who showed that the unifying framework for all these *ad hoc* models can hardly be anything else than a physically consistent model of the catchment mechanisms, i.e., a *causal theory of the hydrologic cycle*. A significant step in this direction has recently been made by Eagleson (1978). For the time being, his and similar attempts must be viewed chiefly as exploratory probes. This line of inquiry seems to be the most promising way out of the unenviable present situation that has been aptly compared (Dooge, 1978) to "a riot of growth reflecting a variety of scale, colour and type and . . . a cacophony of noise . . . confronting . . . a traveller lost in a jungle."

CONCLUSION

Prediction of future states (or ranges thereof) of hydrologic processes is a necessary prerequisite for scientific management of water resources. It requires adequate modeling of the hydrologic cycle in which both the empirical and the causal models have their legitimate place. However, it must be recognized that even at its best a model can be nothing more than a particular system of organization of hydrologic knowledge; at its worst it degenerates into a manipulation of conjectures. The degradation of the first case into the second is inevitable if the body of knowledge to be organized does not increase proportionally to the advancement of the methods for this organization. There are many signs that this degredation process has been taking place in hydrologic modeling in the recent past with active, though unintentional, help from water-resources management, which as a rule has placed models above the knowledge on which they should be based.

For many years, hydrologists have been discouraged, through one-sided training and short-sighted "cost-effective" research financing, from penetrating to the roots of the hydrologic processes and the mechanisms underlying the hydrologic cycle. On the other hand, they have been applauded for serving the same cheap mathematical cocktails offered under a variety of exotic names but always mixed from one or two of the three standard ingredients (often of questionable quality): point precipitation, streamflow, and groundwater level. Naturally, the same hangovers followed only to be "cured" by more of the same mixes. Unless this vicious circle is broken, the cult of mathematical embroidery of sterile concepts abolished, advancement of hydrologic knowledge (rather than fast service for water-management problems) established as the objective of hydrology, there is little hope for a substantial improvement in the scientific basis for water-resource management, an improvement that may be much needed in the years ahead.

REFERENCES

Benson, M. A. (1965). Spurious correlation in hydraulics and hydrology, *Proc. Am. Soc. Civ. Eng. 91*, 35-42.

Bohm, D. (1957). *Causality and Chance in Modern Physics*, Routledge and Kegan Paul, London.

Bronowski, J. (1972). *Science and Human Values*, Harper & Row, New York.

Budyko, M. I., K. Ya. Vinnikov, D. A. Drozdov, and N. A. Yefimova (1979). Impending climatic change, *Soviet Geography XX(7)*, 395-411 (Original in Izv. Akad. Nauk SSSR, ser. geogr., 6, 5-20, 1978).

Dooge, J. C. I. (1972). Mathematical models of hydrologic systems, in *International Symposium on Modeling Techniques in Water Resources Systems, Vol. 1*, Environment Canada, Ottawa, pp. 171-189.

Dooge, J. C. I. (1978). General report on model structure and development, in *International Symposium on Logistics and Benefits of Using Mathematical Models on Hydrological and Water Resource Systems*, IBM, Pisa, October.

Eagleson, P. S. (1972). Dynamics of flood frequency, *Water Resour. Res. 8*, 878-897.

Eagleson, P. S. (1978). Climate, soil, and vegetation, *Water Resour. Res. 14*, 705-776.

Fiering, M. B. (1967). *Streamflow Synthesis*, Harvard University Press, Cambridge, Mass.

Fiering, M. B. (1976). Reservoir planning and operation, in *Stochastic Approaches to Water Management, Vol. II*, H. W. Shen, ed., H. W. Shen, Fort Collins, Colo., pp. 17:1-17:21.

Garrick, M., C. Cunnane, and J. E. Nash (1978). A criterion of efficiency for rainfall-runoff models, *J. Hydrol. 36*, 375-381.

Jackson, B. B. (1975). Birth-death models for differential persistence, *Water Resour. Res. 11*, 75-95.

Jettmar, R. U. and G. K. Young (1975). Hydrologic estimation and economic regret, *Water Resour. Res. 11*, 648-656.

Kartvelishvili, N. A. (1967). *Teoriya veroyatnostnykh protsessov v gidrologii i regulirovanii rechnogo stoka*, Gidrometeorologicheskoe izdatel'stvo, Leningrad.

Kartvelishvili, N. A. (1975). *Stokhasticheskaya gidrologiya*, Gidrometeorologicheskoe izdatel'stvo, Leningrad.

Klemes, V. (1970). Negatively skewed distribution of runoff, in *Symposium of Wellington (N.Z.)*, IASH Publ. No. 96, pp. 219-236.

Klemes, V. (1971). Some problems in pure and applied stochastic hydrology, in *Proceedings of the Symposium on Statistical Hydrology*, Tucson, Ariz.

Klemes, V. (1973). Watershed as a semi-infinite storage reservoir, *Proc. Am. Soc. Civ. Eng. 99 (IR4)*, 477-491.

Klemes, V. (1974). The Hurst phenomenon--a puzzle? *Water Resour. Res. 10*, 675-688.

Klemes, V. (1977). Value of information in reservoir optimization, *Water Resour. Res. 13*, 837-850.

Klemes, V. (1978). Physically based stochastic hydrologic analysis, *Adv. Hydrosci. 11*, 285-356.

Klemes, V. (1979). The unreliability of reliability estimates of storage reservoir performance based on short streamflow records, in *Reliability of Water Resources Management*, Water Resources Publications, Fort Collins, Colo., pp. 193-205.

Laurenson, E. M. (1964). A catchment storage model for runoff routing, *J. Hydrol. 2*, 141-163.

Mandelbrot, B. B. (1970). Comment on "Stochastic models in hydrology" by Adrian E. Scheidegger, *Water Resour. Res. 6*, 1791.

Mein, R. G., and B. M. Brown (1978). Sensitivity of parameters in watershed models, *Water Resour. Res. 14*, 299-303.

NRC Geophysics Study Committee (1977). *Climate, Climatic Change, and Water Supply*, National Academy of Sciences, Washington, D.C., 132 pp.

Novotny, J. (1963). *Dve stolete hydrologicke rady prutokove na ceskych rekach*, Sbornik praci, No. 2, Prague.

Pilgrim, D. H. (1975). Model evaluation, testing and parameter estimation in hydrology, in *Prediction in Catchment Hydrology*, Australian Academy of Science, Canberra, pp. 305-333.

Scheidegger, A. E. (1970). Stochastic models in hydrology, *Water Resour. Res. 6*, 750-755.

Slack, J. R., J. R. Wallis, and N. C. Matalas (1975). On the value of information to flood frequency analysis, *Water Resour. Res. 11*, 629-647.

Todini, E., and J. R. Wallis (1977). Using CLS for daily or longer period rainfall-runoff modelling, in *Mathematical Models for Surface Water Hydrology*, T. A. Ciriani, U. Maione, and J. R. Wallis, eds., John Wiley & Sons, Inc., London, pp. 149-168.

von Mises, R. (1957). *Probability, Statistics, and Truth*, G. Allen and Unwin, London.

WMO (1975). *Intercomparison of Conceptual Models Used in Operational Hydrology Forecasting*, World Meteorological Organization, No. 429, Geneva.

WMO and UNESCO (1974). *International Glossary of Hydrology*, World Meteorological Organization, No. 385, Geneva.

Yevjevich, V. (1963). Fluctuations of wet and dry years, Part I, in *Hydrology Papers 1*, Colorado State University, Fort Collins, Colo.

Yevjevich, V. (1974). Determinism and stochasticity in hydrology, *J. Hydrol. 22*, 225-238.

Ziman, J. (1978). *Reliable Knowledge*, Cambridge University Press, London.

Field Data: The Interface between Hydrology and Geomorphology

9

LUNA B. LEOPOLD
University of California, Berkeley

In 1904 the California Miners' Association adopted in convention a memorial forwarded to the President of the United States, who in turn referred it to the U.S. Geological Survey. It was a request for a study of the relationship between the hydraulic debris in California and problems of flood control and sedimentation. The Geological Survey assigned their most prominent geologist, G. K. Gilbert, to conduct the investigation. He soon became convinced that, although it was possible to make reasonable estimates of the amount of the debris accumulated in the various portions of the valleys and basins of the rivers flowing out of the Sierra Nevada, there was simply not enough known about the principles of sediment transport to determine with surety how the accumulated debris was going to react over time to the passage of floods. Nor was it possible to determine whether the structures that had been built for the containment of this debris were going to operate satisfactorily as planned. Some of the structures were actually under construction at the time Gilbert was making his investigation, and he was able to map the sediment accumulation before and after one of the important structures had been raised in height. The concrete and log dam, called Barrier No. 1 on the Yuba River, was one of those studied in detail by Gilbert. He expressed concern with the erosion at the toe resulting from water deprived of its sediment load pouring over the structure and scouring downstream from the dam.

As a result of these considerations, Gilbert instituted his famous experiments on the campus of the University of California, Berkeley, the data from which have been a major source of information on transport mechanics.

One wonders whether, if the same problem faced modern hydrologists, there would be an equally sophisticated understanding of how the lack of specific data would prevent any useful solution to the problem at hand. On the basis of experience of the last several decades it would be reasonable to forecast that the modern hydrologist would have begun with a dimensional analysis and from that he would construct a theoretical set of equations that would purport to relate sediment transport rate to hydraulic factors. The prospect of 2 yr of laboratory work probably would deter the initiation of an investigation for the collection of basic data.

Even if the modern hydrologist were wise enough to choose a course of action that would involve the collection of a long series of carefully planned basic data, it is unlikely that he would find a place to publish the data themselves. One cannot imagine a modern organization willing to publish a report that totaled 263 printed pages, including not merely the summarized data but also the original observations themselves. In the Gilbert report on his sediment-transport experiments the tabulations of the basic observations involve 16 printed pages, which include the direct observations as they were actually recorded. A whole chapter of the report is devoted to what Gilbert called "Adjustment of Observations," in which 12 printed pages are devoted to a retabulation of the same basic data. In his words, "The observational values were subjected to a process of adjustment, whereby the sequences were freed from irregularities. The irregularities are made manifest by the comparison of the sequences of two variables, and first consideration will be given to those of capacity and slope" (1914, page 55).

There follows in the report retabulated data organized for specific comparisons such as the grouping of a particular discharge at a particular width and tabulated such that the relationships, in this case between capacity and slope, could be plotted easily for further analysis.

Finally, after 218 pages of detailed analyses of the data, there is a chapter in Gilbert's report that deals with the application of these laboratory experiments to natural streams. Further, it must be understood that in addition to the report on his experimental work, there is a separate report on the mining-debris problem, to which he had been assigned originally. This report, U.S. Geological Survey Professional Paper 105, consists of 153 printed pages. In the latter paper, "Hydraulic-Mining Debris in the Sierra Nevada," Gilbert deals with the matter of tides and the tidal prism in San Francisco Bay. He made detailed measurements of the relationship among velocity, water-surface slope, and gauge height, as well as channel geometry of a small tidal estuary in the southern portion of San Francisco Bay. To include a study of tides and their effect on the estuarine geometry in a report on hydraulic mining is an interesting aspect of the bredth of Gilbert's thinking about a particular problem.

Many of the problems that Gilbert faced remain incompletely solved. Some of the most important current problems are those that lie at the interface between hydrology, hydraulics, and geomorphology.

Over the years there has been a series of studies of the gradient of deposition of sediment trapped behind a dam or barrier. From these field observations it is now well known that the gradient of deposition of sediment behind a barrier, not in a reservoir but behind a check-dam, large or small, will in all cases be less than that of the original valley floor. Observations on this matter had not previously been made. Gilbert's statement about this matter is as follows (1917, page 53):

> The initial slope of the river bed in the vicinity of the barrier was about 16 feet to the mile, and there was a rapid increase upstream to 2 feet to the mile . . . At Marysville, near the mouth of the river, where the chief material is sand, the slope is about 6 feet to the mile, and it increases upstream as the material changes successively to a coarse sand, fine gravel, and coarse gravel. The erection of the first 6 feet of the barrier in 1904 caused a filling during the flood stages of the ensuing winter which extended at least 1 1/2 miles upstream and probably somewhat more, reducing the average slope in that region to about 12 feet to the mile. When the second unit was added to the barrier, increasing its height 8 feet, deposition again began, and the flood of January, 1906, in filling the newly formed basin, extended its deposit upstream about 1 3/4 miles, reducing the average grade for that distance to about 9 feet to the mile. The subsequent storms of the same winter extended the area fill somewhat farther upstream and nearly restored the slope of 12 feet to the mile which had been created by the floods of the preceding winter.

Gilbert's diagram (1914, page 59, Figure 7) is the first quantitative observation of the difference in the depositional grade behind a barrier from the gradient of the original stream before the barrier was built. He did not attempt to forecast what would eventually happen if the barrier did not fail, but there is no indication in his report that he expected the depositional wedge to extend upstream indefinitely. He did not specifically say this, but the implication in his report is that a form of equilibrium had been established; he did not take the matter any further. Many subsequent reports have noted the same result, and it has been a matter of concern to geomorphologists ever since consideration of the hydraulic reason why the sediment wedge does not continue to extend itself upstream indefinitely. The latest discussion of this matter is that by Leopold and Bull (1979), which includes the first attempt to provide a hydraulic explanation of this observed fact.

This matter is important because it bears on the interaction of those factors that determine the gradient of a river channel through any particular reach of the river. This involves both hydraulics and geomorphic history and probably is one of the central questions of fluvial geomorphology even at the present time. The slope of a stream in a geologic or geomorphic sense involves the interaction of eight (or possibly more) variables and cannot be completely solved in an analytical fashion. Improved understanding will depend on obtaining a series of field observations that include factors not ordinarily measured at stream-gauging stations, especially water-surface slope and its relation to the pool and riffle sequence and to river curvature. It is nearly impossible to imagine that any so-called model or computational scheme would be effective because there are not requisite field data on which the model might be predicated.

Another problem now taking up the time of many hydrologists and some geomorphologists is the estimation of the effect of forest practices on water quality and sediment production. In recent years the tendency for timber harvesting to be conducted by clear-cutting has posed the possibility of long-term deterioration in water quality, which depends to a large degree on slope stability and channel stability. Formerly it was thought that the best practice for erosion control requires the maintenance of the soil profile in forested areas. It is now claimed that silvicultural efficiency in coniferous forests of the western United States demands that the surface soil be stirred, mixed, and churned. This means the destruction of the forest soil profile. To what extent the destruction of the soil profile is going to alter water quality over a period of years, and to what extent it will change the rate of sediment production, remains to be seen. But theory expressing the leaching by infiltrating water, the cation adsorption on mineral particles, the effect of soil structure on erodibility, and other processes make it impossible to use any known analytical technique to forecast how water quality will change over the next several decades, downstream from areas that have been clear-cut of their forest cover. The amount of water yielded from a clear-cut area and the timing of runoff have been investigated in many areas, but concurrent studies of sediment deposition, its location, form, effect on stream channels, and water quality are few. Long-term field observations will be needed to ascertain wnat principles are the governing ones.

Considering the experimental work done in forested watersheds, the record shows that much effort has been spent in watershed research attempting to forecast the effect of various forest practices on water yield. It is interesting to note that, in the actual management of our national forests, there are few instances of management practices either governed by or even greatly influenced by the need to increase water yield, even in relatively water-short areas. Rather, forest management practices in recent decades have been governed by timber production, recreation, and wildlife--timber production being given the greatest weight. It appears then that most of the hydrologic research on forested watersheds has been and still is being directed toward what is theoretically an important problem but which in actual practice is not given much weight in forest administration.

The point here is that much hydrologic research is directed at problems that are not necessarily the most significant ones in theory, on the ground, or in practice.

Where, for purposes required by recent legislation, it is necessary to assess changes in water quality and sediment yield, especially in connection with environmental impact statements, most of the effort is on office compilations and theoretical constructs rather than on the collection of field data. The lack of surety and even the lack of confidence in most environmental impact assessments come from the fact that the effort is on fulfilling the requirements of the report rather than on the collection of those data that might actually yield the desired answers.

These same remarks might be applied to a whole list of problems in hydrology, but especially those that have an interrelation with geomorphology. A short list of such problems might include the following:

1. *Channel stability* The effect of man's work--agriculture, forest management, grazing--on channel stability is essentially an area of unknowns. A classification of stability based on field measurements with associated information derived from soil mechanics, pedology, and hydraulics is needed.
2. *Rate of bank-cutting and lateral migration of rivers* There is no common body of knowledge dealing with factors controlling bank erosion even though large amounts of money are expended for river revetment, especially at highway crossings.
3. *Erosion by overland flow* Though the hydraulics of thin films are well known, variations in rates caused by nonuniformity of surficial materials and the effect of vegetation cannot be solved by analytical techniques. Direct field observations under various conditions are needed.
4. *The profiles of hills* A great deal of theoretical work has been devoted to constructing equations purported to express the longitudinal profile of hillslopes. This problem is a peculiar combination of hydraulics, hydrology, geomorphology, pedology, and soil mechanics.
5. *The effect of sediment availability in rivers on sediment-transport rates* With the recent development of the Helley:-Smith bedload sampler it appears possible to obtain actual measurements of bedload-transport rates. Difficulty exists, however, in interpreting measurement data because the sediment-transport rate may be limited if the sediment is not available for transport. Therefore, actual river data on sediment-transport rate may not appear to follow the usual empirical relations. Little is known about how to estimate whether sediment is available in the stream channel for movement.
6. *Effect of heterogeneities in geologic materials on groundwater movement* The usual equations dealing with transmissibility, specific yield, permeability, and porosity assume that the geologic material is essentially homogeneous. The inhomogeneities alter both the direction and paths of groundwater flow. Present methods of geologic mapping for groundwater purposes are inadequate to assess the form and effects of heterogeneity in the geologic materials. This is a problem involving a combination of hydraulics, geomorphology and geomorphic history, sedimentology, and hydrology.

It would seem desirable to keep in mind the most important theoretical and practical problems facing the hydrologic science, especially as it relates to associated sciences or subsciences. A study of the title of grant proposals in the field of earth sciences related to hydrology and geomorphology evokes a certain surprise that most of the research projects deal with problems that are not of general interest. One knows from experience the history of most research proposals and how they are often determined by the preferences of the granting agencies. But these preferences are the result of consideration by competent scientists. The scientific view of what is worth working on, however, is usually expressed in terms so general that even poor projects might appear to qualify under the general headings listed as priority items. It would appear far more practical to express priority items for research funding in terms of the questions to be answered rather than a general category of work. As an example, a priority item for a funding agency might be listed as "the problem of urbanization." It would be readily agreed that the hydrologic results of urbanization constitute a large class of problems worthy of study. But that general class could be broken down into specific questions such as, how many tons of sediment per square kilometer will be contributed by housing construction using construction methods that are typical of the area near Washington, D.C.? The answer is not likely to result from present theory, but would require a set of specific field observations designed to answer this specific question.

It appears that the hydrologic science is making only slow progress at the present time because of the choice of the questions being researched and the tendency to study only questions amenable to known theory with the minimum of field measurement.

REFERENCES

Gilbert, G. K. (1914). *The Transportation of Debris by Running Water*, U.S. Geol. Surv. Prof. Pap. 86, 263 pp.

Gilbert, G. K. (1917). *Hydraulic Mining Debris in the Sierra Nevada*, U.S. Geol. Surv. Prof. Pap. 105, 154 pp.

Leopold, L. B., and W. B. Bull (1979). Base level, aggradation, and grade, *Proc. Am. Philos. Soc. 123*, 168:-202.

Geology, Determinism, and Risk Assessment

10

VICTOR R. BAKER
University of Arizona

THE SCIENCE OF RISK ASSESSMENT

The concept of risk implies the probability of undesirable events. Floods comprise a hydrologic risk that is most often analyzed probabilistically. The standard procedure is to develop statistical associations between observed flood events and their estimated probabilities. The derived associations, extrapolated along theoretical distributions, may provide the primary basis for such decisions as the choice of land suitable for certain types of use or the choice of a spillway design for a dam. Although these procedures yield quantitative data, amenable to decision implementation, the underlying assumptions are subject to continual debate and to considerable research by hydrologists. Rather than contributing to this debate, this paper introduces a completely different approach to risk assessment: the use of the geologic record to evaluate risk potential.

The assessment of risk potential has a different goal from the derivation of associations between events and their probabilities. The latter exercise requires an acceptable probability distribution for the events, here an assumed flood-frequency distribution. Various model distributions will always provide quantitative data for decision makers, but those data will probably never inspire universal confidence. In contrast, risk potential is assessed by the analysis of past catastrophes in time and space. The occurrence of one superevent shows that the phenomenon warrants serious consideration. The occurrence of several superevents in a given time span at a given locality shows that such events can be expected sometime within the given time span. Finally, the clustering of events in space and time may indicate important nonrandom elements in the system, such as changes of climate, vegetation, and land use. This information does not constitute an accurate quantitative prediction of a phenomenon, but it does inspire confidence that certain phenomena have occurred and that others have not. Risk prediction is a result that will always be questioned, whereas risk-potential assessment is a continuing study necessary to asking the proper scientific questions.

The present state of the science of hydrologic flood studies is similar to that experienced in other sciences, in which one line of study, successful in its early applications, is stretched to its methodological limits. Speaking of increasing specialization in geology and a seeming preoccupation with the refinement of existing scientific methodology, Davis (1926) observed, "We shall be indeed fortunate if geology is so marvelously enlarged in the next thirty years as physics has been in the last thirty. But to make such progress, violence must be done to many of our accepted principles."

Stated another way, we can observe that many scientists view their profession as concerned with an orderly search for truth. The concept portrayed in scientific textbooks is that knowledge is gained by a steady, cumulative process. One philosophical view, however, holds that science is characterized by long periods of conformity to a commonly held theoretical framework (Kuhn, 1962). This framework, which Kuhn (1962) calls the "paradigm," is the tradition in which new investigations proceed. The science that occurs during paradigm-governed periods is "normal science" and mainly consists of solving puzzles generated by the paradigm. Normal science yields rapid progress through its ability to predict phenomena.

A corollary to Kuhn's paradigm concept is that the new hypotheses generated to explain phenomena in a particular scientific field will be most often conditioned by the paradigm. In essence, a set of unstated "ground rules" dictates the manner in which science will identify puzzles for its study. These rules will also act to exclude many possible explanations for natural phenomena if such explanations lie outside the bounds of normal science.

Kuhn's thesis holds that normal science eventually turns up puzzles that cannot be solved within the existing theoretical framework. Such anomalies are frequently ignored, and attempts may even be made to suppress their study to prevent interference with the rapid progress that appears to occur in normal science. However, these anomalies hold the key to a superior kind of scientific progress. When their importance can no longer be ignored, a scientific field may experience crisis, pronounced change, and eventually a completely new synthesis. This is the "scientific revolution."

The concept of scientific revolutions has been applied mainly to chemistry, biology, and physics. Although Kuhn's thesis may be an oversimplification, it poses fascinating questions for the consideration of any science. The questioning of first principles in a scientific field is one indication of crisis. Does this trend in hydrology indicate that the discipline may be poised for pronounced change? Unfortunately, that question is better answered by a future historian of science than by the actual scientists themselves.

This essay will restrict its examples to the risk assessment of flood hazards. This subject illustrates the crisis posed for the science of hydrology by the need to use hydrologic data as a basis for water-resource management. The introduction of geologic data can assist this goal by extending the data base from which to analyze flood hazards. However, the introduction of this data leads to some questioning of basic principles. In addition, the introduction of geologic methodology perhaps can clarify some philosophical questions concerning the goals of hydrology as a science.

The ultimate risk considered by water-resource management is embodied in the word "catastrophe." However, this word also has unfortunate connotations, e.g., the following American Geological Institute (1972) definition: "*catastrophe*--A sudden, violent disturbance of nature, ascribed to exceptional or supernatural causes, affecting the physical conditions and the inhabitants of the Earth's surface; e.g., the Noahian flood, or an extinction of an entire fauna." The phrase exceptional or supernatural causes implies that the study of catastrophes is something beyond the bounds of science. Quite the contrary, it is the exceptional nature of catastrophes that warrants close scientific scrutiny.

During the last century, when individual communities were largely self-sufficient, the occasional catastrophic flood produced only a local calamity. The present century has been marked by the increasing interdependence of communities through complex transportation, communication, resource, and energy systems. By disrupting these technoecosystems, local disasters now exert a profound impact on society as a whole.

An increasing percentage of the total annual flood loss in the United States is occurring as a result of so-called catastrophic floods. These are floods that either (1) have a return period of 100 yr or more or (2) cause failure of a flood-protection project by exceeding the project design flood (Holmes, 1961). The average amount of flood loss from floods of moderate frequency is decreasing relative to these catastrophic events. Overtopping of levees designed for rare floods accounted for at least 40 percent of the losses from Hurricane Agnes in 1972 (White, 1975). The $3.5 billion damage produced by Agnes's flooding probably represents the trend for all future flood losses in the United States. The current national effort on flood control continues to curb the level of damage from more frequent flooding, while increasing the risk that highly developed, "protected" flood plains will experience a flood that exceeds the design capacity of the control works. These trends show the importance of being able to estimate the potential for achieving catastrophic floods.

THE DILEMMA OF CATASTROPHIC RISK ASSESSMENT

At present two procedures are used routinely in the United States for water-resource decisions based on the magnitudes of catastrophic floods: (1) calculation of the probable maximum precipitation over a drainage basin and conversion of this pattern to stream runoff and (2) statistical flood-frequency computations. Both methods have inherent difficulties and have sparked controversy as to which approach is more valid (Yevjevich, 1968). The probable maximum precipitation (PMP) approach assumes that there is a natural physical limit to the rainfall intensity-duration relationships (Slade, 1936). This limit can presumably be calculated from the most severe rainfall that may be expected from the most severe combination of meteorological and hydrologic conditions that could be expected in a region. In reality the probable maximum flood for an area is a dynamic value that changes as our understanding of great rainstorm hydrometeorology changes. Estimates for the PMP will undoubtedly increase as knowledge of the meteorological phenomena responsible for storms becomes more sophisticated (Paulhus and Gilman, 1953). Moreover, climatic and hydrologic conditions may change over time scales of decades or centuries (Stockton, 1975). For the long time scales, over which very large flood events assume a relatively high probability of occurrence, changes in climate and drainage-basin properties can attain a primary importance.

Proponents of the statistical flood-frequency approach feel that meteorological conditions and antecedent drainage-basin conditions can always be slightly worse than what is assumed for the probable maximum flood. By this philosophy, no physical limits for flood magnitudes or rainfall intensity-duration relationships exist within the probability span used in most water-resource decisions (Yevjevich, 1968). Therefore, even the PMP has a definite probability. Otherwise, structures designed by the PMP concept will imply a no-risk-of-failure situation, which fosters a false sense of security to occupants of the downstream floodplain (Yevjevich, 1968).

The statistical flood-frequency approach attempts to predict the tail of a probability distribution from a small sample of values that usually are not included in this tail (Moran, 1957; Melentijevich, 1969). The use of probability-density functions or frequency distributions assumes that the data base (streamflow records) consists of random samples from the total population of flows with a frequency distribution whose moments we can estimate. Flood hydrologists have devoted inordinate effort to specifying the proper probability-density function for a series of flood flows. The national standard in the United States is the log-Pearson III distribution (U.S. Water Resources Council, 1967, 1976), which has three parameters: the mean, the standard deviation, and the coefficient of skewness of the logarithms of the peak flows. Whatever distribution is chosen, a continuing criticism of the statistical flood-frequency approach is its failure to assess accurately the probability of rare catastrophic floods (Meyers, 1966, 1967; Commons, 1966). The problem with the continuing criticism is that it frequently is as naive as the probability methods that are applied. In fact, given the extraordinary sensitivity of estimates of tail probabilities, it is virtually impossible to meaningfully criticize one flood frequency or another, because the levels of statistical significance associated with tests of goodness of fit typically apply equally well (or badly) to a wide range of candidate functions. Such errors have been explained as a failure to properly define the mathematical form of the flow population (Benson, 1962). However, more recent literature by Matalas *et al.* (1975), Wallis *et al.* (1977), and Slack *et al.* (1975) points to a more basic problem by noting the issue to be the simultaneous selection of the distribution and estimation of its parameters. The problem cannot be parsed into its component parts, assessed independently and then combined. Slack *et al.* (1975) showed that the use of a logarithmic distribution for annual floods, applied to typical record lengths, produces instabilities in the estimation of a frequency-discharge function far out on the tails and that for a given class of economic-regret functions it is more efficient (in an economic sense) to consciously use a density function that avoids the calculation of antilogarithms (such as the normal distribution) even though it might be known that the true density function is (for example) log-normal with unknown parameters. This is not an attempt to influence the estimation of "true" probabilities, but rather to reflect the economic realities associated with being wrong.

There exist other theoretical limits to flood-frequency analysis, including the establishment and identification of bounds on sample estimates of the moments. These bounds are functions of sample size, and typical record lengths impose serious constraints on the ability of a sample to reliably estimate population values. Regional data are useful for increasing the effective record length, but they frequently reveal pronounced variation in flood-frequency distributions in areas of diverse climate and physiography. Figure 10.1 shows several graphically fitted curves from western Gulf Coast rivers draining nearly equal areas. Streams in areas of higher mean annual precipitation, dense vegetation, permeable soils, and relatively flat topography (curves 1, 2, and 3) exhibit relatively little contrast between rare, large floods and very frequent, small ones. However, rare high-magnitude events obviously highly distort the shapes of flood-frequency curves in the dry interior of Texas, a region of sparse vegetation, impermeable soils, and steep slopes (Patton and Baker, 1976). These distorted curves (curves 4 and 5) are examples of highly skewed flood series.

Another example of a troublesome flood record is illustrated in Figure 10.2. The record contains a spectacular outlier. On June 24-29, 1954, an extratropical disturbance resulting from Hurricane Alice in the Gulf of Mexico migrated far inland and produced rainfall of up to 1067 mm on the Pecos-Devils River divide (Figure 10.3). This storm produced the largest flood on record in Texas and peaked at 27,440 m^3/sec near Comstock, Texas, nearly eight times the maximum discharge of the other 54 recorded annual flood peaks. A naive extrapolation of the frequency curve of the 54 lesser floods (Figure 10.2,

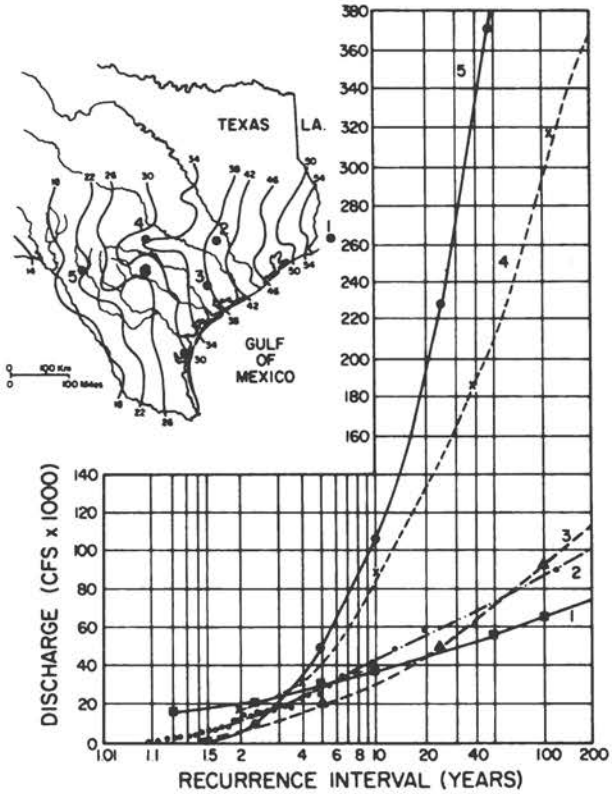


FIGURE 10.1 Flood-frequency curves for rivers of the western Gulf Coast with drainage areas of approximately 2000 km^2. The inset shows the locations of the gauges and isohyets (in inches) of mean annual precipitation [from Baker, (1977) with permission of the Geological Society of America].

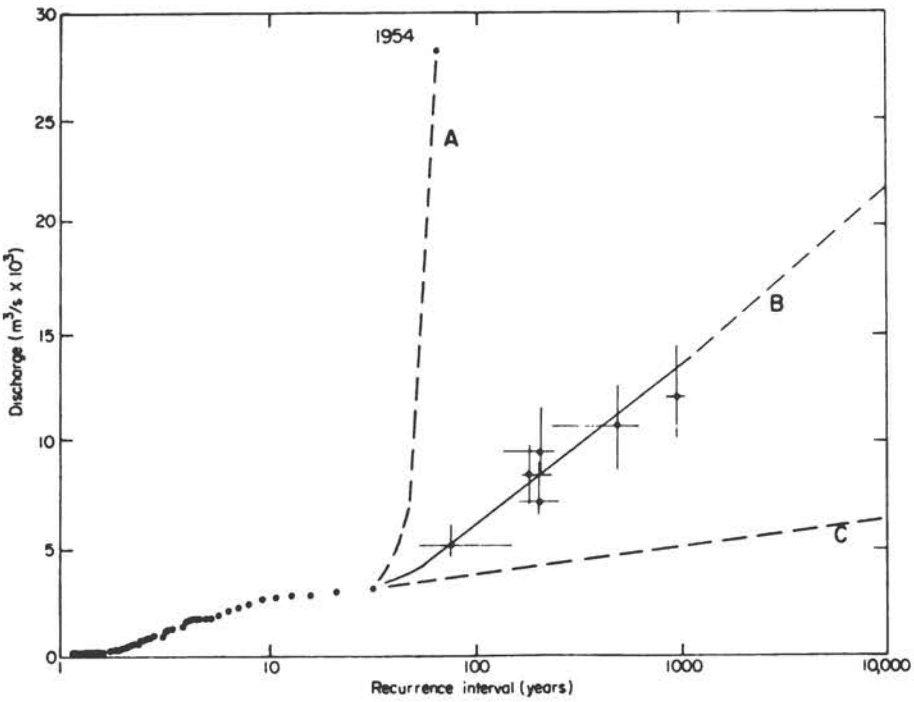


FIGURE 10.2 Flood-frequency curves for the Pecos River near Comstock, Texas. Curve A shows the frequency curve based on the 1954 flood and the other 39 yr of recorded annual flood peaks. Curve C is the extrapolated flood record, excluding the 1954 event. Curve B is a paleoflood-frequency curve estimated from the alluvial stratigraphy at Arenosa Shelter (Figure 10.3, site A). Vertical bars for the Arenosa data represent the estimated error in determining flood discharge from the sediment elevations. The horizontal bars are the standard deviations of the radiocarbon dates.

curve C) would estimate the recurrence interval of the 1954 flood to be on the order of millions of years. Similarly, inclusion of the outlier (Figure 10.2, curve A) is equally troublesome because its use changes all the ranks of the smaller flood events and hence all the plotting positions.

The above difficulties have been used to justify appraisals and reappraisals of flood-frequency analysis. This paper suggests that the problem may be recast from one in analytical statistics to one in applied geology. Geologic information from critical locations may provide the key information on extreme events against which to judge various institutional, societal, and economic issues. The use of geologic data involves a philosophical shift in the science of floods. Rather than analyzing large populations of observed events, the geologic method considers the individual superfloods that have left a record on the landscape. The goal is to reconstruct the actual flood history of a specific location over thousands of years as a basis for informed institutional and political decisions.

USE OF THE GEOLOGIC RECORD IN ASSESSING LONG-TERM RISK

Figure 10.2 (curve B) illustrates how the geologic record occasionally provides auxiliary information to enhance an existing hydrologic record. The 1954 event becomes a less surprising occurrence when one considers the geologic data from a spectacular alternating sequence of cultural habitation layers and flood slack-water

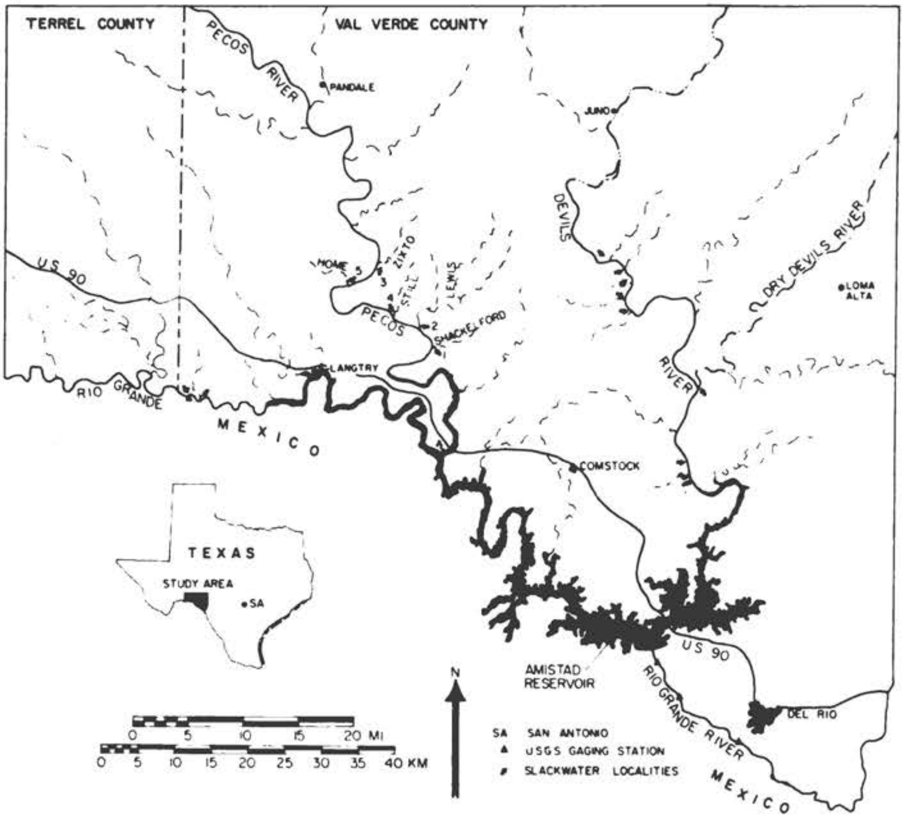


FIGURE 10.3 Index map showing the west Texas study area. The letter "A" shows the location of Arenosa Shelter, and the numbers refer to slack-water sedimentation sites. The Comstock gauging station is at the US 90 Bridge over the Pecos River (based on 1:250,000 scale maps of the U.S. Geological Survey of Del Rio and Sonora, Texas).

deposits (overbank sediments) at the Arenosa Rock Shelter (Figure 10.3, locality A) on the Pecos River near its confluence with the Rio Grande (Patton, 1977; Patton and Baker, 1977). Arenosa Shelter is a bedrock overhang that began filling with flood sediment approximately 9500 yr ago, following the diversion of the Pecos River channel away from the western bedrock wall. Sediment from the 1954 event is the uppermost deposit that formed at this protected locality. Analysis of the alluvial stratigraphy and radiocarbon dating of the intercalated cultural horizons provided the means of establishing a paleo-flood-frequency record for the Pecos River (Patton, 1977). The alluvial stratigraphy of Arenosa Shelter provided real data on ancient flood peaks. The discharge for each peak was estimated by assuming that the highest deposit elevation corresponded to the flood stage and by converting this stage to discharge by slope-area hydraulics calculations. The frequency of flooding, however, had to be estimated from the radiocarbon dates on the individual floods. Thus, the geologic data in Figure 10.2 (curve B) are not statistical-recurrence intervals in the sense of the annual series of measured streamflows. Plotting the geologic data on Figure 10.2 is meant as a heuristic device, not a design criterion for a water-resource project.

Many additional slack-water sedimentation sites occur along the lower Pecos River and the nearby Devils River. Baker *et al.* (1979) examined these, demonstrating that buried charcoal from surficial fires, buried soil A horizons, and archeological materials can provide sufficiently detailed time stratigraphy to establish flood-recurrence intervals. The buried soils are particularly common and result from the fact that floods with recurrence intervals in excess of several hundred years allow sufficient time for incipient A horizon development prior to burial by the sedimentation of a brief flood event. However, if the slack-water deposit is not sufficiently elevated above more frequent flood stages, then sedimentation by the more frequent events will be so rapid that soil profile development is inhibited (Patton *et al.*, 1979). The ideal locations for producing these buried A horizons are those that have accumulated thick slack-water sequences (certainly to an elevation in excess of 100-yr flood stage) along streams that have highly right-skewed, flood-frequency distributions and that flow in deep, narrow valleys.

Some slack-water sequences are sufficiently detailed that changes can be discerned in flood magnitude-frequency relationships over relatively long time scales. The Arenosa Shelter record contains evidence that the more humid periods during the last 9500 yr were characterized by numerous floods of intermediate magnitude but no superfloods. Other more arid periods, including the last 2000 yr, show relatively few floods of intermediate magnitude but have one or two immense floods.

Slack-water flood deposits in deep, narrow valleys appear unique in their ability to preserve a detailed record of both flood magnitude and frequency. However, many other stratigraphic situations can yield useful information on the occurrence of one or more great floods during a long span of geologic time (Costa, 1978). Even without analyzing the detailed geologic record of ancient floods at Arenosa Shelter, the long recurrence interval of the 1954 flood is implied by its burial with sediment of a 1300-yr old surface (Patton, 1977; Patton and Baker, 1977). For Elm Creek in central Texas, Baker (1977) described three bouldery deposits of superfloods, each separated by buried soils that have been radiocarbon dated. The last of the three events was the historic flood of 1972, which yielded runoff of 90 m^3 sec^{-1} km^{-2} from a catchment of 12.5 km^2. The radiocarbon dates indicate that the three superevents were separated by approximately 400 yr (Patton and Baker, 1977). In this latter example, precise magnitudes of the ancient events are difficult to assess. Nevertheless, the key point of this type of analysis is that *real* data on approximate flood magnitudes and temporal spacing can be used to assess risk potential.

The geologic examples presented here represent the results of recent research into an underutilized tool for flood assessment. The approach is likely to be alien to most hydrologists. It involves a largely deterministic analysis of selected individual phenomena, rather than the usual assumptions of randomness applied to large populations of phenomena. Aside from the philosophical implications, discussed in the next section, the method warrants consideration by hydrologists as an alternative source of input into effective water-resource decisions. What decision maker concerned with the risk of earthquakes or volcanic eruptions would base his analysis solely on statistical studies of regional seismicity and volcanism? The geologic record of a specific fault or of an individual volcano can point to important periodicities that are only ignored to the perils of the analyst. Similarly, as geophysical phenomena, floods, especially the largest floods, deserve individual scientific scrutiny. For selected localities, geology may provide a new source of data that will help stabilize the decision-making process so that it does not become a meaningless exercise in expected value calculation.

RANDOM OBSERVATION AND CAUSAL EXPLANATION

Hydrology is a science that deals largely with fluctuations and with the adjustment of complex systems to those fluctuations. As the time base of hydrologic concern is extended, so must hydrology increasingly concern itself with "catastrophic" fluctuations. Geology not only provides a method for documenting past catastrophes but it also provides a long philosophical tradition for contending with the scientific study of catastrophes.

The complexity of the geologic record has led to continued philosophical reflection among geologists. Among many concerns has been the question of whether the observed (empirical) regularity of the record is of a statistical (random) or deterministic (causal) nature. Observational data clearly support the former view, although combined deterministic-stochastic models are required to explain some phenomena (Mann, 1970). On the other hand, the most cherished first principle of geology is that of causal explanation through the principle of uniformity, i.e., the familiar concept of simplicity, or Occam's razor (see Albritton, 1967). The prospect of a randomly generated catastrophe is anathema to this tradition. Attempts were even made to suppress the study of great geologic catastrophes, such as the Lake Missoula floods (Baker, 1978). Consider the comments of Brown (1974):

> If catastrophe is not a uniform process, there is no rational basis for understanding the past . . . the only justifiable key to the past is probability and the orderliness of natural process; if uniformity is not the key, there is no key in the rational sense. . . .

In essence, the orderliness and causality of nature is a matter of faith. This is no better illustrated than in the classic exchange between Albert Einstein and Niels Bohr. Einstein's assertion was, "God does not play dice." Bohr supposedly replied, "Stop telling God what to do." The Einstein-Bohr debate arose from considerations of electrons. Heisenberg's famous "uncertainty principle" was applied to the indeterminacy of specifying the position and momentum of electrons. Bohr extended that principle to a philosophical one: There is no such thing as an electron with definite positon and momentum. To this, Einstein's rational mind rebelled: "The uncertainty principle was not an inherent feature of reality, but rather it was merely a shortcoming of our current theory for reality."

The Einstein-Bohr debate was never resolved. Indeed it may be the ultimate dilemma of all science. The need to predict certain phenomena often requires assumptions of randomness to predict at least probability distributions of large populations of those phenomena. However, this expedient has also been termed "the philosophy of scientific desperation." Sometimes it is the individual phenomena that require scrutiny.

There is some evidence that nature itself compromises in its achievement of order versus disorder in phenomena. The great Missoula floods, the largest documented in the geologic record (Baker, 1973), might be conceived of as isolated events, best considered as part of a larger population, and no more worthy of individual scrutiny than the thousands of flood peaks observed on thousands of streams that dissect our planet. Most floods do little to the landscape. Like most human ideas, they have little influence on the larger system of which they are a part. Nevertheless, a small number of ideas may lead to revolution and to a new order. Biological evolution involves mutations, most of which have no effect on the ecosystem. Given sufficient time, however, a mutant may appear that will have a profound effect, changing the entire operation of the ecosystem. The sequence of flood events is not different. A finite possibility exists that fluvial systems may react in an analogous manner when perturbed by a phenomenal excess of input. The usual factors for damping out such a fluctuation are overwhelmed, and the system enters a high-energy state of disequilibrium.

How does one study systems that have been driven far from equilibrium? Does equilibrium simply lead to chaos? Such questions are fundamental in thermodynamics. Prigogine (1977) offers the following approach: "Instead of . . . looking (in Physics) for simplicity, you can now also go the other way, and try to find the laws of complexity.

"How can order be generated out of chaos? . . . the usual theory was to calculate the average number of collisions . . . which would produce new species . . . but these new structures arise from violations of the laws of large numbers. This breakdown of the average . . . leads to the new structures."

The new structures that develop from nonequilibrium are called "dissipative structures" (Prigogine, 1978). They are giant fluctuations stabilized by exchanges of mass and energy. The Missoula floods created a landscape by an immense fluctuation in fluvial system behavior, but that landscape comprises an orderly assemblage of landforms consistent with the high-energy state of its formation (Baker, in press).

The "success" of a science might be measured by its ability to predict phenomena. Certainly the clients of scientific knowledge regard this as the paramount goal of science. Hydrology, indeed any physical science, exhibits soundness and prestige through its quantitative predictions. Although we scientists must concern ourselves with this public perception of scientific success, we must not let the scientific image conflict with the scientific process.

Kuhn (1962, 1978) argued that laboratory measurements in science generally refine principles that are presupposed by the prevailing consensus of "normal science", i.e., the paradigm. In geology and hydrology the "laboratory" is often the earth itself, but the principle remains that the quantitative regularity sought is conditioned by the regularity that is expected. Stated another way, Kuhn's argument holds that prediction with "loaded dice" is far more prevalent in science than frustrating attempts to fathom the unknown. Thus, one can demonstrate that the predictive capabilities of many hydrologic models arise because the tests of those models are severely constrained (Chapter 8).

A story of Alexander the Great illustrates that a choice must be made. King Gordius of Phrygia supposedly tied an immensely intricate knot and stated that the knot could only be undone by the future ruler of Asia. Faced with the knot's complexity, Alexander summarily cut it with his sword.

Faced with the complexity of the geologic record, many geologists have tried to apply a rational method of working hypotheses (Gilbert, 1886; Chamberlin, 1897). Gilbert (1886) even tried to avoid the pitfalls of causality by conceiving of "antecedent and consequent" relations constituting a "plexus" that pervades nature. Hypotheses are used to penetrate this plexus, but it is not their predictive ability that the scientist must seek. Rather it is the points at which models (hypotheses) fail to predict accurately. Science will only progress by this concern with the failings of its own accomplishments. The scientific process, as opposed to the scientific image, has been clearly contrasted by Bayly (1968, p. 120): "Science is not the orderly accumulation of facts; it is the orderly accumulation of rejected hypotheses."

Can hydrology maintain a dualistic philosophy on these issues? The realist position dictates a combined deterministic-stochastic approach (Yevjevich, 1974). Certainly the practical application of hydrology requires the refinement of statistical procedures. However, hydrology's scientific advancement requires a continuing process of hypothesis development and testing. Human nature will dictate that many of these hypotheses will be formulated in deterministic terms. Conflict between these two approaches arises largely when the "success" criteria of one are indiscriminantly applied to the other. Like geology as a whole, hydrology seems fated to a dilemma.

CONCLUSIONS

1. Hydrologic risk assessment is increasingly concerning itself with very rare events of high magnitude. The possible disruptive effects of these events on our increasingly complex society leads to an increased concern with "catastrophic" phenomena.

2. Many techniques utilized in normal hydrologic science are not adequate for the confident assessment of catastrophic risk. For example, normal hydrologic science must estimate rare, great floods by statistical analysis of limited sets of observational data. The moment measures of these data sets may not be adequate for hazard assessment in regions of highly variable flood behavior.

3. The geologic record has been underutilized as a source of independent data on the magnitude and frequency of hydrologic phenomena. Recent work on the geologic effects of individual floods shows that, for some circumstances, geology can play a key role in assessing catastrophic risk. More research needs to be devoted to the geologic approach and to the blending of the hydrologic and geologic methodologies.

4. The need to accurately assess hydrologic risk for society and the application of geology to that goal illustrate several fundamental concerns regarding hydrology as a science. Assumptions of randomness or determinacy in hydrologic phenomena depend largely on the scale at which the phenomena are being studied and on the goal of investigation. Hydrologists must remember that the empirical base for their science de-

rives from their implicit conception of that science. The goal of prediction requires an orderly conception of science within which to measure and extrapolate. The goal of understanding relationships among phenomena requires the continued scrutiny and reassessment of the conception of science that made prediction possible.

ACKNOWLEDGMENTS

Peter C. Patton worked closely with me on the geologic studies of floods cited in this report. Comments by William B. Bull, Myron B. Fiering, Peter C. Patton, and Thomas M. Usselman led to improvements in the manuscript. The geologic flood studies reported here were supported in part by the Division of Earth Sciences, National Science Foundation, NSF Grant EAR 77-23025.

REFERENCES

Albritton, C. C., ed. (1967). Uniformity and simplicity, *Geol. Soc. Am. Spec. Pap. 89*, 99 pp.

American Geological Institute (1972). *Glossary of Geology*, American Geological Institute, Washington, D.C., 805 pp.

Baker, V. R. (1973). Paleohydrology and sedimentology of Lake Missoula flooding in eastern Washington, *Geol. Soc. Am. Spec. Pap. 144*, 79 pp.

Baker, V. R. (1977). Stream channel response to floods with examples from central Texas, *Geol. Soc. Am. Bull. 88*, 1057-1071.

Baker, V. R. (1978). The Spokane flood controversy and the Martian outflow channels, *Science 202*, 1249-1456.

Baker, V. R., ed. (in press). *Catastrophic Flooding: The Origin of the Channeled Scabland*, Dowden, Hutchinson and Ross, Inc., Stroudsburg, Pa.

Baker, V. R., R. C. Kochel, and P. C. Patton (1979). Long-term flood frequency analysis using geological data, *Proceedings International Association Hydrological Sciences Symposium on the Hydrology of Areas of Low Precipitation*, Canberra, Australia, December 1979, IAHS-AISH Publ. No. 12, pp. 3-9.

Bayly, B. (1968). *Introduction to Petrology*, Prentice-Hall, Englewood Cliffs, N.J., 371 pp.

Benson, M. A. (1962). Factors influencing the occurrence of floods in a humid region of diverse terrain, *U.S. Geol. Surv. Water-Supply Pap. 1580-B*, 64 pp.

Brown, B. W. (1974). Induction, deduction, and irrationality in geologic reasoning, *Geology 2*, 456.

Chamberlin, T. C. (1897). The method of multiple working hypotheses, *J. Geol. 5*, 837-848.

Commons, G. G. (1966). Some observations on flood formulas and flood frequencies, in *Symposium of Some Aspects of Storms and Floods in Water Planning, Texas Water Development Board Report 33*, pp. 156-163.

Costa, J. E. (1978). Holocene stratigraphy in flood frequency analysis, *Water Resour. Res. 14*, 626-632.

Davis, W. M. (1926). The value of outrageous geological hypotheses, *Science 63*, 463-468.

Gilbert, G. K. (1886). The inculcation of scientific method by example, *Am. Jour. Sci. (3rd ser.) 31*, 284-299.

Holmes, R. C. (1961). Composition and size of flood losses, in *Papers on Flood Problems*, G. F. White, ed., University of Chicago, Department of Geography Research Paper 70, pp. 7-20.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, Ill., 172 pp.

Kuhn, T. S. (1978). *The Essential Tension*, University of Chicago Press, Chicago, Ill., 366 pp.

Mann, C. J. (1970). Randomness in nature, *Geol. Soc. Am. Bull. 81*, 95-104.

Matalas, N. C., J. R. Slack, and J. R. Wallis (1975). Regional skew in search of a parent, *Water Resour. Res. 11*, 816-818

Melentijevich, M. (1969). Estimation of flood flows using mathematical statistics, *Proceedings International Association Hydrological Sciences Symposium on Floods and Their Computations*, Leningrad, IAHS-AISH Publ. No. 84, pp. 162-174.

Meyers, V. A. (1966). Criteria and limitations for the transposition of large storms over various size watersheds, in *Symposium of Some Aspects of Storms and Floods in Water Planning, Texas Water Development Board Report 33*, pp. 4-17.

Meyers, V. A. (1967). The estimation of extreme precipitation as the basis for design floods resume of practice in the United States, *Proceedings International Association Hydrological Sciences Symposium on Floods and Their Computations*, Leningrad, IAHS-AISH Publ. No. 84, pp. 84-104.

Moran, P. A. P. (1957). The statistical treatment of flood flows, *EOS Trans. Am. Geophys. Union 38*, 519-523.

Patton, P. C. (1977). Geomorphic criteria for estimating the magnitude and frequency of flooding in central Texas, unpublished Ph.D. dissertation, University of Texas, Austin, 225 pp.

Patton, P. C., and V. R. Baker (1976). Morphometry and floods in small drainage basins subject to diverse hydrogeomorphic controls, *Water Resour. Res. 12*, 941-952.

Patton, P. C., and V. R. Baker (1977). Geomorphic response of central Texas stream channels to catastrophic rainfall and runoff, in *Geomorphology in Arid Regions*, D. O. Doehring, ed., Publications in Geomorphology, State University of New York, Binghamton, N.Y., pp. 189-217.

Patton, P. C., V. R. Baker, and R. C. Kochel (1979). Slack-water deposits: A geomorphic technique for the interpretation of fluvial paleohydrology, in *Adjustments of the Fluvial System*, D. D. Rhodes and G. Williams, eds., Kendall-Hunt Publishing Co., Dubuque, Iowa, pp. 225-252.

Paulhus, J. L. H., and C. S. Gilman (1953). Evaluation of probable maximum precipitation, *EOS Trans. Am. Geophys. Union 34*, 701-708.

Prigogine, I. (1977). Order out of chaos, Public lecture, Nov. 18, 1977 (printed in *Texas Times*, University of Texas System, Austin, December 1977, pp. 6-8).

Prigogine, I. (1978). Time, structure, and fluctuations, *Science 201*, 777-785.

Slack, J. R., J. R. Wallis, and N. C. Matalas (1975). On the value of information to flood frequency analysis, *Water Resour. Res. 11*, 629-647.

Slade, J. J., Jr. (1936). The reliability of statistical methods in the determination of flood frequencies, in *Floods in the United States, Magnitude and Frequency*, C. S. Jarvis, ed., U.S. Geological Survey Water-Supply Paper 771, pp. 421-432.

Stockton, C. W. (1975). Long-term streamflow reconstruction in the upper Colorado River basin using tree rings, in *Colorado River Basin Modeling Studies*, C. G. Clyde *et al.*, eds., Utah Water Research Laboratory, Utah State University, Logan, Utah, pp. 401-441.

U.S. Water Resources Council (1967). A uniform technique for determining flood flow frequencies, *Bull. 15, U.S. Water Resources Council*, Washington, D.C.

U.S. Water Resources Council (1976). Guidelines for determining flood-flow frequency, *Bull. 17, U.S. Water Resources Council*, Washington, D.C.

Wallis, J. R., N. C. Matalas, and J. R. Slack (1977). Apparent regional skew, *Water Resour. Res. 13*, 159-182.
White, G. F. (1975). *Flood hazard in the United States: A Research Assessment*, University of Colorado Institute for Behavioral Science, Boulder, Colo., 140 pp.
Yevjevich, V. (1968). Misconceptions in hydrology and their consequences, *Water Resour. Res. 4*, 225-232.
Yevjevich, V. (1974). Determinism and stochasticity in hydrology, *J. Hydrol. 22*, 225-238.

11

Prediction in Water Management

NICHOLAS C. MATALAS *and* JURATE MACIUNAS LANDWEHR
U.S. Geological Survey

M. GORDON WOLMAN
The Johns Hopkins University

INTRODUCTION

Decisions regarding the management (i.e., planning, design, and operation) of a water-resource system are dependent on predictions as to the state of nature and society at some future time. Predictions may be qualitative rather than quantitative statements, and the dependence may be less than an analytical kind. The temporal scope of a prediction may be short, hours or days, as in the case of decisions pertaining to reservoir releases to lessen downstream damages from predicted flood waters. In contrast, investment decisions as to the type, size, and location of the structural and nonstructural measures composing a water-resource system depend on predictions of water supply and demand decades hence. The time span of a prediction may well be in excess of normal life expectancy, e.g., in reference to predictions of environmental effects on the groundwater system resulting from the burial of toxic or radioactive wastes.

Increasing attention is being focused on the future, reflected, implicitly at least, by recent environmental (e.g., Public Law 92-500, 1972), endangered species (Public Law 93-205, 1973), and climate (Public Law 95-367, 1978) legislation. Aspects of such statutes as well as derived regulations have been questioned. For example, it has been noted that the stated goal in Public Law 92-500 of "zero discharge" is not physically realizable in 1985 or at any other time. Similarly, Public Law 93-251 (Water Resources Development Act of 1974) requires "the development of measures and programs to protect and enhance water quality" within an area of "natural integrity" while concurrently calling for the development of the area's "potential for healthful outdoor recreation." Whether such paradoxical mandates result from incorrect or (socially and politically) ineffective scientific arguments is debatable. But almost certainly, future statutes will reflect the intent to achieve social ends by hedging against uncertainty regarding future events.

Moreover, the time from conception to realization of large-scale, federally funded water projects, which has never been short, has become longer, and the decision process more involved, as social and environmental issues have come to compete with more traditional economic and technological concerns. It can be argued that national problems (e.g., those of energy and food) bearing on water resources might well lead to decisions that set the further exploitation of these resources on a costly and almost irreversible course, resources that in many regions are sparse or already nearly fully committed.

If the attendant economic, environmental, and social costs of future legislation are to be reduced, if the time from conception to realization of water projects is to be shortened, and if more resilient water-management options are to be proposed (see, e.g., Fiering and Holland, 1974), then the basis of hydrologic prediction will need to be strengthened. Most would agree that this is in the public interest and can be achieved only through research. No doubt, an increased research budget would improve the probability of doing so. But no less important is a change in the focus of research, from its present emphasis on resolving immediate operational problems to that of furthering a fundamental understanding of hydrology.

Hydrology deals with the distribution, properties, and movement of water through the natural system, conceived to be the hydrologic cycle. But it must be recognized that human activity is inherent in the "natural system," both influencing it and affected by it. This is in direct contrast to the widely held, traditional view that the effects of human enterprise can be dealt with as separable from the hydrologic cycle. Furthermore, the chemical, biological, and physical attributes of water must be seen to be as appropriate to the concern of hydrology as the quantity of the water mass. The necessity of such a viewpoint has been anticipated by legislation, such as Public Law 95-87 (Surface Mining Control and Reclamation Act of 1977), which requires the capability to characterize the quality as well as the quantity of water in surface-water and groundwater systems and to describe the hydrologic balance in relation to human activity (mining and reclamation).

In the following discussions, reference is made to the prediction of events, where the events as well as the predictions may be of a quantitative or qualitative nature. It has been said that ". . . prediction of future events is a major aim of all geophysical sciences" (NRC Geophysics Study Committee, 1978, page 5). Distinction must be made, however, between two types of predictions: those that are inherent to and deduced from axiomatic theories are referred to below as "consequences," so that "predictions" will only be used to refer to statements that are derived to answer specific, generally operational, questions, tangential to any particular theory.

Predictions, however derived, cannot be said to be right or wrong until after the fact. The *a priori* judgment that they are good or bad, plausible or not, rests on the manner by which they are obtained; i.e., the credibility of predictions derives from their supporting scientific and philosophical arguments. These arguments, rooted in the consequences of hydrologic theories advanced through fundamental research, assure a scientific foundation in the conduct of public affairs regarding water resources.

HYDROLOGIC CYCLE

The hydrologic cycle--the basis of hydrologic theories and predictions--describes the distribution and characteristics of water as it moves through the atmosphere, to and beneath the earth's surface, and returns to the atmosphere. The word "cycle" is meant to convey the notion that the movement is circular, not that it is necessarily periodic. The movement itself is manifest in various interdependent processes, such as precipitation, evapotranspiration, infiltration, and runoff, tempered by atmospheric, surface, and subsurface storage. (Note that the hydrologic cycle is seen as a collection of connected processes, not just a set of events.) The range in spatial and temporal dimensions entailed in describing the movement is large, but no larger than those of processes in other empirical sciences. The general acceptance of the hydrologic cycle as a scientific concept dates back to the late seventeenth century (see, e.g., White, 1968). Three tenets characterize the present most widely held conception of the hydrologic cycle (herein called the "traditional paradigm"): (1) the quantity of the water mass as it is distributed and moves through the cycle is of primary importance; (2) the quantity of the water mass can be described independently of the quality; and (3) human activity is an external perturbation of the cycle.

The driving energy of the cycle is primarily solar. Estimates of the principal energy inputs to the life system have been reported by Hubbert (1976) as (1) 174,000 x 10^{12} thermal watts from solar radiation intercepted by the earth's diametrical plane, (2) 32 x 10^{12} thermal watts conducted and convected to the earth's surface from its interior, and (3) 3 x 10^{12} from the combined kinetic and potential energy of the earth-moon-sun system. Of the total solar input, about 30 percent is directly reflected and scattered into outer space; 47 percent is directly absorbed and converted into heat; and 23 percent (40,000 x 10^{12} thermal watts) is dissipated in circulating through the atmosphere and the land and water bodies of the earth in driving the hydrologic cycle.

The total mass of water in circulation is treated as constant: At any one time, it is the same as it was at any preceding time or will be at any future time. The amount of water taken out of circulation, at least for an indefinitely long period of time, by geochemical processes (e.g., in the formation of anhydrides) and the number of hydrogen atoms escaping the earth's gravitational pull are negligible within terms of human experience. Nonetheless, there is no reason to believe that the mass has been constant since the beginning or that it will remain so over the life of the solar system.

Nace (1960) estimated global water supplies on a volume basis. The specific magnitudes of the volume estimates are not meant to be taken exactly--they are ball park numbers--but the relative magnitudes and percentages of the total supply are informative. Although the total supply is within the bounds of hydrologic inquiry, it is mainly the total land supply (less than 2 percent of total supply) that has concerned hydrologists. Even so, the polar icecaps and glaciers (greater than 70 percent of total land supply) receive less than proportionate hydrologic attention. Primary interest is focused on the remaining land supply (less than 0.6 percent of total supply), that which is within the traditional scope of water management. Therein, the volume of surface water, excluding polar icecaps and glaciers, is about 3 percent of the total land supply and the ratio of stream channel supply to subsurface supply is estimated to be on the order of one ten-thousandth.

Note that these percentages are slightly at variance with the more recently published global volume estimates--see, e.g., figures provided in studies from the Soviet Union published by UNESCO (1978)--but this only emphasizes the approximate nature of such figures.

The qualified percentages of global supplies cited above are not meant to imply that water management has been preoccupied with a negligible portion of the total, ignoring the larger sources represented by the polar icecaps, glaciers, and oceans. The potential of these larger supplies for augmenting conventional ones has on occasion been discussed, particularly in terms of transporting glacier ice to temperate, water-short coastal regions. And on a limited scale, desalinization plants along coastal reaches are used to augment local water supplies.

That the potential of the nonconventional sources has not been realized to a significant extent, because of economic, environmental, political, and technological constraints, does not diminish their importance regarding the management of conventional supplies. Not only are nonconventional supplies much larger than conventional ones, but they exert a significant effect on and are affected by climate. However, this relationship is not understood well enough to permit extensive support for credible predictions of future climate or of the subsequent effects that would be relevant to the management of conventional supplies. Whatever other problems are presented, "a further complication to a rational study of climatic change and water supply is the aura of uncertainty that hovers over all who try to predict the future from limited data and understanding" (NRC Geophysics Study Committee, 1977a, page 2). In part, an appreciation of the potential importance of climatic predictions to national problems of water supply motivated the establishment of the National Climate Program (Public Law 95-367, 1978).

The traditional concept of the hydrologic cycle does not serve as an adequate paradigm (in the sense of Kuhn, 1962) for guiding hydrologic research or for the management of water resources as human activity is not separable from the "natural system." Human enterprise invariably affects the movement, distribution, and characteristics of water, whether or not the effect was intentional, because no part of the natural system can be isolated from these activities. Limitations on the use of water resources are determined by rules of man (e.g., water laws) and human behavior (e.g., land use) as much as by vagaries of nature. Water-resource management seeks to effect a change in the system through the use of structural and nonstructural measures in order to obtain a specific goal. And, in some areas, such as the southwestern United States, reservoirs and irrigation channels comprise the major portion of the surficial water system, but this represents only a portion of human involvement in the cycle. Hydrologic effects may be completely inadvertent, such as changes in geomorphology resulting from the networks of roads built in an area (Bannister, 1979) and the contribution of industrial effluents to the carbon balance of an estuarine ecosystem (Knapp *et al.*, 1979).

Furthermore, not only is the cycle affected by human involvement with its terrestrial phase (i.e., surface and subsurface water) but also with its atmospheric phase. For example, the increased acidity of rain and snow is a problem that is not confined to the industrialized areas in which it arises (Likens *et al.*, 1979). It has been estimated that the anthropogenic contributions to the atmosphere of all particulates is about 12 percent and that of organic aerosols ranges from 0.2 percent to 17 percent of the "natural" contribution (Zenchelsky and Youssefi, 1979), whereas that of some trace metals exceeds the "natural" contribution by several orders of magnitude (Nriagu, 1979). As much as half of the total carbonyl sulfide, the principal sulfur-bearing compound in the atmosphere and of climatic significance, is thought to be of anthropogenic origin (Turco *et al.*, 1980). Similarly, the climatic implications of perceived increases in the concentration of atmospheric carbon dioxide due to the combustion of fossil fuels remain a matter of public and scientific concern (NRC Geophysics Study Committee, 1977b).

The active paradigm of the hydrologic cycle to be adopted should be characterized by the following three tenets: (1) human activity is inseparable from the natural system, (2) quality is no less a concern than quantity of the water mass as it is distributed and moves through the cycle, and (3) the quantity of the water mass affects and is affected by the quality of the water.

PERSPECTIVE ON PREDICTION

Hydrology seeks to describe and explain the connectedness between events in terms of axiomatic theories within the paradigm of the hydrologic cycle. However, research is not conducted in a philosophic vacuum since the content of any theory exceeds what is strictly given or implied by the data. Observations may tell us that things are different, but they do not tell us why. With regard to any system (i.e., a collection of objects), what can be physically sensed are changes in the system at successive times. The necessary connection itself cannot in any obvious way be sensed, and the inference that changes are necessarily connected has both a philosophical and a scientific basis. This point has been made by others: in the spirit of Hume, Northrop (1958, page 13) stated, ". . . one does not sense any relation of necessary connection. Nor does one directly sense probability." Explanations of necessary connection are conditioned on one's philosophical predisposition.

The above discussion has centered on the notion of connectedness rather than causality, for which there is no generally accepted definition. Bunge (1979, page 26) spoke of determination in discussing connectedness, i.e., "everything is determined in accordance with laws by something else," distinguishing between different categories that include statistical as well as causal determination. In contrast, Northrop (1958) chose to speak of a weak and a strong form of causality, as characterized by the statics of the system. Assuming the system is closed, the postulates of a mechanistic theory specify both a state function, whose independent variables define the state of the system at any time (i.e., the statics of the system), and a time function, relating the empirical values of the variables at one time to those at another (i.e., the dynamics of the system). If none of the independent variables defining the state of the system refer to probabilities, then causality is of the strong form; otherwise causality is weak.

In the case of weak causality, probability enters into the definition of the state of a system, i.e., into the theoretical statement of knowledge. Probability may enter into the description of the statics of a strongly causal system only indirectly, through the theory of errors. Thus, in the case of strong causality, probability expresses the limitation of knowledge; i.e., the empirical limitation in defining the initial or deduced state of the system.

The theory of errors recognizes that the ability to account for any system is limited by time as well as physical constraints. Narrowly interpreted, it can be said to describe the occurrence of measurement errors in the empirical data. It can be argued that any system is operationally indeterminate, as there exists some finite time period over which any time function is rendered ineffective by the cumulative measurement error (for further discussion of this point, see Smart, 1979, and Feynman *et al.*, 1965). However, even in the absence of measurement error, there is a broader implication for large (with respect to the number of objects in the collection) systems: The statics or dynamics of the system may not be definable within a lifetime unless connectedness is formulated in probability terms. This is the operational motivation for statistical mechanics. However, the difference between classical statistical mechanics and statistical quantum mechanics should be noted: The underlying system of the former is strongly causal and that of the latter is weakly causal, even though both systems are treated as if weakly causal.

These definitions, though convenient, do not resolve the philosophical and scientific debate regarding what the nature of connectedness is. Notably, in physics, efforts to find the "hidden variable" that would explain away Heisenberg's (1930) principle of uncertainty continue despite the success of quantum mechanics in dealing with subatomic phenomena. And though such efforts to date have not been fruitful, attempts have been made to establish the philosophical legitimacy of the concept of hidden variables (see, e.g., Bohm, 1957). Thus, however satisfactorily the deduced theoretical consequences accord with the experimental data, the acceptance of the theories from which these consequences are based cannot be fully decided based on the empirical experimental evidence alone.

In tracing the evolution of the concepts of determinism and randomness in science (as opposed to causal and statistical determination), Smart (1979) drew attention to how the study of geomorphology has been affected by philosophical inconsistencies in their use. He pointed out that the issues involved cannot be settled on empirical grounds and that in any case they should not be allowed to impede investigations.

Parallels in hydrology can be drawn easily--the arguments as to whether the notion of infinite memory (i.e., long-term persistence) is physically unrealistic cannot be resolved on an empirical basis. Mandlebrot (1980) discussed this point with respect to the adoption of a stationary versus nonstationary viewpoint concerning hydrologic processes. An initial look at a hydrologic record may suggest a behavior, e.g., upward trend or a point of discontinuity, not inconsistent with a prior conception of a nonstationary process. Thus, it is possible to postulate this process and develop analytical methods to describe the nature of the nonstationarity. However, it is also possible to accept the postulate of stationarity (albeit not necessarily as historically perceived in hydrology), maintaining that it is only the "shortness" of the data record that gives it the appearance of nonstationarity. Neither position can be dismissed as false on the basis of the data alone, and the adoption of either admits a completely different analysis than the other. It is possible to invoke the principle of parsimony for deciding between postulates--i.e., to adopt the notion of stationarity as it leads to a simpler explanation of the process, not subject to the infinite possible forms of nonstationarity. These arguments are also applicable to questions regarding the effect of policies with respect to water quality. For example, an apparent trend in the concentration of a particular constituent might be perceived as a manifestation of a persistent stationary process, rather than as an induced nonstationarity, and such a perception would cast trend-detection activities in an irrelevant position.

Hydrology lacks a cohesive axiomatic theoretical structure with which to describe the quantity and quality of water as it is distributed and moves through the hydrologic cycle. In attempting to explain the experience contained in the extant hydrologic data base, various empirical relationships have been developed. Although not supported by a cohesive deductive theory, many of these relationships have been codified as "empirical models" that have been so finely tuned that they have proven useful to water management in predicting the near future at particular locales although not generally elsewhere. Their limited utility is the motivation for the development of so-called conceptual models, microabstractions of the hydrologic cycle under the traditional paradigm, attempting to describe aspects of physical processes (e.g., gravity, capillarity, friction) as conceived to govern the flow of water through elemental storage components and conditioned on the principle of continuity.

A distinction is drawn herein between the terms "model" and "theory." A theory represents a synthesis of understanding, which provides not only a description of what constitutes the states of the system and their connectedness (i.e., postulated concepts), but allows deduction of consequences from these postulates. A model is an analogy or an abstraction, which may be drawn in verbal, mathematical, electrical, even physical or mechanical terms, for rationalizing the necessary connections between changes in a system--it may be derived intuitively and without formal deductive capability. The term "model" has been used so broadly in hydrologic studies that do not differentiate between empirical relationships, intuitive concepts, and formal theories that its meaning has become ambiguous; great care must be taken to discern its intent if the intuitive and postulated concepts of hydrology are to be distinguished. However they are proposed, theories evolve and are accepted as their deduced consequences survive critical examination made with reference to empirical evidence. A model may serve the purpose of examining the consequences of the theory. However, the purpose of hydrologic research is the advancement of scientific knowledge, as embodied in deductive theories, not the construction of models *per se*.

Until the middle of this century the prevailing philosophical perception of scientific inquiry was that it proceeded in a manner described as logical positivism. Positivism, an outgrowth of the German scientific movement in the late nineteenth and early twentieth centuries, was structured by the Vienna Circle in the 1930's and 1940's (see Suppe, 1977). Basically, it holds that science is a strictly rational process: Theories are proposed through inductive logic, and the proposed theories are confirmed or refuted on the basis of critical experiments designed to verify the consequences of the theories. And through theory reduction or adoption of new or modified theories, science is able to approach truth. Furthermore, the most important task of science was held to be the ability to make predictions in preference to those of description and explanation.

Popper (1959) challenged positivism on two main points, claiming that science was deductive rather than inductive and that theories cannot be verified, only "falsified." Toulmin (1953) argued that accurate "prediction" (in the sense of deduced consequences) is not a sufficient condition for the acceptance of a theory: The explanation offered by a theory must be consistent with data and also "pleasing to the mind." More recently Kuhn (1962), Feyerabend (1963), and others have strongly criticized positivism, suggesting that science is a social enterprise, that the human process cannot be disassociated from scientific inquiry. This suggestion is consistent with discussions in fields other than the natural sciences (e.g., DeMillo *et al.*, 1979). Critics of positivism have placed the philosophy of science somewhat in disarray, though a dominant alternative to positivism has not yet taken hold. However, the positivistic view continues to pervade the conduct of hydrologic research--strong emphasis is placed on such pursuits as "model verification" on the basis of the data, i.e., the empirical evidence alone (even though "model verification" is not pursued strictly in accordance with the notion of "criticality" or formulated in terms of critical experiments), and advances in hydrology are judged primarily in terms of predictive capability.

PERSPECTIVE ON RESEARCH

It is easy enough to say that more fundamental research is needed to provide a firmer basis for hydrologic predictions. More to the point is the need for assuring a scientific basis in the conduct of public affairs concerning water resources. A prerequisite for such assurance is the ability to discern scientific and transscientific questions (in the sense of Weinberg, 1972)--those that may be stated in terms of scientific language but which cannot be answered by science as their resolution transcends scientific methods and concerns. For example, directives to place a value on research cannot be answered in strict economic terms, for there is no way to anticipate the results or even the exact course of research, much less the subsequent benefits, both direct and indirect, that will be realized, nor any way of trading off its costs to the cost to society if the work were not to be done. Since many questions arising in public affairs are of a transscientific nature and since only aspects of any predictions that address these questions can be supported scientifically, the requisite predictions will necessarily have nonscientifically derived supporting arguments.

There is no lack of predictions of future and recurrent events--the making of predictions is undertaken by all who have an interest in the development of water resources, prompted by the awareness that a particular course of development may well limit the options of future generations. All predictions emanating from outside recognized scientific circles cannot be said to be erroneous or misleading. They are numerous and diverse enough so that the "law of large numbers" assures that at least one will accord with the outcome. By the same token, there is no guarantee that predictions emanating from within the scientific community will always accord with the outcome. The question comes down to which prediction society is to choose, and, having made the choice, what course of action to take to counter any implied or perceived hazards or risks concomitant with that prediction.

That the question is not to be taken lightly is exemplified by the March 1979 accident at the Three Mile Island nuclear facility near Harrisburg, Pennsylvania (President's Commission on the Accident at Three Mile Island, 1979). The nature of the accident and the threat it posed to surrounding communities have been extensively reported in the news media and need no further elaboration. To be noted is that the probability of such an accident reoccurring has not changed (assuming no technological or managerial changes). But the public's perception of the posterior probability is considerably larger than of the prior probability--no doubt attributed to greater awareness of human fallability. What was accepted as a low risk (from a technological viewpoint) prior to the accident is now perceived as a high risk (from a behavioral viewpoint)--a change in probabilistic attitude not confined to the Three Mile Island accident. However the risk is eventually assessed, it will have a bearing on the national energy policy.

Also to be noted is that flood protection at the facility was designed on the basis of the probable-maximum flood--a basis used extensively in spillway designs, particularly in those cases where the failure of the structure would pose a direct threat to lives. The hydrologic merit of this basis has been questioned on more than one occasion. The arguments pro and con might be deemed academic were it not for the awareness that the protection provided is illusory (Kirby, 1978). The values of probable-maximum floods are revised upward on those occasions when they are exceeded or nearly equaled by the magnitudes of floods. There is nothing to say that future floods will not prompt upward revision at Three Mile Island as has occurred elsewhere.

Assessment of flood risks is a traditional concern of water management and remains an important one: the potential and realized costs of floods to society continue on an upward trend. The local and immediate nature of flood hazards is in sharp contrast to those of droughts, whose effects are regionally extensive and of long duration. Although much work has been devoted to quantifying the effects of droughts on agriculture, a general methodology for quantifying drought hazards has not been developed to an extent that the mitigation of attendant losses can be brought within the mainstream of water management. Although predictions of occurrences and effects of floods and droughts are of major concern to society, of no less importance are those concerning water quality. Like those of droughts, the associated hazards and costs have not been well defined. Furthermore, water-quality management shares common environmental concerns with traditional water management but has developed and continues to be pursued separately. Outside the concerns of either water quality or traditional management but, nonetheless, of significance to society are predictions of water-related geophysical hazards such as ice breaks, channel erosion, and landslides.

Predictions of future and recurrent events may enter into decision making in a formal, analytical manner but their importance derives from the rational basis they provide for risk assessment. In turn, the credibility of the predictions derives from the scientific and philosophical arguments on which they rest. Fundamental (basic) research provides the theoretical base from which these arguments can be advanced and articulated.

Short-term prediction is essentially a technological matter, supported by established physical principles (e.g., those of hydraulics). Even so, any number of examples can be found of predictions based on empirical relationships rather than theoretically derived ones. That they should be conversely structured is a reasonable expectation, but such approaches have been impeded by the fact that over the years many "empirical models" have been so finely tuned that their theoretical counterparts are not likely to give much better predictions.

However, errors in prediction increase rapidly as the time difference from present to future is lengthened, and it is apparent that less credibility can be attached to long-range predictions derived from empirical relationships. As the difference in time is increased, the water-resource system becomes, in effect, an open system, and whatever principles and theories might have guided near-term predictions now become ineffective. The limitations of scientific knowledge are manifest in our general inability to delineate the boundaries of the system (i.e., to make the free-body cut) such that the system may be regarded as effectively closed.

The boundaries to be delineated are not solely of a physical kind, but economic, sociological, ecological, and political as well. Furthermore, depending on the context in which they are made (e.g., traditional water management versus water-quality management), the free-body cuts of the system are likely to differ. Although it may be that these cuts can be made independently (e.g., as exhibited by various "empirical" and "conceptual models"), the separately made predictions may be inconsistent in their implications for common environmental issues. A task of research is to gain an understanding of how to make such delineations that are not contradictory among the diverse interests of water-resource management and that are not inconsistent with hydrologic theory.

Correspondent with this task is the identification and quantification of the hazards and risks that would be imposed on society by the realization of specific events. That is, unless something is known about how the events translate into hazards (a subject of research), it is difficult to determine what it is that is to be predicted. The translation is understood reasonably well in the case of floods, a process that has been dealt with under the traditional paradigm, but less so in the case of droughts and in matters relating to water quality, both of which may be better suited for description under the active paradigm. With respect to the former, this is attested to by the many definitions of a drought that have been proposed, none of which have proven completely satisfactory, in particular those strictly limited to the quantity of the water mass. With respect to the latter, although there are many secondary facets to the concern about environmental degradation, such as aesthetics and economics, the primary concern is for public-health effects, as exemplified by the fishable-swimmable criteria stated in Public Law 92-500 (1972), by the Safe Drinking Water Act (Public Law 95-190, 1977) and by the Clean Water Act of 1977 (Public Law 95-217). Unfortunately, the state that constitutes "good health," or rather the factors that impair good health, are not well defined. Therefore, standards for myriad surrogate factors--water-quality constituents--are promulgated. Such standards are purposefully constructed to "err on the side of safety," because the translation from specific water-quality constituent levels to public-health effects is generally not well understood (Krenkel, 1979). Hence, predictions, however good they may be, about these surrogate variables are irrelevant unless the variables bear on the matter of concern, public health.

Predictions for water-resource management should be based on the consequences of hydrologic theory, which in turn seeks to describe and explain the connectedness between events. However, the systematically collected portion of the extant hydrologic data base has evolved largely in response to questions of management of the water resource rather than being motivated by scientific inquiry, a situation not unique to the United States (see e.g., Chapman, 1975). This data base has come to constitute the principal empirical (i.e., experimental) evidence for the examination of hydrologic theories because the past cannot be revisited and time between events of interest mitigates against the construction of another. This situation has forced a particular perception on hydrology. Attention has focused almost exclusively on the capability to make predictions that duplicate this data, rather than on the critical examination of the theories from whose consequences these predictions arise.

In the aforementioned context, matters of prediction cannot be disassociated from those of estimation, particularly in the case of recurrent events. Thomas (1978) noted, however, that the use of conventional statistical methods in hydrologic analysis may be called into question because hydrology is concerned with persistent processes and phenomena resulting from these processes are typified by probability distributions with long straggly tails that do not accord well with classical probability laws. (But such properties inferred from the extant hydrologic data base as persistence and straggly tail structure have heretofore not generally been considered in examinations of the consequences of hydrologic theories, e.g., such as those of global-water balances.) Thomas (1978) suggested that robust estimators (see Chapter 7; and Andrews *et al.*, 1972) offer promise in this respect as they can provide high-estimation efficiency while making few assumptions regarding the underlying distributions of the phenomena. Robust estimators are being used more widely in areas other than hydrology, although the *ad hoc* nature of their derivations is more disturbing to some (see, e.g., Box, 1979) than to others (see, e.g., Tukey, 1977).

A number of criteria (e.g., unbiasedness, consistency, minimax) have been proposed for deciding among alternative estimators. That of admissibility underlies current interest in Stein-rule estimators (see, e.g., Judge and Bock, 1978). Although derived in a non-Bayesian context, the Stein (1956) estimator of the location parameter may also be derived by empirical Bayes arguments, whereby the Bayes prior distribution is estimated from the data. The risk (i.e., the summed-mean-square error) of the empirical Bayes estimator is only slightly larger than that of the Bayes estimator with a perfectly specified prior distribution and protects against the larger Bayes risk that would arise through misspecification of the prior (see, e.g., Efron and Morris, 1973). Stein-rule estimators combine seemingly related phenomena to improve the estimate of the location parameter for a specific phenomenon. In this context, relatedness is defined in terms of the distance between the location parameters rather than the stricter characterization of dependence. Such a perception of relatedness underlies the simple idea of scaling observations by their mean, which in hydrology dates back at least to Fuller (1914).

Nothing in the above discussion negates the potential utility of historical data (e.g., ancient-flood marks), geomorphological data (e.g., current and past channel geometries), and geochronological data (e.g., tree rings and stratigraphy). These data have not been incorporated into the mainstream of hydrologic analysis. For example, there remains the promise of relating channel morphology and stratigraphy of fluvial deposits to extreme floods, as suggested by the description of major, flood-induced shifts in the course of the Yellow River in China given by Shen (1979). The realization of this potential--together with an understanding of the interplay of climate, geology, and hydrology--would provide a rationale for assessing extreme-flood risks. In any case, it is these data and not the systematically collected data that constitute the empirical basis by which predictions of future climates and their effects on regional water resources may be judged.

The attention on "duplicating" the data, rather than assessing it within the context of a hydrologic theory, has encouraged the explanation of the differences between data and predictions in terms of the peculiarities of each particular situation. Thus, increasingly detailed structures are proposed to encompass what are perceived to be complex situations, almost a collection of special cases. What is needed is a unifying theory, one that accounts for the disparate hydrologic situations.

SUMMARY AND CONCLUSIONS

Hydrology deals with the distribution and properties of water moving through the natural system, conceived to be the hydrologic cycle. But it must be recognized that human activity is inseparable from these concerns, both influencing and affected by them, both intentionally and inadvertently. This is in direct contrast to the traditional paradigm, which holds human enterprise to be a thing apart from the hydrologic cycle. The active paradigm to be adopted is one that recognizes that human activity is inherently a part of the natural system. Furthermore, the quality (physical, chemical, biological) of water is as central a concern in hydrology as the quantity of the water mass, affecting and being affected by it.

Water-resource management seeks to effect a change in the system through the use of structural and nonstructural measures to obtain a specific goal. Predictions of future and recurrent events enter into the risk-cost assessment on which water-management decisions are based. It is only after the fact that a prediction can be said to be right or wrong. However, predictions must be judged *a priori* to be good or bad and acted upon accordingly. The judgment is based on the credibility of predictions derived from their supporting scientific and philosophical arguments, which must distinguish between the scientific and transscientific questions asked with respect to the decision. These arguments, rooted in fundamental research, assure a firm scientific basis in the conduct of public affairs regarding water resources.

The economic, social, and environmental costs resulting from these decisions are a function of the errors in the predictions on which they rely. However, the principles that guide short-term predictions become ineffective as the difference in time from prediction to event becomes large: "Quantitative predictions of economic and social trends are made obsolete by qualitative changes in the rules of the game" (Dyson, 1979, p. 192). Therefore, research must change its focus from its present emphasis on resolving immediate operational problems, which at best can only refine near-term predictive ability, to furthering the study of hydrology on a more fundamental level under the active paradigm. Only from such study can long-range questions of water management be addressed.

Within the scope of research is the translation of events into identifiable hazards—events that remain outside the capability of traditional water management, notably those relating to problems of droughts and questions of water quality. The current societal inclination for a risk-free world in which natural laws are to be circumvented via legislative fiat does not mitigate the hard task of research. Precedence may rule that water-management systems in the United States will remain distinct, separated according to the different uses of the water resource. Therefore, when dealing with environmental and societal issues that are common to these diverse management interests, particular attention must be paid to the natural system in delineating the free-body cuts imposed.

The primary empirical (i.e., experimental) basis for the examination of hydrologic theories is the extant hydrologic data base, the systematic portion of which has evolved largely in response to questions of the management of the water resources rather than being motivated by scientific inquiry. This has forced a particular perception on hydrology as attention has focused almost exclusively on the capability to make predictions that duplicate that data. In this context, prediction cannot be disassociated from estimation. However, conventional statistical methods may be inappropriate to hydrologic analysis because of the persistent nature of the phenomena and the straggly tails typifying the distributions of the resulting phenomena. The inherent difficulties in discerning the underlying probability law argue for the adoption of robust estimators that are not overly dependent on assumptions regarding the law and for Stein-rule estimators that allow for an extended notion of hydrologic relatedness.

The purpose of hydrologic research is the advancement of scientific knowledge embodied in theory. However they are proposed, theories evolve and are accepted as their deduced consequences withstand examination in light of empirical evidence. But, however satisfactorily the experimental data accord with the consequences, the acceptance of the theories from which they are deduced cannot be fully decided based on the empirical evidence alone. The necessary connections between sensed changes in the states of a system cannot be inferred in a philosophical vacuum so that the philosophical aspects of science cannot be ignored.

Within narrow temporal and spatial dimensions, characteristic of more immediate operational water-management decisions for which quantitative, short-term predictions are to be obtained (e.g., those of runoff from a specific storm) such that a free-body cut of the natural system can effectively be made in terms of specific physical boundaries, strong causality or notions of causal determination may be invoked. However, regardless of whether the traditional or active paradigm is adopted, within the broader spatial and temporal dimensions of the natural system, more introspection is needed to interpret the scientific issues of hydrology within a philosophical context.

Hydrology does not seek to describe the history of any single molecule as it moves through the cycle; rather, its concern is with collections of molecules. (This is reflected by the fact that measurements of hydrologic phenomena do not give precise information about any one molecule, be it a water molecule or a "pollutant," but rather imply only the probability of finding some collection of such molecules at a particular location.) It is suggested that operationally there may be a little recourse to the use of probability in describing hydrologic processes regardless of the nature of the state variables.

Under the active paradigm, causal determination cannot strictly apply as man is inherent to the system. Strong causality also cannot be supported because the system of interest cannot be observed without interfering with or disturbing it. However, particular concepts of determination or causality cannot be imposed externally on hydrology: Adoption will occur if they are successful in providing an explanation of the empirical evidence that is acceptable to the scientific community. Furthermore, this explanation of the necessary connection between changes in the state of the system in time, i.e., causality, cannot be strictly inferred from the empirical evidence alone.

These statements should not be viewed as posing a hindrance to hydrologic research: Advances in modern physics did not await consistent philosophical support. Rather, they should be regarded as optimistic that hydrology can evolve in accordance with a more comprehensive accounting of the hydrologic cycle under the active paradigm, limited neither by "systems" descriptions nor by expedient "conceptual constructs." To be kept in perspective is that the principal purpose of research is not the development of models nor the enhancement of predictive capability per se; rather, it is the furthering of scientific knowledge, as embodied in the description and explanation of hydrologic connectedness from which consequences may be deduced in a manner consistent with the principle of determination. It is from this knowledge that support is drawn for the predictions on which water-management decisions are conditioned, assuring a scientific foundation in the conduct of public affairs regarding water resources.

REFERENCES

Andrews, D. F., *et al.* (1972). *Robust Estimation of Location: Survey and Advances*, Princeton University Press, Princeton, N.J.

Bannister, E. N. (1979). Impact of road networks on southeastern Michigan lakeshore drainage, *Water Resour. Res. 15*, 1515-1520.

Bohm, D. (1957). *Causality and Chance in Modern Physics*, Harper & Row, New York.

Box, G. E. P. (1979). *Sampling and Bayes' Inference in the Advancement of Learning*, MRC Technical Survey Report No. 1969, Mathematic Research Center, University of Wisconsin, Madison, Wis., 28 pp.

Bunge, M. (1979). *Causality and Modern Science*, Dover Publications, Inc., New York.

Chapman, T. G. (1975). Trends in catchment modeling, in *Prediction in Catchment Hydrology*, T. G. Chapman and F. X. Dunin, eds., Australian Academy of Science, The Griffin Press, Netley, South Australia, pp. 459-482.

DeMillo, R. A., R. J. Lipton, and A. J. Perlis (1979). Social processes and proof of theorems and programs, *Communications of the ACM 22*, 271-280.

Dyson, F. (1979). *Disturbing the Universe*, Harper & Row, New York.

Efron, B., and C. Morris (1973). Stein's estimation rule and its competitors--an empirical Bayes approach, *J. Am. Statistical Assoc. 68*, 117-130.

Feyerabend, P. (1963). How to be a good empiricist--a plea for tolerance in matters epistemological, in *Philosophy of Science: The Delaware Seminar, Vol. II*, B. Baumrin, ed., John Wiley & Sons, New York.

Feynman, R. P., R. B. Leighton, and M. Sands (1965). *The Feynman Lectures on Physics*, Addison-Wesley Publishing Co., Reading, Mass.

Fiering, M. B., and C. S. Holling (1974). Management and standards for perturbed ecosystems, *Agro-ecosystems 1*, 301.

Fuller, W. E. (1914). Flood flows, *Trans. Am. Soc. Civil Eng. 77*, 564-694.

Heisenberg, W. (1930). *The Physical Principles of the Quantum Theory*, University of Chicago Press, Chicago, Ill.

Hubbert, M. K. (1976). Energy resources: A scientific and cultural dilemma, *Bull. Assoc. Eng. Geol. 12*, 81-124.

Judge, G. C., and M. E. Bock (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, North-Holland Publishing Company, New York.

Kirby, W. (1978). Predictions of streamflow hazards, in *Geophysical Predictions*, Geophysics Study Committee, National Research Council, National Academy of Sciences, Washington, D.C., pp. 202-215.

Knapp, A. H., P. J. Le B. Williams, and I. Tyler (1979). Contribution of votatile petroleum hydrocarbons to the organic carbon budget of an estuary, *Nature 279*, 517-519.

Krenkel, P. A. (1979). Problems in the establishment of water quality criteria, *J. Water Pollut. Control Fed. 51*, 2168-2188.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago, Ill.

Likens, G. E., R. F. Wright, J. N. Galloway, and T. J. Butler (1979). Acid rain, *Sci. Am. 241*, 43-51.

Mandlebrot, B. B. (1980). *Stationarity, Ergodic, Settled and Vagrant Time Series and the Hurst Effect*, (unpublished manuscript), 10 pp.

Nace, R. L. (1960). *Water Management, Agriculture, and Groundwater Supplies*, U.S. Geol. Surv. Circ. 415, 12 pp.

NRC Geophysics Study Committee (1977a). *Climate, Climatic Change, and Water Supply*, National Academy of Sciences, Washington, D.C., 132 pp.

NRC Geophysics Study Committee (1977b). *Energy and Climate*, National Academy of Sciences, Washington, D.C., 172 pp.

NRC Geophysics Study Committee (1978). *Geophysical Predictions*, National Academy of Sciences, Washington, D.C., 213 pp.

Nriagu, J. O. (1979). Global inventory of natural and anthropogenic emissions of trace metals to the atmosphere, *Nature 279*, 409-411.

Northrop, F. S. C. (1958). Introduction, in *Physics and Philosophy*, authored by W. Heisenberg, Harper & Row, New York.

Popper, K. R. (1959). *The Logic of Scientific Discovery*, Hutchison and Co., Ltd., London, England (translated from the 1935 German edition, *Logik der Forschung*, J. Springer, Vienna, Austria).

President's Commission on the Accident at Three Mile Island (1979). *Report on the Accident at Three Mile Island*, U.S. Government Printing Office, November 22, 1979, Washington, D.C., 201 pp.

Shen, H. W. (1979). Some notes on the Yellow River, *EOS Trans. Am. Geophys. Union 60*, 645-647.

Smart, J. S. (1979). Determinism and randomness in fluvial geomorphology, *EOS Trans. Am. Geophys. Union 60*, 651-655.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multi-variate normal distribution, in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, Calif., pp. 197-206.

Suppe, F., (1977). The search for philosophic understanding of scientific theories, in *The Structure of Scientific Theories*, F. Suppe, ed., University of Illinois Press, Urbana, Ill.

Thomas, H. A., Jr. (1978). Acceptance: Robert E. Horton Medal, *EOS Trans. Am. Geophys. Union 59*, 793.

Toulmin, S. (1953). *The Philosophy of Science: An Introduction*, Hutchison and Co., Ltd., London, England.

Tukey, J. W. (1977). *Exploratory Data Analysis*, Addison-Wesley Publishing Company, Reading, Mass.

Turco, R. P., R. C. Whitten, O. B. Toon, J. B. Pollack, and P. Hamill (1980). OCS, stratospheric aerosols and climate, *Nature 283*, 283-285.

United Nations Educational, Scientific, and Cultural Organization--UNESCO (1978). World Water Balances and Water Resources of the Earth, UNESCO Press, Paris, 663 pp. (translated from the 1974 Russian monograph prepared by the U.S.S.R. Committee for the International Hydrologic Decade under the Chief Administrator of the Hydrometeorological Service under the Council of Ministers of the U.S.S.R.).

U.S. Congress (1972). *Federal Water Pollution Control Act Amendments of 1972*, Public Law 92-500, 92nd Congress, October 18, 1972.

U.S. Congress (1973). *Endangered Species Act*, Public Law 93-205, 93rd Congress, December 28, 1973.

U.S. Congress (1974). *Water Resources Development Act of 1974*, Public Law 93-251, 93rd Congress, March 7, 1974.

U.S. Congress (1977). *Surface Mining Control and Reclamation Act of 1977*, Public Law 95-87, 95th Congress, August 3, 1977.

U.S. Congress (1977). *Safe Drinking Water Amendments of 1977*, Public Law 95-190, 95th Congress, November 16, 1977.

U.S. Congress (1977). *Clean Water Act of 1977*, Public Law 95-217, 95th Congress, December 27, 1977.

U.S. Congress (1978). *National Climate Program Act*, Public Law 95-367, 95th Congress, September 17, 1978.

Weinberg, A. M. (1972). Science and trans-science, *Minerva X*, 209-222.

White, G. W. (1968). John Keill's view of the hydrologic cycle, 1698, *Water Resour. Res. 4*, 1371-1374.

Zenchelsky, S., and M. Youssegi (1979). Natural organic atmospheric aerosols of terrestrial origin, *Rev. Geophys. Space Phys. 17*, 459-464.