IIASA PROCEEDINGS SERIES

Real-Time Forecasting/Control of Water Resource Systems

Selected Papers from an IIASA Workshop

Eric F. Wood, Editor
With the Assistance of András Szöllösi-Nagy

International Institute for Applied Systems Analysis

IIASA PROCEEDINGS SERIES

VOLUME 1 — CARBON DIOXIDE, CLIMATE AND SOCIETY
Proceedings of an IIASA Workshop Cosponsored by WMO, UNEP, and SCOPE, February 21–24, 1978
Jill Williams, Editor

VOLUME 2 — SARUM AND MRI: DESCRIPTION AND COMPARISON OF A WORLD MODEL AND A NATIONAL MODEL
Proceedings of the Fourth IIASA Symposium on Global Modeling, September 20–23, 1976
Gerhart Bruckmann, Editor

VOLUME 3 — NONSMOOTH OPTIMIZATION
Proceedings of an IIASA Workshop, March 28–April 8, 1977
Claude Lemarechal and Robert Mifflin, Editors

VOLUME 4 — PEST MANAGEMENT
Proceedings of an International Conference, October 25–29, 1976
G.A. Norton and C.S. Holling, Editors

VOLUME 5 — METHODS AND MODELS FOR ASSESSING ENERGY RESOURCES
First IIASA Conference on Energy Resources, May 20–21, 1975
Michel Grenon, Editor

VOLUME 6 — FUTURE COAL SUPPLY FOR THE WORLD ENERGY BALANCE
Third IIASA Conference on Energy Resources, November 28–December 2, 1977
Michel Grenon, Editor

VOLUME 7 — THE SHINKANSEN HIGH-SPEED RAIL NETWORK OF JAPAN
Proceedings of an IIASA Conference, June 27–30, 1977
A. Straszak and R. Tuch, Editors

VOLUME 8 — REAL-TIME FORECASTING/CONTROL OF WATER RESOURCE SYSTEMS
Selected Papers from an IIASA Workshop, October 18–20, 1976
Eric F. Wood, Editor, with the Assistance of András Szöllösi-Nagy

VOLUME 9 — INPUT–OUTPUT APPROACHES IN GLOBAL MODELING
Proceedings of the Fifth IIASA Symposium on Global Modeling, September 26–29, 1977
Gerhart Bruckmann, Editor

VOLUME 10 — CLIMATIC CONSTRAINTS AND HUMAN ACTIVITIES
Selected Papers from an IIASA Task Force Meeting, February 4–6, 1980
Jesse Ausubel and Asit K. Biswas, Editors

IIASA PROCEEDINGS SERIES

Volume 8

Real-Time Forecasting/Control of Water Resource Systems

REAL-TIME FORECASTING/CONTROL OF WATER RESOURCE SYSTEMS

Selected Papers from an IIASA Workshop,
October 18–21, 1976

Edited by
ERIC F. WOOD
International Institute for Applied Systems Analysis
and
Department of Civil Engineering, Princeton University

With the Assistance of
ANDRÁS SZÖLLÖSI-NAGY
International Institute for Applied Systems Analysis
and
Research Institute for Water Resources Development, Budapest

PERGAMON PRESS

OXFORD · NEW YORK · TORONTO · SYDNEY · PARIS · FRANKFURT

U.K.	Pergamon Press Ltd., Headington Hill Hall, Oxford OX3 0BW, England
U.S.A.	Pergamon Press Inc., Maxwell House, Fairview Park, Elmsford, New York 10523, U.S.A.
CANADA	Pergamon of Canada, Suite 104, 150 Consumers Road, Willowdale, Ontario M2J 1P9, Canada
AUSTRALIA	Pergamon Press (Aust.) Pty. Ltd., P.O. Box 544, Potts Point, N.S.W. 2011, Australia
FRANCE	Pergamon Press SARL, 24 rue des Ecoles, 75240 Paris, Cedex 05, France
FEDERAL REPUBLIC OF GERMANY	Pergamon Press GmbH, 6242 Kronberg-Taunus, Hammerweg 6, Federal Republic of Germany

First edition 1980

British Library Cataloguing in Publication Data

Real-time forecasting/control of water resource systems. - (International Institute for Applied Systems Analysis. IIASA proceedings series; 8).
1. Water resources development - Data processing - Congresses
2. Real-time data processing - Congresses
I. Wood, Eric F II. Szöllösi-Nagy, András
III. Series
333.9'1'002854404 TC409 79-41070

ISBN 0-08-024486-6

Printed and bound in Great Britain by William Clowes (Beccles) Limited, Beccles and London

PREFACE

In the day-to-day operation of water resource systems, forecasting is extremely important for establishing optimal design and control strategies. In the past few years, considerable attention has been paid to real-time control of water resource systems in an on-line operation mode. Several techniques have been developed to approach this important practical problem. However, there were, and still are, some basic problems needing further clarification. For example, the nonlinear, spatial, stochastic characteristics of water resource systems cause problems in the implementation of algorithms adapted from other fields. Progress in overcoming these problems has been made throughout the world. At the International Institute for Applied Systems Analysis (IIASA), the Water Resource Project had a research effort devoted to the methodology of real-time forecasting and control of water resource systems. Research was being done in several places, but without mutual cooperation. It was felt that a small international workshop would serve as an important forum to exchange experiences and ideas. Thus, IIASA hosted a workshop from October 18 to 20, 1976, in which 52 scientists from 17 countries participated. Twenty-six papers were presented, of which 17 are included in this volume. The papers in the workshop and in this volume were divided into four areas: overviews of forecasting methodologies, forecasting of hydrological systems, control, and operation experiences. The work in these proceedings represents an important interchange of ideas and approaches, and forms the basis for future developments in the area of real-time forecasting and control of water resource systems.

CONTENTS

1 Introduction

This book represents the edited proceedings of a small workshop held at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, from October 18 to 20, 1976. The workshop was held within the framework of the research study "Methodology of Real-Time Forecasting and Control of Water Resource Systems".

Current operational forecasting procedures for streamflows have developed from deterministic hydrology and are based on either index catchment models or conceptual catchment models. Such models tend to simulate the basin response to hydrologic events and do not fully utilize information collected during an event.

Motivation for the research task and subsequent workshop was twofold: In the day-to-day management of water resources, greater demands are being made on systems to increase yields or provide better flood protection through better operation rather than by expansion. Poor management of reservoirs can lead to the waste of over 20 per cent of the available reservoir supply (that is, water that can be delivered to users) because of needless discharges. This waste could be reduced by better forecasts of demands and downstream supplies. The most cost-effective procedure for reducing flood damage in the United States (which still runs at about $1.5 billion annually and is increasing at about 6% per year in real terms) and many other countries is through better forecasting techniques. Therefore, our first motivation came from the knowledge that improved forecasting and control can lead to the better use of water resource systems. The interest in the last few years in operational, on-line control systems supports this claim.

The second motivation for our research was a belief that currently applied hydrologic forecasting and control procedures are not the best available. During a flood, data arrives at the forecast center. This in-coming data contains valuable information that the forecasting procedures and models must extract and utilize. It is not evident that current operational models, like those presented in Part Three, are formulated and developed to process this in-coming data adequately.

Lately, considerable interest has developed in forecasting techniques based on state-space models. Most of the practical applications have been in the aerospace field, and industrial applications are becoming more common. Recently, forecasting algorithms, based on statistical estimation techniques such as the Kalman filter, have been developed by hydrologists, water quality modelers and meteorologists. Many important difficulties, both of a practical and of a methodological nature, remain to be resolved before operational state-space models can be applied.

The subject of the IIASA workshop and this book encompasses recent developments of statistically estimated models for hydrologic forecasting in real-time for control. By real-time, we mean a mode of operation in which a forecast is made for a given time horizon (which may be a few hours for flood discharges from a catchment that responds quickly, to a few days for water demands or supplies, to a month or more for groundwater systems) in order to make a control decision. During the forecast interval, data is collected and processed and new forecasts are issued so that the control action can be modified.

STATE ESTIMATION

This volume is concerned with developing state-space models of hydrologic systems. The concept of the "state" of the system is introduced as a mathematical convenience in order to use the notions of causality and internal structure in the description of the state (Casti 1977). The states need not represent a physically measurable quantity. The only variables that do have physical representation are the inputs and outputs, since these can be observed and measured.

The concept of the state is important since it permits models for the system dynamics (often called the state equation) that describe the response of system states to inputs, to be separated from models that relate the observed outputs to the system states through an output equation (or measurement system).

In hydrologic analysis, the state may refer to the "correct" value of a measurable quantity, for example, river levels, subcatchment discharges, precipitation rates, or reservoir volumes. A measurement system would provide, through time, observations on these variables by using sensors or other measuring devices. The state variable can also be a system parameter like the friction coefficient for losses in overland flow or the permeability coefficient of a soil. These variables cannot be directly observed but can be related to observed outputs; discharge rates from overland flow are a function of the friction coefficient and well drawdowns in groundwater systems are a function of soil permeability (through an output function).

The concept of states and state-space models allows great flexibility in describing and modeling systems. The state equation models the behavior of the system by a finite dimensional Markov process that has been "noise corrupted", that is, it is subject to random error disturbances. Because of these random disturbances, the states are random variables whose probabilistic properties are related to the probabilistic properties of the error disturbances.

The measurement system observes (or measures) the precipitation in a catchment, the stage of a river, the groundwater and so forth. The measurements of the output variables are also in error due to the complex nature of the hydrologic systems. For example, areal precipitation measurements will be in error because only a small number of point measurements are taken and because the measurements may be in error due to shortcomings in equipment. The (measured) outputs are related to the state variables through the measurement (or output) equation that accounts for the measurement errors through an additive error term.

Thus, the state-space model approach has two sets of equations. One, the state equation, is a dynamic model that describes the behavior of the system over time and is subject to modeling errors; the other is the measurement or output equation that relates the observed output (from the measurement system) to the state variables of the system.

Using both the state equation and the available measurements, it is possible to estimate the state of the system. This procedure is regarded as state estimation. Given measurement up to and including time t, three kinds of problems in estimating

the state of the system, $x(\tau)$, at time τ, are of interest: filtering, estimating $x(\tau)$ when $\tau = t$; smoothing, estimating $x(\tau)$ when $\tau < t$; and forecasting, estimating $x(\tau)$ when $\tau > t$. The estimator for $x(\tau)$ is often denoted as $\hat{x}(\tau|t)$, showing that the estimator uses measurements up to and including time t. The most commonly applied filtering technique was developed by Kalman (1960) for estimating the state of a linear system with known system dynamics and knowledge of the probability properties of the error distributions.

Figure 1.1 illustrates the procedure of system measurement and estimation. The survey paper by Mehra (Chapter 2) gives a review of state-space models and discusses a number of estimation techniques. The introduction to Part One of the book includes a derivation of the Kalman filter since many papers in that section apply the technique. Filtering algorithms are emphasized because optimal forecasting is based on filtering (Gelb 1974).

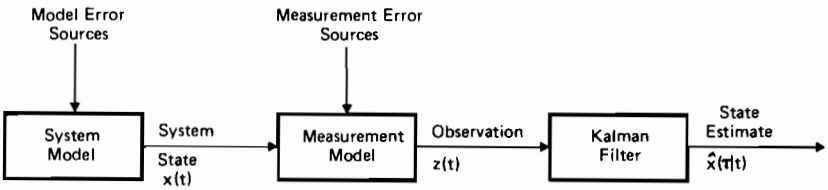


Fig. 1.1 Procedure of system measurement and estimation (after Gelb, 1974).

APPLICATION OF STATE ESTIMATION

For hydrologic systems, a number of practical problems arise when the Kalman filter is applied. The Kalman filter provides an algorithm to process the measurement data given, a (linear) model of the state dynamics, the probabilistic properties of the model error and measurement error terms, and initial conditions consisting of an (unbiased) estimate of the state at $t = 0$, $x(0)$, and the error covariance matrix of this estimator. The dynamics of hydrologic systems are often nonlinear (or approximately linear), estimators for the parameter values are unreliable, and the magnitude of the noise variances in the state equation and the measurement equation are unknown.

The last problem is of critical importance because of the forecasting and updating property of the Kalman filter algorithm. Prior to the processing of the measurement at time t, an estimate of the state at time t (based on past measurements) is made. The uncertainty of this estimate depends partially on the variance of the noise terms that are influencing the state equation. After the (noise corrupted) output variables are measured with the measurement equation, the estimate of the state based on past measurements is updated to include the current measurement. The updated estimate of the state is a weighted estimate of the uncertain prior estimate of the state and the uncertain measurements, where the weight is a function of the relative magnitudes of the variances of the state equation and measurement equation noise terms. If the variance of the state equation noise term is much smaller than the variance of the measurement equation noise term, then less weight will be given to the new measurement in updating the state estimator.

Therefore, before the correct Kalman filter can be applied to hydrologic problems, the problems of model structure, parameter values, and probabilistic properties of the noise terms must be resolved. Part One addresses these issues specifically.

The aim of a forecasting algorithm is to fully utilize the information from the incoming data. If it is fully utilized, the one-step-ahead residual errors will be

uncorrelated and there will be cross-correlation between forecast error and (prewhitened) input. The first property implies that the noise terms (in the state and measurement equations) have been correctly modeled and estimated and the second property implies adequacy in the structure and parameter values of the transfer function between inputs and outputs (Box and Jenkins 1970).

The problem of determining the structure of the model and estimating the parameters is often referred to as model identification. Part One of the volume is concerned with this problem and a number of techniques are applied to models of hydrologic systems. The problem of estimating probabilistic properties of the noise terms (especially their variances) has not been adequately solved for those models applied to hydrologic systems. Pioneering work by Mehra in this field is described in Chapter 2 and a recursive algorithm (Wood and Szöllösi-Nagy 1978) has been applied to real-world data (results were presented at the workshop but are not reported in this volume).

OPTIMAL CONTROL

The aim of water resource management is to control the system in a manner that is "most preferred." To determine what is meant by most preferred would require the determination of appropriate cost, benefit, and loss functions for the water resource system. This is a difficult task and one that is not addressed in this volume. It will suffice, for our discussion, to assume that the appropriate functions exist and have been adequately estimated.

The control strategy determines, over time, target levels for the preferred operation of the system. To develop meaningful control strategies for hydrologic systems, the stochastic nature of the system must be taken into account. Stochastic control theory is concerned with developing strategies for systems that are subject to stochastic disturbances.

Control requires the adjustment of a variable whose effect on the state variable is well known through the state equation; the purpose of the adjustment is to maintain the established target level. The filtered and forecasted values of the state equations are very useful for control problems. In fact, for certain types of problems, these filtered and forecasted states of the hydrologic system can be utilized in finding control policies when stochastic disturbances are present.

A fuller discussion of the important connection between the state estimation problem and the control problem of stochastic hydrologic systems is presented in the introduction to Part Two, Control of Water Resource Systems. Applications presented in Part Two are for reservoir control, waste water treatment, and the determination of in-stream water quality.

ARRANGEMENT OF THE BOOK

The papers have been grouped in such a manner that readers with varying interests and mathematical backgrounds may approach the book in some logical manner. The emphasis of the book is on hydrologic models that fit into the state-space formulation. The papers describe the behavior of a dynamic system by a finite-dimensional Markov process whose state is the output variable of the stochastic differential or difference equation describing this process. This formulation is subsequently embedded in a state-estimation filtering algorithm usually of a recursive nature, for the processing of data from the real-time, sequential measuring system.

For the reader who is unfamiliar with many of the concepts of state-estimation

theory, Mehra's survey paper (Chapter 2) presents some of the fundamental problems and approaches to state-space models, system identification, and forecasting. Almost all of the major techniques represented in Part One are discussed.

Part One is the focus of the work and it is here that hydrologic problems are formulated and analyzed within the state-estimation framework. This portion of the volume are represented by Chapters 3 through 10 and can be divided into two groups. The first group of papers focuses on system identification and parameter estimation techniques. These papers describe many state-of-the-art techniques for modeling and forecasting hydrologic phenomena. The second group of papers emphasize the state estimation problem of hydrologic systems, and a variety of approaches besides the Kalman filter approach are discussed.

Part One should be of great interest to both the researchers interested in applying new methodology, and the operational or field hydrologists who want to evaluate the potential of the proposed techniques. The paper by McLaughlin (Chapter 7) is especially effective in the latter respect.

Part Two considers the problem of the control of water resource systems. The paper by Beck (Chapter 11) relates many of the topics discussed elsewhere in the book to the problems in the control of hydrologic systems and is an effective introduction to readers who are not familiar with the concepts. Many hydrologists may be able to profit from the formulation of problems of reservoir control (Chapter 13), in-stream water quality (Chapters 11 and 14) and waste treatment (Chapter 11). A more technical discussion of the relation between forecasting and control of stochastic systems is provided by Szöllösi-Nagy and Wood (Chapter 12) and may be of interest to research hydrologists.

Both Parts One and Two have introductions in which the basic aims and concepts developed within that section are discussed. The introductory material also includes a literature review to assist those readers who are not familiar with the historical development of the subject.

One of the aims of the workshop was to relate the state-estimation techniques to operational forecasting. We attempted to do this by having, at the workshop, papers describing some operational forecasting experiences in various countries. These papers are presented in Part Three and represent an important link between illustrative applications (of an academic type) and operational or field forecasting. The research hydrologist must understand the problems inherent in large, complex systems and the constraints that exist. Similarly, the field hydrologist, who is probably using techniques described in Part Three, should attempt to apply the techniques of Part One (or at least support their application) to see if improved forecasting can be achieved.

It is possible for the hydrologist who usually works with deterministic models to relate his methodology to the state-space modeling approach if he recognizes that his approach will give identical results to the state-space model when only the state equation of the system dynamics model is used for forecasting. On the other hand, the state-space approach allows one to estimate and forecast the state of the (stochastically disturbed) hydrologic system with the use of a filter algorithm. This procedure will lead to results that are equal or better than a procedure based on the state equation model alone. This invites further investigation of the state-space approach. The editors hope that this volume will encourage many hydrologists to consider the techniques described and to apply those that appear promising to the day-to-day management of water resource systems.

REFERENCES

Box, G. E. P., and G. Jenkins (1970). Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco.

Casti, J. L. (1977). Dynamical Systems and Their Application: Linear Theory. Academic Press, New York.

Gelb, A. (1974). Applied Optimal Estimation. M. I. T. Press, Cambridge, Massachusetts.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. J. Basic Eng. Trans. ASME 82:35-44.

Wood, E. F., and A. Szöllösi-Nagy (1978). An adaptive algorithm for analyzing short-term structural and parameter changes in hydrologic prediction models. Water Resour. Res. 14(4): 577-581.

2 A Survey of Time Series Modeling and Forecasting Methodology

Raman K. Mehra

This paper reports on certain recent developments in time series analysis, system identification, and forecasting in the hope that they may find applications in hydrology. A number of concepts discussed here may be familiar to hydrologists, though the details of the state-space formulation and the associated results may not be widely known. The emphasis will be on model building using real data; there will generally be a specific objective for modeling such as forecasting, control, or simulation. Time series analysis is too vast to be covered in a single paper; the selection of material, no doubt, reflects the author's own interests and viewpoints on the subject.

HISTORICAL BACKGROUND

The systematic development of time series analysis started with the work of Yule (1927), who showed that Wolfer's sunspot data could be represented satisfactorily by a second-order autoregressive (AR) model. Wold (1938) showed that any non-deterministic stationary time series can be decomposed into a moving average of independent random variables known as "innovations" or the one-step-ahead prediction errors.

The estimation of parameters in AR and other linear difference equation models was considered in depth by Mann and Wald (1943); and Whittle, in a series of ground-breaking papers (1951, 1952, 1952a, 1954, 1954a, 1954b), explored various aspects of time series analysis, including spectral representation, maximum likelihood estimation of parameters, hypothesis testing, tests of fit, and random processes on a plane. From 1950 to 1970 several developments took place in the field of spectral estimation and nonparametric modeling of time series; they are reported by Bartlett (1966), Hannan (1960), Blackman and Tukey (1959), Parzen (1957), and Jenkins and Watts (1968). The importance and usefulness of parametric modeling was re-emphasized by Box and Jenkins (1970), who also showed that a number of practical nonstationary time series could be modeled by using integrated and seasonable autoregressive moving average (ARMA) models.

The work reported here is based on research sponsored by the Office of Naval Research under Contract #N00014-76-C-1024.

With the exception of the work of Quenouille (1957), the literature of time series analysis until 1970 was largely confined to scalar time series. Since 1970, however, multiple time series analysis has received increasing attention in both the statistical and control literature (see Hannan 1970, Akaike 1974, Kashyap and Rao 1976). The state-space theory of linear systems, developed during the early 1960s by Kalman (1960) and others, is described briefly below. Since the theory is applicable to both physical and black box models, one can present a unified treatment of modeling and forecasting methologies.

Further references on the development of time series analysis may be found in Wold (1965), IEEE (1974), and Mehra and Lainiotis (1976).

TIME SERIES ANALYSIS AND ITS APPLICATIONS

System identification and time series analysis involve developing mathematical models and identifying relevant physical and mathematical characteristics of systems using noisy operational input output data. The procedure may be divided into the four subproblems shown schematically in Figure 2.1:

- Design of experiment
- Model structure determination
- Parameter estimation
- Model verification and comparison

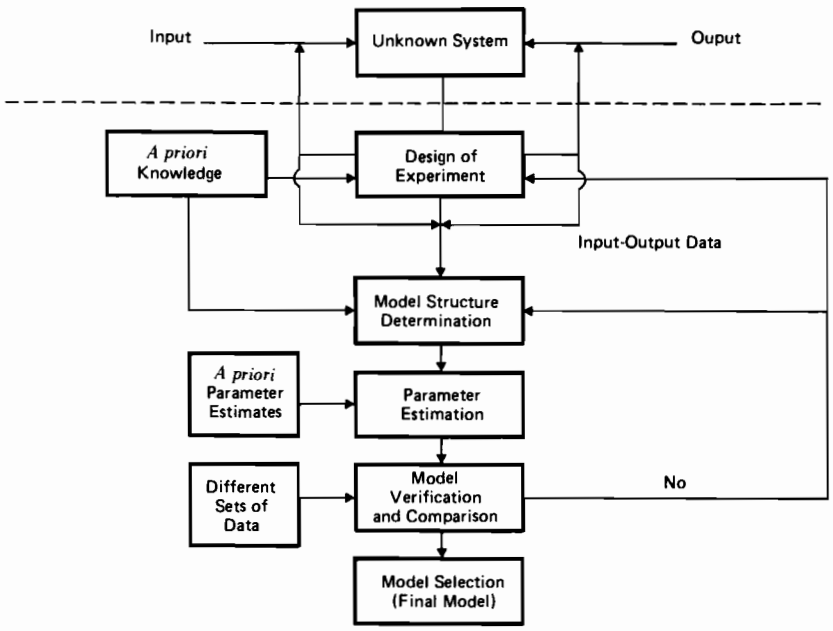


Fig. 2.1 Steps in system modeling and identification (from Mehra and Lainiotis, 1976).

The model-building process involves several passes through the above four steps. Implicit in the model verification step is the objective of identification. For example, if the model is being developed for real-time forecasting of river flow

one day ahead, then the model verification and selection is based on the ability of the model to provide the best one-day-ahead forecasts. Several other criteria, such as simplicity and the relation of the variables to reality, will be discussed in later sections.

APPLICATIONS

According to Parzen (1976), one may distinguish six basic applications of time series analysis:

- Forecasting (or extrapolation)
- Spectral analysis (or interpolation by harmonics)
- Parameterization (or data compression)
- Intervention analysis (detection of significant changes in forecasts or parameters)
- Filtering and smoothing (signal-plus-noise decomposition)
- Control

For hydrology, and fields similar to hydrology, there are two other applications:

- Simulation or generation of synthetic time series for design purposes
- Understanding of the internal mechanism of a process by the estimation of parameters in mechanistic (rather than black box) models

Some major areas of application system identification and time series analysis outside hydrology are:

- Industrial processes: paper mills (Aström 1967); basic oxygen furnace (BOF) (Mehra and Wells 1971); aircraft (Mehra and Tyler 1973)
- Energy systems: nuclear reactors (Olsson 1976); turboalternators (Jenkins and Watts 1968); electricity demand forecasting (Galiana et al. 1974)
- Biomedical systems: EKG and EEG analysis (Gustafson et al. 1977, Gersch 1972); respiratory models (Swanson 1972); human operator modeling (Phatak et al. 1975); compartmental models for drug distribution (see survey by Beckey 1976 for other applications and for problems peculiar to biomedical modeling)
- Econometrics: (Theil 1971) and business forecasting (Management Science 1977)

MODELS USED IN SYSTEM IDENTIFICATION AND TIME SERIES ANALYSIS

General State Vector Model

A general state vector model can be used to represent mechanistic, conceptual, or black box models. A state vector model is specified in terms of (a) three vectors of input, output, and state variables; (b) a rule for transformation of the state vector from one instant in time to the next; (c) a relationship between the output variable and the input and state variables; (d) the initial state; and (e) joint

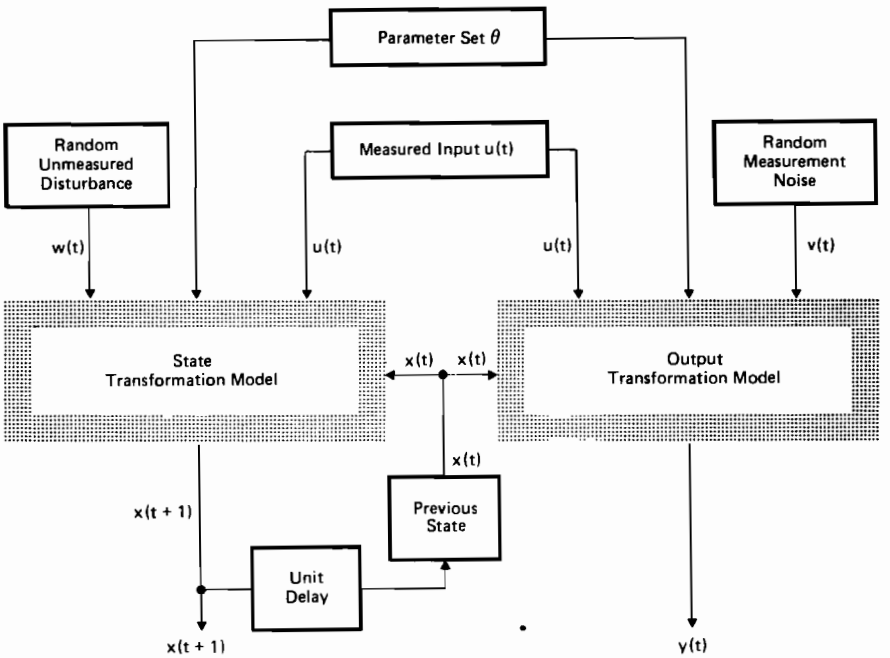


Fig. 2.2 State vector model.

statistics of unknown parameters and random variables (Fig. 2.2). Mathematically, in discrete time

$$x(t+1) = f(x(t),u(t),\Theta,t) + w(t) \qquad (2.1)$$

$$y(t) = h(x(t),u(t),\Theta,t) + v(t) \qquad (2.2)$$

$$t = 0,1,2...$$

where x(t) is an n-dimensional state vector, u(t) is an r-dimensional input vector, w(t) is an n-dimensional process noise vector, Θ is an m-dimensional parameter vector, y(t) is a p-dimensional output vector, and v(t) is a p-dimensional measurement noise vector. The noise vectors w(t), v(t) are usually assumed to be uncorrelated white noise sequences with known distributions. Similarly, the distribution of x(0) is assumed known.

All mathematical models, including the state vector model, are only approximations of reality. It has been shown by Szöllösi-Nagy (1976) that a number of models used in hydrology can be written in the form of Eqs. (2.1) and (2.2). If the model is mechanistic or conceptual, the state vector x(t) has a close relation to reality. In black box or time series models, the state vector need have no real correlation; however, it possesses a mathematical meaning that may be simply stated, for heuristic purposes, as follows:

State vector x(t) collects information, from the past and present behavior of the system, that is sufficient to predict its future behavior for a specified input sequence. For mathematically rigorous definitions of the state vector, see Kalman et al. (1969) and Akaike (1974).

The linear form of (2.1) and (2.2) is the Gauss-Markov model, written as:

$$x(t+1) = \Phi.x(t) + G.u(t) + \Gamma.w(t) \tag{2.3}$$

$$y(t) = H.x(t) + v(t) \tag{2.4}$$

The parameter vector θ is included in Φ and H, and y(t) is no longer dependent on the input vector u(t). The vectors w(t) and v(t) are assumed to be Gaussian white noise (GWN) sequences with zero mean and covariances Q and R. The initial state x(0) is normally distributed with mean $\hat{x}_0$ and covariance P_0. The matrices Φ, G, H, Γ, Q, R, and P_0 are deterministic but may be time varying. The main advantage of the Gauss-Markov model is that the mean, covariance, and correlation functions for x(t) and y(t) can be computed recursively by solving a set of first-order vector-difference equations. The general equations may be found in Bryson and Ho (1969) and are summarized below:

Mean: Let $\bar{x}(t) = E[x(t)]$. Then

$$\bar{x}(t+1) = \Phi\, \bar{x}(t) + G\, u(t) \tag{2.5}$$

$$\bar{y}(t) = H\, \bar{x}(t) \tag{2.6}$$

$$\bar{x}(0) = \bar{x}_0 \tag{2.7}$$

Covariance: Let $\Sigma_x(t) = E[(x(t) - \bar{x}(t))(x(t) - \bar{x}(t))^T]$. Then

$$\Sigma_x(t+1) = \Phi\, \Sigma_x(t)\, \Phi^T + \Gamma Q \Gamma^T \tag{2.8}$$

$$\Sigma_y(t) = H\, \Sigma_x(t)\, H^T \tag{2.9}$$

$$\Sigma_x(0) = \Sigma_0 \tag{2.10}$$

Correlation function: For the time-invariant case, let

$$C_x(k) = E[(x(t+k) - \bar{x}(t+k))(x(t) - \bar{x}(t))^T] .$$

Then

$$C_x(k) = \Phi^k\, \Sigma_x \qquad k \geq 0 \tag{2.11}$$

$$C_y(k) = \begin{cases} H\, \Sigma_x\, H^T + R & k = 0 \\ H\, \Phi^k\, \Sigma_x\, H^T & k > 0 \end{cases} \tag{2.12}$$

and

$$C_x(-k) = C_x^T(k) \tag{2.13}$$

The Fourier transform of $C_y(k)$, known as the spectral density function $S_y(w)$, is

a rational polynomial matrix and is given by

$$S_y(w) = H(e^{iw}I - \Phi)^{-1} \Sigma_x H^T + H(e^{iw}I - \Phi^T)^{-1} \Sigma_x H^T + R \qquad (2.14)$$

$$w \in [-\pi,\pi] .$$

Notice that a complete external description of the system from the knowledge of inputs and outputs is provided by $\{\bar{y}, C_y(k), k \geq 0\}$ or $\{\bar{y}, S_y(w), w \in [-\pi,\pi]\}$. Now consider parametric time series models that are based on the above external description.

Stationary Time Series Models

If in (2.3), the transition matrix Φ has all eigenvalues inside the unit circle and u(t) is stationary, the processes x(t) and y(t) are asymptotically stationary processes. Furthermore, by a change of basis of the form $x'(t) = Tx(t)$ where T is a nonsingular n x n matrix, the system equations can be written as:

$$x'(t+1) = \Phi' x'(t) + T \Gamma w(t) \qquad (2.15)$$

$$y(t) = H'x'(t) + v(t) \qquad (2.16)$$

where

$$\Phi' = T \Phi T^{-1}$$

$$H' = H T^{-1}$$

and u(t) = 0 for simplicity.

The form of T can be chosen in such a way that the matrices $T \Phi T^{-1}$ and HT^{-1} have the canonical form. For single-output systems, by choosing T as the observability matrix, i.e.,

$$T = \begin{bmatrix} H \\ H\Phi^{n-1} \\ \vdots \\ H\Phi^{n-1} \end{bmatrix} \qquad (2.17)$$

one gets

$$H' = [1,0,0,\ldots, 0] \qquad (2.18)$$

$$\Phi' = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & & \ldots & 1 \\ -\alpha_1 & -\alpha_2 & & \ldots & -\alpha_n \end{bmatrix} \qquad (2.19)$$

where $(\alpha_1,\ldots,\alpha_n)$ are the coefficients of the characteristic polynomial of Φ. For single-input, single-output systems, the transition matrix of the transformed system (which is the same as that of the original system) is a ratio of two polynomials corresponding to an ARMA model of the following form:

$$y(t) + \sum_{i=1}^{n-1} \alpha_i\, y(t-i) = \nu(t) + \sum_{i=1}^{n-1} \beta_i\, \nu(t-i) \qquad (2.20)$$

where $\nu(t)$ is the innovation or the one-step-ahead prediction error $(y(t) - \hat{y}(t|t-1))$ where $\hat{y}(t|t-1)$ is the prediction of $y(t)$ utilizing measurements up to time t-1. Furthermore, $\{\nu(t)\}$ is linearly related to $\{w(t)\}$ and $\{v(t)\}$ processes. Since, given a realization of the process $y(t)$, only the parameters $(\alpha_i,\beta_i,\ i=1,\ldots,n)$ and $\mathrm{var}(\nu)$ are estimable, the state vector model is identifiable if and only if the parameters in $\{\Phi,\Gamma,H,Q,R\}$ can be obtained uniquely from the ARMA parameters. The state vector canonical models are identifiable since they satisfy this requirement (Popov 1972).

The AR and moving average (MA) models are special forms of the ARMA model with $\beta_i = 0$ or $\alpha_i = 0$, respectively. Furthermore, an ARMA model may be written as either an infinite-order AR or an infinite-order MA model.

Canonical models for multi-input, multi-output (MIMO) systems are discussed by Mayne (1972), Mehra (1973), and Akaike (1974), and multiple ARMA models are discussed by Hannan (1970) and by Kashyap and Rao (1976). The single-input, single-output ARMA model is identifiable, but the multiple ARMA model requires restrictions on parameters to be identified.

Nonstationary Time Series Models

In many applications the class of stationary time series is too restrictive to be useful. For example, the series of monthly river flows have means and variances with seasonal dependence that is an essential part of the Thomas-Fiering (1962) model, and many population and sales series have trend components. Models of the following two classes are generally used to represent such time series.

Covariance Stationary Models In covariance stationary models, (Parzen 1976, Kashyap and Rao (1976), the mean is time varying, but the zero-lag covariance function is constant and the higher lag covariance function depends only on the lag variable. The time-varying mean is known as the trend and is generally represented as a polynomial-sinusoidal function of time. The subtraction of the mean, known as detrending, leads to a stationary time series.

A generalization of the covariance stationary models is achieved by allowing the covariance function to be the product of two functions, one dependent on time and the other not dependent on time. The Thomas-Fiering model uses a periodically varying covariance structure, while Kashyap and Rao (1976) use parametrically parsimonious models of the same type.

Autoregressive Integrated Moving Average Models Models obtained through repeated integrations of stationary ARMA models are known as autoregressive integrated moving averages (ARIMA) models (Box and Jenkins 1970). If the original ARMA series has a nonzero mean, then repeated integrations result in a polynomial time trend. Furthermore, irrespective of the mean, the variance of an ARIMA model would grow with time. In many cases, such behavior may be observed by plotting the data, but a better method is to examine statistical properties of the differenced time series.

Box and Jenkins (1970) also consider seasonal ARIMA series, which can be obtained

from stationary ARMA series or nonstationary ARIMA series by periodic summing of terms. For example, a series with an annual cycle may be produced by shifting a monthly series by 12 lags and adding it to the original series.

In practical time series analysis and system identification, the order of integration, the period, and the multiplicity of the seasonal component are unknown. It is often possible to represent the series by models from different classes. The final choice of a model is based on the nature of the situation being modeled and the purpose of identification.

Tests of Stationarity Practical time series methods require transformation of nonstationary time series into stationary time series because correlation and spectral methods have been developed mainly for stationary time series. In accordance with the discussion above, one may either use detrending or differencing, depending upon the assumptions regarding the form of the nonstationarity. In either case, a test of stationarity is required to make sure that the transformed series is stationary. According to Parzen (1976), the purpose of time series analysis is to transform a given time series into white noise. Since an AR, an ARMA, or a state vector model will convert most time series (in particular, linear time series with rational spectra) into white noise, Parzen (1976) claims that the approaches of detrending and differencing are asymptotically equivalent. However, for finite sample sizes, Kashyap and Rao (1976) have demonstrated that one of the approaches may be superior.

Several tests of stationarity have been proposed in the literature, but perhaps the most general is the T*-test, discussed by Zaremba (1967, 1972) and Pleszczynska (1971). The T*-statistic, defined below, tests the null hypothesis H_0: $\{Z_t\} = \{X_t\}$ against the alternative hypothesis H_1: $\{Z_t\} = \{X_t + y_t\}$, $(y_t \neq 0)$ where $\{X_t\}$, $t = 1, 2,\ldots,N$ is a linear purely stochastic process and y_t is a deterministic real function defined on $t = 0, \pm1, \pm2,\ldots$ satisfying the following conditions:

$$\lim_{N\to\infty} \frac{1}{NL_n^4} \sum_{k=1-N}^{N-1} (1 - \frac{|k|}{N}) Q_{k,N}^2 = Q \tag{2.21}$$

where

$$L_n = \max_{1\leq t\leq N} |y_t| \tag{2.22}$$

$$Q_{k,N} = \frac{1}{N-|k|} \sum_{t=1}^{N-|k|} y_t y_{t+|k|} \tag{2.23}$$

and Q is a positive constant.

Zaremba (1967) and Pleszczynska (1971) have shown that all polynomial and periodic functions of time satisfy condition (2.21). The T*-statistic is defined as follows:

For each N, choose two integers μ and ν such that the product $\mu\nu$ approximates N as closely as possible and ν is the smallest integer satisfying $N^{2/5} \leq \nu \leq N^{1/2}$.

Define

$$C^*_{s,N} = \frac{1}{N-|s|} \sum_{t=1}^{N-|s|} z_t z_{t+|s|} \qquad (|s| = 0,1\ldots,N-1) \tag{2.24}$$

$$S^*_N = \frac{1}{2} \sum_{s=1-N}^{N-1} \left(1 - \frac{|s|}{N}\right)^2 C^{*2}_{sN} \tag{2.25}$$

$$C^*_{0,\nu,p} = \frac{1}{2} \sum_{t=1}^{\nu} Z^2_{t+p\nu} \qquad (p = 0,1\ldots,\mu-1) \tag{2.26}$$

$$U^*_{\mu,\nu} = \frac{\nu}{\mu-1} \sum_{p=0}^{\mu-1} (C^*_{0,\nu,p} - C^*_{0,N})^2 \tag{2.27}$$

$$T^*_{\mu,\nu}(k) = \left(\frac{\mu}{2}\right)^{1/2} \left(1 - \frac{2S^*_N + kC^{*2}_{0,N}}{U^*_{\mu,\nu}}\right) \tag{2.28}$$

where $-2 \leq k \leq R_4 R_2^{-2}$, R_k being the kth cumulant of the innovation process of $\{X_t\}$. For a normal process, $R_4 R_2^{-2} = 0$. Since the exact distribution of $\{X_t\}$ is unknown, $k = -2$ is recommended. Zaremba (1967) and Pleszczynska (1971) show that under H_1

$$\operatorname*{plim}_{\substack{N \to \infty \\ \mu,\nu \to \infty}} T^*(k) = -\infty \,. \tag{2.29}$$

On the other hand, under H_0 (or $y_t \equiv 0$), $T(R_4 R_2^{-2})$ tends to normal distribution with zero mean and unit variance and

$$\operatorname{plim} T_{\mu,\nu}(k) = \begin{cases} +\infty & \text{for } k < R_4 R_2^{-2} \\ -\infty & \text{for } k > R_4 R_2^{-2} \end{cases} \tag{2.30}$$

Thus, for $k = -2$, a test in which the critical set is defined by $T^*_{\mu,\nu}(k) < \xi_\alpha$ has a level of significance asymptotically equal to zero. Furthermore, this test can be shown to have an asymptotic power (i.e., $\text{Prob}(T^*_{\mu,\nu}(k) < \xi_\alpha | H_1)$ of 1) if $\{X_t\}$ has finite moments up to order 8. The main advantage of the T*-test, besides its attractive type I and II errors, is that very little is assumed a priori regarding the process $\{X_t\}$. Pleszczynska (1971) also discusses a technique for polynomial trend removal using the T*-test that has been extended to more general stationarity-inducing transformations by Lissagor and Fihman (1976). Simpler but less powerful tests for detecting trends and periodicities have been given by Fisher (see Kashyap and Rao 1976), Bartlett (1966), and Jenkins and Watts (1968).

STATE PREDICTION AND PARAMETER ESTIMATION

The methods that have been used for state estimation and prediction for the models described in the preceding section are discussed below since these methods form the basis for parameter estimation and for combined state and parameter estimation. The basic assumption made here is that the model of the system is known in the form of (2.1) and (2.2) except for a vector of constant unknown parameters Θ. The results stated here are generally exact for linear models but are only approximate for the nonlinear model of (2.1) and (2.2). We start with the simplest linear case in (2.3) and (2.4) with no unknown parameters. The results for this case were first derived by Kalman (1960) in the form of a recursive filter, now commonly known as the Kalman filter.

Kalman Filter

Let $\hat{x}(t|\tau)$ denote the best linear unbiased estimator of $x(t)$ based on $\{y(1),\ldots y(\tau)\}$, the state model (2.3) and (2.4), and the a priori statistics of the initial state $x(0)$. Let $P(t|\tau)$ be the covariance matrix of the estimation error $(x(t) - \hat{x}(t|\tau))$. The Kalman filter computes $(x(t|\tau),P(t|\tau))$ for $t\geq\tau$ using the following equations recursively:

Update equations:

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)\nu(t) \tag{2.31}$$

$$\nu(t) = y(t) - H\hat{x}(t|t-1) \tag{2.32}$$

$$K(t) = P(t|t-1)H^T\Sigma_\nu^{-1}(t) \tag{2.33}$$

$$\Sigma_\nu(t) = HP(t|t-1)H^T+R \tag{2.34}$$

$$P(t|t) = (I-K(t)H)\ P(t|t-1) \tag{2.35}$$

One-step-ahead prediction equations:

$$\hat{x}(t|t-1) = \Phi\hat{x}(t-1|t-1) + Gu(t) \tag{2.36}$$

$$P(t|t-1) = \Phi P(t-1|t-1)\Phi^T + \Gamma Q\Gamma^T \tag{2.37}$$

Initial conditions:

$$\hat{x}(0|0) = x_0 \tag{2.38}$$

$$P(0|0) = P_0 \tag{2.39}$$

k-step-ahead prediction:

$$\hat{x}(t+k|t) = \Phi\hat{x}(t+k-1|t) + Gu(t+k) \tag{2.40}$$

$$P(t+k|t) = \Phi P(t+k-1|t)\Phi^T + \Gamma Q\Gamma^T \tag{2.41}$$

$$k = 2,3,4,\ldots$$

Not all the theoretical properties of the Kalman filter will be discussed since the information is readily available in texts such as those of Bryson and Ho (1969), Meditch (1969), Jazwinski (1970), Gelb (1974), and Schweppe (1974). Instead, we focus on some of the practical aspects such as unknown noise covariances, numerical problems necessitating square root formulations, dimensionality problems, robustness, and sensitivity properties.

Adaptive Kalman Filters

In practical applications of Kalman filters, the noise covariance matrices Q and R are generally not known. A systematic study of adaptive Kalman filters estimating Q and R recursively, in real-time, was started by Mehra (1968) by using the crucial fact that the innovation sequence $\nu(t)$ Eq. (2.32) contains all the information necessary for the estimation of the steady-state Kalman gain K and the measurement noise covariance matrix R. Mehra (1968) also described a whiteness test based on the innovation sequence to test the optimality of a Kalman filter.

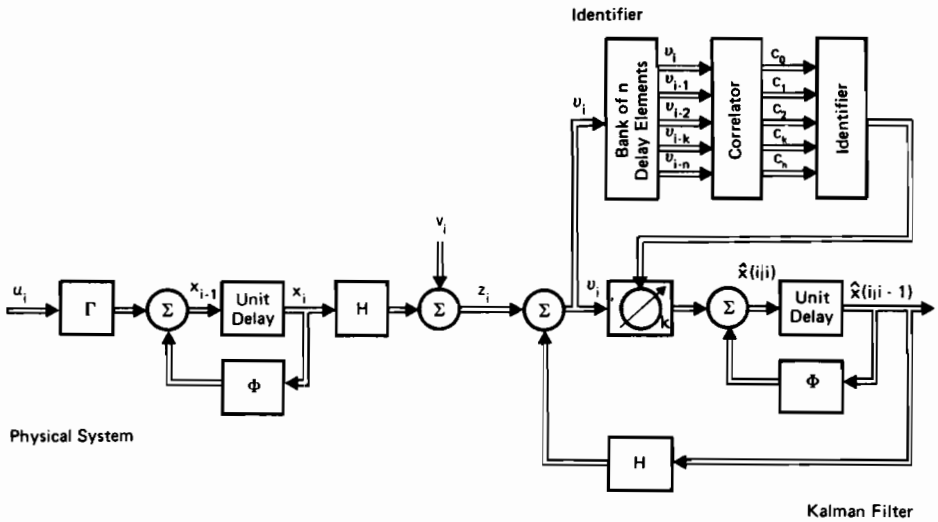


Fig. 2.3 An adaptive Kalman filter (from Mehra 1970).

A block diagram of the adaptive Kalman filter derived by Mehra (1970) is reproduced here as Fig. 2.3. The following equations must be solved:

Consider a suboptimal Kalman filter with steady-state gain K_0 and the innovation sequence $\{\nu(t)\}$. Let $\hat{C}_\nu(k)$ denote the k-lag sample correlation function after N time steps

$$\hat{C}_\nu(k) = \frac{1}{N} \sum_{t=k}^{N} \nu(t)\nu(t-k) \ . \tag{2.42}$$

The actual covariance matrix of the suboptimal filter is denoted by M_1. Notice that it is different from the covariance matrix computed by (2.35) and (2.37) for the suboptimal Kalman filter. Mehra (1968) shows that M_1H^T may be calculated uniquely from the innovation covariance sequence $(\hat{C}_\nu(1),\ldots,\hat{C}_\nu(n))$ where n is the order of the system. Then R is estimated using the equation

$$\hat{R} = \hat{C}_\nu(0) - HM_1H^T \tag{2.43}$$

The complete Q matrix is not identifiable, and only those elements of Q that are related to M_1H^T by a one-to-one transformation can be estimated. Alternatively, the optimal gain K may be estimated uniquely as follows:

$$\hat{K} = (\hat{M}_1H^T + \delta MH^T)\ (\hat{C}_0 + H\delta MH^T)^{-1} \tag{2.44}$$

where $\delta M = (M-\hat{M}_1)$ is the negative semidefinitive solution to the quadratic matrix equation

$$\hat{M} = \Phi[\delta M - (\hat{M}_1 H^T + \cdot\delta M H^T)(\hat{C}_0 + H\delta M H^T)^{-1}(H\hat{M}_1 + H\delta M) + K_0 H\hat{M}_1 + \hat{M}_1 H^T K_0^T - K_0 \hat{C}_0 K_0^T]\Phi^T \quad (2.45)$$

The estimates $\hat{K}$ and $\hat{R}$ are consistent and unbiased, but in general, they are not efficient. The efficiency, however, may be improved by repeated iterations through the data or by using the maximum likelihood method discussed in the introduction to Part One and by Mehra and Krishnaprasad (1974). It should be pointed out that simpler algorithms for estimating Q and R proposed by Sage and Husa (1969) and Jazwinski and Bailie (1967) are inconsistent and should be used with caution.

A survey of other techniques that can be used for adaptive filtering is contained in Mehra (1972). They may be classified as follows:

- Bayesian techniques, based on the computation of the posterior density function $p(x(t),\Theta|y(1),\ldots,y(t))$ (Smith 1967, Magill 1965)
- On-line maximum likelihood techniques using one Gauss-Newton iteration (Mehra, 1972)
- Output correlation techniques (Mehra, 1971a)
- Covariance matching techniques (Sage and Husa 1969, Jazwinski and Bailie 1967)

Kalman Filter Sensitivity and Robustness

Figs. 2.4a and b show the sensitivity of root-mean-square (RMS) estimation error

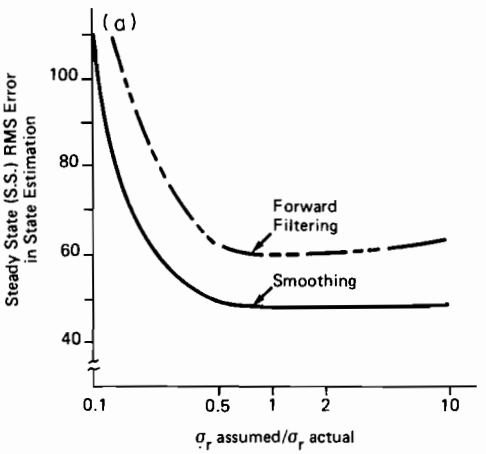


Fig. 2.4a Effect of errors in measurement noise variance on steady-state RMS error in state estimation. σ_r = measurement noise standard deviation. (from Mehra 1968).

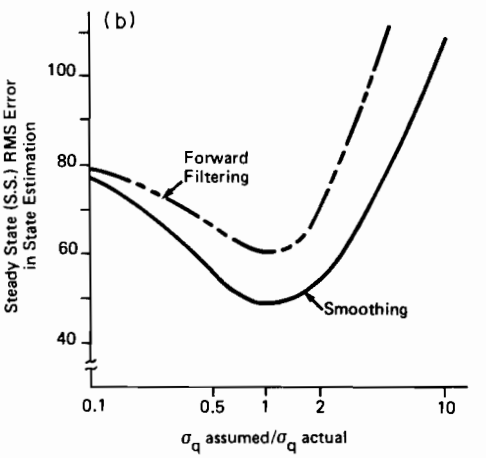


Fig. 2.4b Effect of errors in process noise variance on steady-state RMS error in state estimation. σ_q = process noise standard deviation. (from Mehra 1968).

to errors in the assumed values of measurement noise (R) and process noise (Q) variance. It is seen that in filter design, it is better to assume that R is larger than its expected value since there is extreme sensitivity to R on the lower side. Computing Kalman filter sensitivity (both small-scale and large-scale) by solving linear matrix equations (see, e.g., Gelb 1974) is a relatively straightforward procedure, and the information gained is very useful in practical filter design and in deriving reduced-order filters. A particularly serious problem is that of divergence, which can occur if some of the process noise terms are neglected in filter design (Fitzgerald 1967).

VandeLinde et al. (1972) and Masreliez and Martin (1977) have derived robust filters for non-gaussian statistics, including some simple techniques for reducing the Kalman gain K for unusually large innovations ν. Schweppe (1974) discusses a number of practical aspects of using Kalman filters for power systems, but the problems are common to other large systems and measurement devices. Two important considerations in such cases are computational efficiency and reinjection of bad data. The former is achieved by simplified covariance calculations and the latter is achieved by statistical tests on the innovations.

Numerical Problems

A straightforward use of the Kalman filtering equations (2.31) - (2.41) can lead to negative-definite covariance matrices and filter divergence, particularly when the spread of eigenvalues of the error covariance matrix P is large. Square-root filters have been used successfully in these cases (Potter 1963, Schmidt 1970, Dyer and McReynolds 1969, Kaminski et al. 1971). Bierman (1976) gives a complete description of these techniques and their numerical properties. It is also possible to compute the Kalman gain directly without first computing the filter covariance by using the so-called Chandrashekhar-type filters for time-invariant systems (see Kailath 1973, Linquist 1974, and Morf and Kailath 1975 for further details).

Nonlinear Filters

Conceptual hydrological models contain nonlinearities that would necessitate the use of advanced nonlinear filters; extended Kalman filters would most probably be inadequate. On the other hand, computational considerations would forbid the use of optimal or nearly optimal nonlinear filters. A useful compromise could be achieved by the use of the iterative sequential filters discussed in Jazwinski (1970) and Mehra (1971). In these filters, the measurement update step is repeated several times by linearizing around the latest best estimate until convergence is achieved. Other more advanced nonlinear filters (second order filters, iterative smoothing filters, and so on) are discussed in Wishner et al. (1971).

A serious problem with nonlinear filters is estimation of their performance, since the filter covariance matrix may be very different from the actual error covariance matrix. The only satisfactory way to judge the performance of nonlinear filters is through Monte Carlo simulations, but these can be quite expensive because of the computation time they require.

Parameter Estimation

The problem of parameter estimation in state vector models has received considerable attention in the last decade. Next to least squares, the most widely used method is the maximum likelihood (ML) method (Aström and Bohlin 1965, Kashyap 1970, Mehra 1969, Gupta and Mehra 1974). The main advantage of the ML method is that it can be shown to be asymptotically unbiased, consistent, and efficient under certain mild regularity conditions. In fact, these properties make the ML method more attractive than least squares, even though the computation is more expensive.

Furthermore, ML can be used as a starting point for devising simpler techniques, often generically called approximate ML, which trade statistical efficiency for computational simplicity and recursiveness.

For on-line parameter estimation the most versatile technique is the instrumental variable (IV) technique (Young 1970). The basic concept of the IV technique is explained easily in terms of the following linear model:

$$Y = X\Theta + V_\tau \tag{2.46}$$

where Y is the N-dimensional vector of the dependent variable, X is an N x m matrix of independent variables, Θ is an m-dimensional vector of unknown parameters, and V is an N-dimensional vector of random errors. Here N denotes sample size and V is normally distributed with zero mean and covariance Σ_V. The ordinary least square (OLS) estimate of Θ is

$$\hat{\Theta}_{OLS} = (X^T X)^{-1} X^T Y \ . \tag{2.47}$$

The generalized least squares (GLS) estimate of Θ is

$$\hat{\Theta}_{GLS} = (X^T \Sigma_V^{-1} X)^{-1} X^T \Sigma_V^{-1} Y \ . \tag{2.48}$$

Both OLS and GLS estimates of Θ are unbiased and consistent as long as X and V are uncorrelated (GLS has the additional advantage that it is also efficient). However, when the state vector model of (2.3)-(2.4) or the ARMA model of (2.20) is expressed in the form of (2.46) (Mehra 1973), X and V are found to be correlated. To obtain unbiased and consistent estimates for this case, the IV estimate is computed as follows:

$$\hat{\Theta}_{IV} = (Z^T X)^{-1} Z^T Y \tag{2.49}$$

where Z is an N x m matrix with the property that

$$Z^T V = 0$$

and $(Z^T X)$ is nonsingular. In other words, the columns of Z, known as instruments, are uncorrelated with the error but are correlated with the independent variables in such a way that the correlation matrix is invertible. The optimal choice of the instruments is important for efficiency of the estimates and is discussed by Wong and Polak (1967) and Finigan (1975). Finigan discusses in detail the on-line estimates of parameters in linear dynamic models using the IV technique. It is interesting that the optimal IV estimator involves a Kalman filter to obtain one-step-ahead predictions of the independent variables, which are found to be uncorrelated with the noise terms in the model. This can be demonstrated by multiplying both sides of (2.3) by H and using (2.4) to obtain

$$y(t+1) = H\Phi x(t) + (HGu(t) + H\Gamma w(t) + v(t)) \ . \tag{2.50}$$

Now replace x(t) by ($\hat{x}(t|t-1)$ + e(t) where $\hat{x}(t|t-1)$ is the one-step-ahead predicted state estimate obtained from the Kalman filter equations (2.31) through (2.39) using measurements {y(1),...,y(t-1)} and inputs {u(1),...,u(t-1)}. Since $\hat{x}(t|t-1)$ is based on measurements up to time t-1 and since, by the orthogonality principle, e(t) and $\hat{x}(t|t-1)$ are uncorrelated, we can choose $\hat{x}(t|t-1)$ as the instruments to estimate unknown parameters in matrices (HΦ,HG). This argument assumes that the Kalman filter and the parameter estimator are used iteratively such that each requires the other's output. Finigan (1975) examines the stability question and suggests the use of a third filter to ensure both stability and optimality of the IV estimator.

MODEL STRUCTURE DETERMINATION

The problem of model structure determination occupies a central role in system identification. In its most general form, it may be thought of as a hierarchical problem (Fig. 2.5).

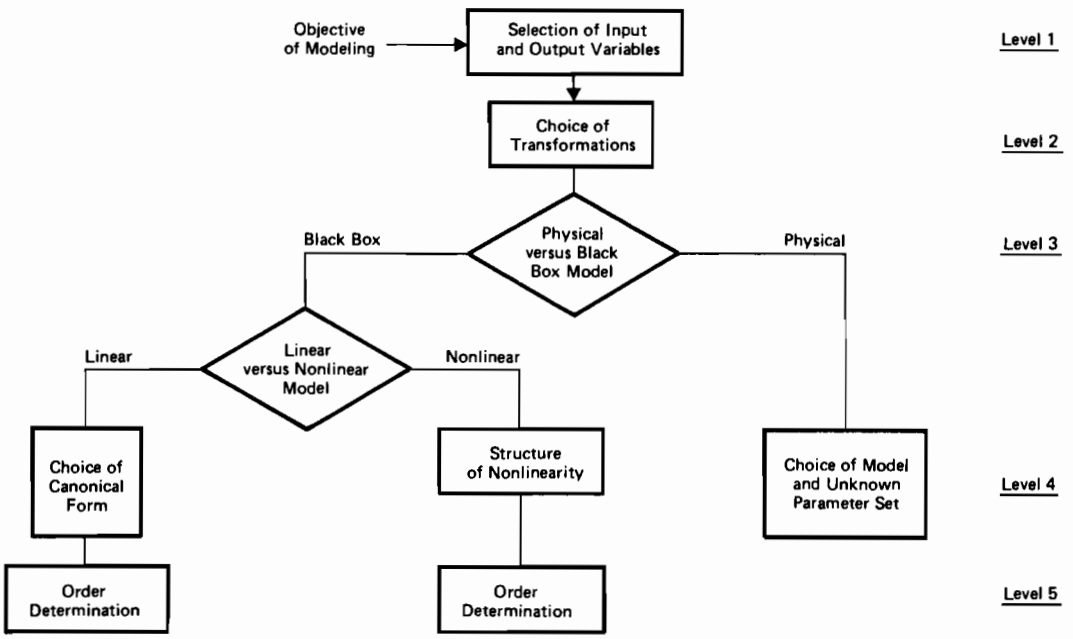


Fig. 2.5 Hierarchical levels in model structure determination.

The selection at the first level is done subjectively and is based on the objectives of modeling, on the understanding of the process to be modeled, and on the scope of the modeling effort. The purpose of level 2 is to transform the measured variables into a form suitable for different classes of models. For example, if an ARMA or a linear state vector model is to be used, the transformations are chosen to induce linearity and normality. The selection at level 3 between a mechanistic and black box model is based on such factors as the objective of modeling, the degree of belief in the mechanistic model, and the relative complexities of the mechanistic and black box models. In many cases, a combination of the models, known as a gray box or conceptual model, may be used, or both types of model may be fitted and their performances compared before the final selection. Since linear models are simpler to fit, one generally starts with a linear model and adds nonlinearities if it is found to be inadequate. Mathematical and statistical techniques are

available for the analysis in levels 4 and 5 and will be discussed later in this section and in the next section. However, it may be observed that experienced judgment is required at each level of model structure determination, and, for this reason, the process cannot be fully automated. Certain guiding principles of model building have evolved from experience and are summarized below.

Parsimony The model should contain as few parameters as possible for adequate representation; for example, if a particular time series requires a tenth-order AR model to produce as good a fit to the data as a second-order ARMA model, then the latter is preferable since it involves five parameters instead of eleven. The reason for parsimony is that the sample size N from which the parameters are to be estimated is finite, and the fewer the parameters, the lower the variance of the estimates. Akaike (1974) has demonstrated this elegantly in terms of the final prediction error (FPE) criterion for an AR model with m parameters and N samples (Fig. 2.6).

$$\text{mean square FPE} = \frac{N + m}{N - m} \times \text{mean square fit error} \tag{2.51}$$

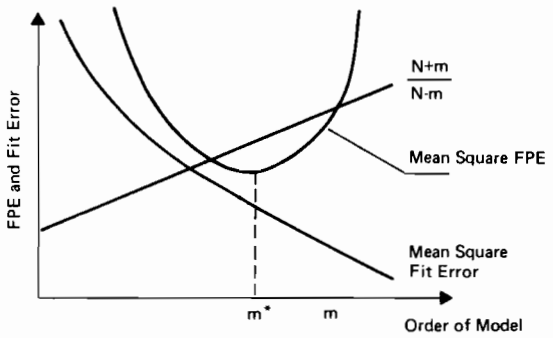


Fig. 2.6 Final prediction error and fit error versus order of the AR model. M* is the optimal order of the models.

Identifiability The unknown parameters in the model must be identifiable from the given measurements. For linear state vector models, this necessitates the use of canonical forms such as the companion form of (2.18) and (2.19) or the more general forms derived by Popov (1972) and Akaike (1974) for multi-input, multi-output systems. The problem is also of great importance in the simultaneous equation models of econometrics, which have the form

$$Ay(t+1) = \Phi y(t) + Gu(t) + \Gamma w(t) \ . \tag{2.52}$$

In (2.52), A is nonsingular and only the parameters of $A^{-1}\Phi$, $A^{-1}G$, and $A^{-1}\Gamma$ are identifiable. Fisher (1966) and Goldberger (1968) discuss various conditions for the identifiability of parameters in A, Φ, G, and Γ.

Representation The model should reflect reality as much as possible, that is, the state variables should represent real conditions. Where it is necessary or desirable to use black box models, the identified models should be subjected to rigorous tests for reasonableness of behavior.

Application The objective of the modeling should be stated clearly e.g. forecasting, control, understanding, and should influence the choice of the model. It should be

kept in mind that a model suitable for short-term forecasting will generally not be suitable for long-term forecasting and vice versa. Similarly, for control of feedback situations, a black box model may suffice even though it may not provide understanding of the internal workings of the system.

A specific procedure for model structure determination in MIMO linear systems was devised by Akaike (1976a). A flow diagram of the procedure is shown in Fig. 2.7. For simplicity, we consider the case u(t) = 0 (i.e., no deterministic input).

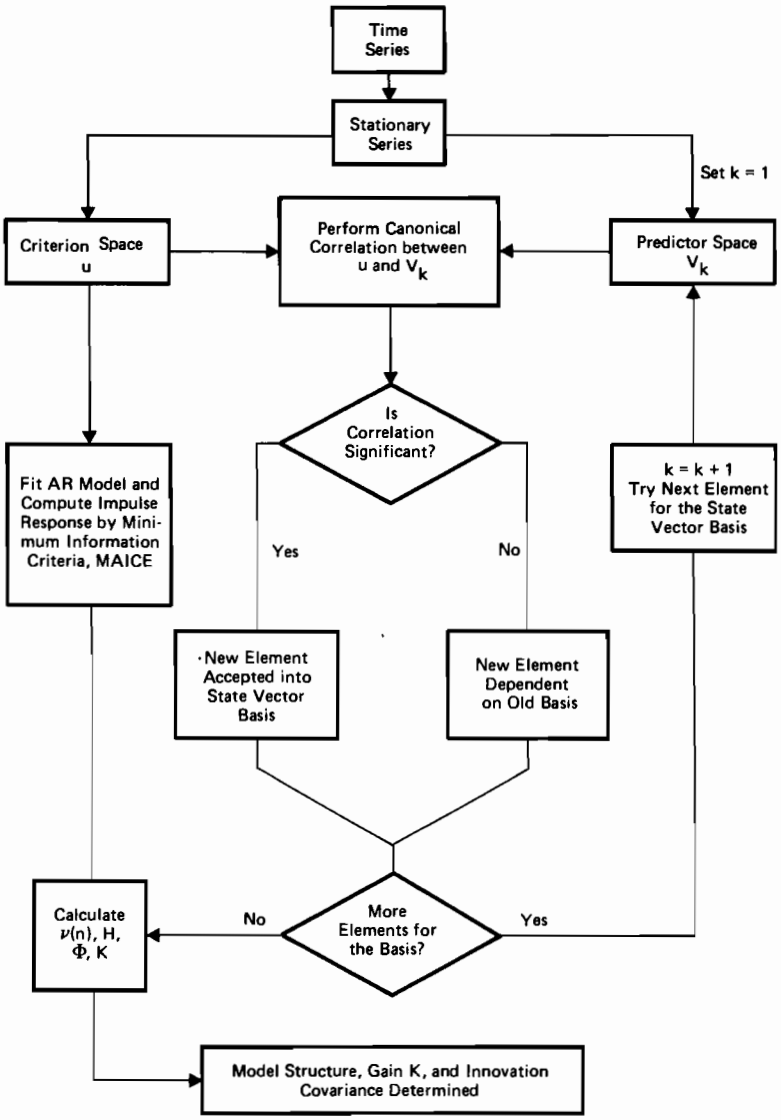


Fig. 2.7 Akaike's procedure for model structure determination in MIMO linear systems.

The basic concept behind the method is that the state vector of a system, at current time t, may be defined as a basis vector for the space spanned by the current output

and future predictions:

$$\{y(t),\ \hat{y}(t+1|t),\ldots\ \hat{y}(t+k|t)\ldots,\}\ .$$

This space is referred to as the predictor space of a system at time t. It can be shown that this space is finite dimensional for a finite dimensional system and that the dimension of the system can be determined by examining the canonical correlations between the sets of variables U and V_k for k = 1,2,3,... where $U = \{y^T(t),y^T(t-1),\ldots y^T(t-k)\}$, K is sufficiently large, and $V_k = \{y^T(t),\hat{y}^T(t+1|t),\ldots\hat{y}^T(t+k|t)\}$

By increasing k by 1 in each iteration and by considering elements of y in a given order, the state vector x(t) of the system is determined as a subset of V_k as shown in Fig. 2.7. Once the state vector is identified, the matrix Φ is determined simultaneously with x(t) from the canonical variables. The purpose of fitting the AR model and computing the impulse response function is to estimate the Kalman gain matrix K and the innovation covariance matrix Σ_v. Notice that the procedure identifies the stationary Kalman filter model directly and does not involve solving Ricatti equations. The computation of the AR model is based on the Levinson-Whittle-Wiggins-Robinson-Burg (LWWRB) algorithm, and the order of the model is selected using Akaike's minimum information criterion (MAICE) (Akaike 1976a). It has been shown by Shibata (1976) that Akaike's Information Criterion (AIC) tends to over-estimate the system order. This may also be true of the F-test, which has been shown to be asymptotically equivalent to AIC by Soderstrom (1976). Akaike (1976b) is investigating a new criterion, called BIC, which is a consistent estimator of the system order. (Further discussion of Akaike's technique is given in Chapters 3 and 4.)

MODEL VALIDATION AND COMPARISON OF MODELS

Model validation involves testing an identified model and comparing it with other candidate models. To validate a model, one can

- Simulate the identified model and generate synthetic time series; compute various statistical characteristics of these time series such as correlograms, power spectra, threshold crossing frequencies, rescaled range (or Hurst range) characteristics, or histograms; and compare these characteristics with those of the process being modeled. (Fiering 1967, Kashyap and Rao 1976, Lawrence and Kattegoda 1977).

- Test the null hypothesis that the identified model cannot be rejected as the true model against the alternate hypothesis that some other model or class of models cannot be rejected as the true model of the system (Wald 1943, Whittle 1952). The difficulty with this approach is that there is generally no simple way of defining the alternate hypothesis.

- Use an information-theoretic or entropy-related criterion to select the model that optimizes the selected criterion (see Akaike 1972, Rissanen 1976).

The Bayesian decision theory approach This approach seems to be particularly suitable for model validation and comparison (Kashyap 1976, Peterka 1976b). Assume that K different models denoted by M_1, $M_2,\ldots,M_K$ are entertained *a priori* as candidates for the final model. The *a priori* probabilities $p(M_i)$, i = 1,...,K are first specified. Let each model M_i be of the form

$$f_i(y(t)) = g_i^T(t,y(t-1),\ldots,y(t-m_i))\ \theta_i + w_i(t)$$

$$t = 1,\ldots,N, \quad \theta_i \in R^{n_i} \tag{2.53}$$

where $w_i(t)$ is a zero mean, white noise sequence with variance σ_i^2. It is assumed that the parameters $(\theta_i\sigma_i, m_i, n_i)$ are unknown, but prior distributions (informative or noninformative) for (θ_i, σ_i) can be specified.

Kashyap (1976) derives the posterior distributions $p(M_i|y(1),\ldots,y(N))$ for $i = 1,\ldots,K$ based on which, with the help of a suitable risk function, the final model is selected. For an AR model, the Bayes risk function for minimizing the probability of error is given by

$$R_i' = N\ \ell n\ \hat{\sigma}_i^2 + n_i\ (\ell n\ N\ \sigma_p^2/\hat{\sigma}_i^2) \tag{2.54}$$

where

$\hat{\sigma}_i^2$ = sample variance of residuals for model M_i

σ_p^2 = variance of the process

Notice that in (2.54), the relative importance of the second term involving the number of parameters n_i varies with sample size N. In the AIC criterion given by Akaike (1974), the factor multiplying n_i is fixed at 2, but the BIC criterion of Akaike (1976b) has a form similar to (2.54).

An interesting consequence of (2.54) and the associated probability of error (PE) computation is developed by Kashyap (1976) for the case of two AR models M_1 and M_2 with parameters n_1 and n_2 respectively, where M_1 is the alternate model and M_2 is the true model. Two cases arise

$$\text{(a)}\quad \text{If } n_1 > n_2, \quad PE = N^{-(n_1-n_2)/2} \tag{2.55}$$

$$\text{(b)}\quad \text{If } n_1 < n_2, \quad PE = K^{-N} \tag{2.56}$$

Notice that if the true model M_2 has fewer parameters than the alternate model M_1 such that M_2 can be obtained by setting some parameters in the alternate model to zero (case a), then the convergence rate of PE is much slower than in case b, where the true model M_2 is compared with a model with fewer parameters and PE goes to zero exponentially as $N \to \infty$.

The advantages of using the Bayesian decision theory approach are that:

- More than two models of widely differing natures can be compared.
- The decision rule is transitive, i.e., if M_i is preferred to M_j and M_j is preferred to M_k then M_i is preferred to M_k. Most hypothesis tests do not possess this property.
- The loss function is well defined since both type I and type II errors are

considered. The objective of identification, such as spectral estimation or prediction error, may be considered directly.

MISCELLANEOUS TOPICS IN TIME SERIES ANALYSIS

Time-Varying Parameter Models

In many applications (Bohlin 1976, Peterka 1976a, Mehra and Krishnaprasad 1974), a simple regression model is suitable if the coefficients are allowed to vary randomly over time. The following Gauss-Markov model for parameters $\Theta(t)$ has been found useful in these applications

$$\Theta(t+1) = F\,\Theta(t) + w(t) \tag{2.57}$$

$$y(t) = C(t)\,\Theta(t) + v(t) \tag{2.58}$$

It is possible to estimate parameters in F, Σ_w, and Σ_v, which are covariance matrices of $w(t)$ and $v(t)$, using the maximum likelihood technique. The on-line tracking of parameters is performed easily by using a Kalman filter for the above model.

Applications of the above technique to dryer control in a paper mill, EEG with changing spectra, machine failure forecasting, and power load prediction have been considered by Bohlin (1976). Peterka (1976a) and Mehra (1978) consider the problem of detecting sudden changes in the parameters of a system using the above model.

Nonlinear System Identification

There is a distinct lack of methods for identifying the structure of nonlinear systems. One method that has been reported to be successful on applications is Ivakhnenko's group method of data handling (GMDH). This method uses a network of quadratic polynomials in a layering arrangement based on perception. At each level, heuristic criteria are used to select models. Applications of this method are described by Ivakhnenko (1975), Duffy and Franklin (1975), Ikeda et al. (1976), and Barron (1975). An application of GMDH is given in Chapter 10.

Random Fields

Since environmental and hydrological processes are multidimensional in character, the rapidly developing theory of random fields is particularly relevant for these applications. The monograph by Bartlett (1976) covers the theory of random fields from a statistical and modeling viewpoint. Two classes of models, nearest neighbor models and Markov fields, are discussed with numerical examples. The literature on random fields in the control and communication areas is discussed in Willsky (1977). A class of models with many properties similar to Gauss-Markov models discussed in previous sections has been proposed by Attasi (1976).

Fractional Noise Models

The so-called Joseph and Noah effects observed in hydrological time series (Mandelbrot and Wallis 1968) have been modeled by using fractional noise models, which are discussed in a survey paper by Lawrence and Kattegoda (1977). However, the development of these models from finite sample data is quite involved and the results obtained by Kashyap and Rao (1976) question the usefulness of these

models for empirical time series analysis.

CONCLUSIONS

Time series analysis and system identification are arts. However, they possess a very rich set of tools that have been used successfully for numerous applications over the last fifty years. In recent years, further impetus has been provided by developments in control, communication, and system theory. In particular, the state vector models of control theory have emerged as a unifying link between different classes of models. An attempt has been made in this paper to present a comprehensive treatment of state vector models. Further research is required in the following areas:

- Model structure determination for nonlinear systems
- Validation for models involving both process and measurement noise
- Nonstationarities present in real systems
- Random fields.

REFERENCES

Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle. Int. Symp. Informat. Theory, USSR.

Akaike, H. (1974). Stochastic theory of minimal realizations. IEEE Trans. Autom. Control 19(6): 667-674.

Akaike, H. (1976a). Canonical correlation analysis of time series and the use of an information criterion, in R.K. Mehra and D.G. Laintiotis, Eds., System Identification, Advances and Case Studies. Academic Press, New York.

Akaike, H. (1976b). On the entropy maximization principle. Symposium on Applications of Statistics, Dayton, Ohio, June 14-18.

Aström, K.J. (1967). Computer control of a paper machine in an application of linear stochastic control theory. IBM J. Res. Dev. 11: 389-405.

Aström, K.J., and T. Bohlin (1965). Numerical identification of linear dynamic systems from normal operating records. Proc. 2nd IFAC Symp. on Self-Adaptive Systems, pp. 96-111.

Aström, K.J., T. Bohlin, and S. Wensmark (1967). Automatic construction of linear stochastic dynamic models for stationary industrial processes. Technical Report TR-18-150, IBM Nordic Laboratory.

Aström, K.J., and P. Dykhoff (1971). System identification in a survey. Automatica 7: 123-162.

Attasi, S. (1976). Modelling and recursive estimation for double indexed sequences, in R.K. Mehra and D.G. Laintiotis, Eds., System Identification, Advances and Case Studies. Academic Press, New York.

Barron, R.L. (1975). Learning networks improve computer-aided prediction and control. Comput. Des. August.

Bartlett, M.S. (1966). An Introduction to Stochastic Processes. Cambridge University Press, London and New York.

Bartlett, M.S. (1976). The Statistical Analysis of Spatial Pattern. Chapman and Hall, London.

Beckey, G.A. (1976). Identification of biological and medical systems. Fourth IFAC Symp. on Identification and System Parameter Estimation, Part I., Tbilisi, USSR, September.

Bierman, G.J. (1976). Factorization Methods for Discrete Sequential Estimation. Academic Press, New York.

Blackman, R.B., and J.W. Tukey (1959). The Measurement of Power Spectra. Dover, New York.

Bohlin, T. (1976). Four cases of identification of changing systems from discrete time series, in R.K. Mehra and K.G. Laintiotis, Eds., System Identification, Advances and Case Studies. Academic Press, New York.

Box, G.E.P., and G.M. Jenkins (1970). Time Series Analysis in Forecasting and Control. Holden-Day, San Francisco.

Box, G.E.P., et al. (1976). NIBER-CENSUS Conference on Seasonal Analysis of Economic Time Series, Washington, D.C., September.

Bryson, A.E., and Y.C. Ho (1969). Applied Optimal Control. Xerox Publishing Company, New York.

Duffey, J.J., and M.A. Franklin (1975). A learning identification algorithm and its application to an environmental system. IEEE Trans. Syst. Man. Cybern. 5(2), March.

Dyer, P., and S.R. McReynolds (1969). Extension of square root filtering to include process noise. J. Optimization Theory and Applications 3(6): 444-459.

Eykhoff, P. (1974). System Identification: Parameter and State Estimation. Wiley, New York.

Fiering, M.B. (1967). Streamflow Synthesis. Harvard University Press, Cambridge, Massachusetts.

Fiering, M.B. (1971). Synthetic stream flows. Water Resour. Monogr. 1, Am. Geophys. Union, Washington, D.C.

Finigan, B. (1975). Strongly consistent parameter estimation using the instrumental variable approach. Pd.D. Thesis, University of Toronto, Canada, October.

Fisher, F.M. (1966). The Identification Problems in Econometrics, McGraw-Hill, New York.

Fitzgerald, R.J. (1967). Error divergence in optimal filtering problems. Presented at the 2nd IFAC Symp. Automatic Control in Space, Vienna, Austria.

Galiana, F.D., E. Handschin, and A. Fiechta (1974). Identification of stochastic load models from physical data. IEEE Trans. Autom. Control, December.

Gelb, A., Ed. (1974). Applied Optimal Estimation. M.I.T. Press, Cambridge, Massachusetts.

Gersch, W. (1972). Causality or driving in electrophysiological signal analysis. Math. Biosci.

Goldberger, A.S. (1968). Topics in Regression Analysis. MacMillan, New York.

Gupta, N.K., and R.K. Mehra (1974). Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. IEEE Trans. Autom. Control, Special Issue on System Identification and Time Series Analysis, December.

Gustafson, D.E., A.S. Willsky, J.Y. Wang, M.C. Lancaster and J.H. Triabwasser (1977). A statistical approach to rhythm diagnosis of cardiograms. IEEE Proc., Special Issue on Biomedical Signal Processing, April.

Hannan, E.J. (1960). Time Series Analysis. Methuen, London.

Hannan, E.J. (1970). Multiple Time Series. Wiley, New York.

IEEE (1974). System Identification and Time Series Analysis. IEEE Trans. Autom. Control. December.

Ikeda, S., M. Ochaia, and Y. Sawaragi (1976). Sequential GMDH algorithm and its application to river flow prediction. IEEE Trans. Syst. Man. Cybern. 6(7), July.

Ivakhnenko, A.G. (1970). Heuristic self-organization in problems of engineering cybernetics. Automatica.

Ivakhnenko, A.G. (1971). Polynomial theory of complex systems. IEEE Trans. Syst. Man. Cybern. 1(4), October.

Ivakhnenko, A.G. (1975). Control of Dniepl hydroelectric power station water reservoirs with two optimality criteria based on self-organization. Sov. Autom. Control.

Ivakhnenko, A.G. and N.A. Ivakhnenko (1974). Long-term prediction of random processes by GMDH algorithm using unbiasedness criterion and balance-of-variables criterion. Sov. Autom. Control.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Jazwinski, A.H., and A.E. Bailie (1967). Adaptive filtering. Analytical Mechanics Associates, Inc., Rep. 67-6, March.

Jenkins, G.M., and D.G. Watts (1968). Spectral Analysis and Its Applications. Holden-Day, San Francisco.

Johnston, J. (1963). Econometric Methods. McGraw-Hill, New York.

Kailath, T. (1973). Some new algorithms for recursive estimation in constant linear systems. IEEE Trans. Inf. Theory. 19(6): 750-60, November.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Trans. ASME, J. Basic Eng. 82: 35-44.

Kalman, R.E. (1963). New methods in Wiener filtering theory, in J. Bogdanov and F. Kozin, Eds., Proc. Symp. on Engineering and Applied Random Functions. Wiley, New York.

Kalman, R.E., P.R. Falb, and M.A. Arbib (1969). Topics in Mathematical System Theory. McGraw-Hill, New York.

Kaminski, P.G., A.E. Bryson, and S.F. Schmidt (1971). Discrete square root filtering: A survey of current techniques. IEEE Trans. Autom. Control. 16(6): 727-36.

Kashyap, R.L. (1970). Maximum likelihood identification of stochastic linear systems. IEEE Trans. Autom. Control. 15: 25-34.

Kashyap, R.L. (1976). A Bayesian Comparison of Different Classes of Dynamical Models Using Empirical Data, Internal Report, Purdue University, Lafayette, Indiana.

Kashyap, R.L., and A.R. Rao (1976). Dynamic stochastic models from empirical data. Academic Press, New York.

Lawrence, A.J., and N.T. Kattegoda (1977). Stochastic modeling of riverflow time series. J.R. Stat. Soc. A140, Part I.

Levinson, N. (1949). The Wiener RMS criterion in filter design and prediction, in N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series. Wiley, New York.

Lindquist, A. (1974). A new algorithm for optimal filtering of discrete-time stationary processes. SIAM J. Control. 12(4): 736-746.

Lissagor, M., and L. Fihman (1976). Parametric bias in forecasting in alternative method to Box-Jenkins, in Automatiz Provanie Systemi Ipravlenia v Neftanoi i Gazovoi Promishlenosti (in Russian).

Magill, D.T. (1965). Optimal adaptive estimation of sampled stochastic processes. IEEE Trans. Autom. Control. 10: 434-439, October.

Management Science (1977). Special Issue on Forecasting.

Mandelbrot, B.B., and J.R. Wallis (1968). Noah, Joseph and Operational Hydrology. Water Resour. Res. 4(5): 900-18.

Mann, A.B., and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. Econometrica 11: 173-220.

Masreliez, C. J., and R.D. Martin (1977). Robust Bayesian estimation for the linear model and robustifying the Kalman filter. IEEE Trans. Autom. Control, June.

Mayne, D. (1972). A canonical model for identification of multi-variable linear systems. IEEE Trans. Autom. Control 17: 728-9.

Meditch, J.S. (1969). Stochastic Optimal Linear Estimation and Control. McGraw-Hill, New York.

Mehra, R.K. (1968). On optimal and suboptimal linear smoothing. Proc. Nat. Electron. Conf. XXIV, December.

Mehra, R.K. (1969). Identification of linear dynamic systems. IEEE Symposium on Adaptive Processes, November (also AIAA J. January 1971).

Mehra, R.K. (1970). Identification of variances and adaptive Kalman filtering. Joint Automatic Control Conf., Boulder, Colorado, August 1969 (also IEEE Trans. Autom. Control, 15(2), February 1970).

Mehra, R.K. (1971a). On-line identification of linear dynamic systems with applications to Kalman filtering. Joint Automatic Control Conf., Atlanta, Georgia, June 24-26, 1970 (also IEEE Trans. Autom. Control, February 1971).

Mehra, R.K. (1971b). A comparison of several nonlinear filters for reentry vehicle tracking. IEEE Symposium on Decision and Control, Austin, Texas, December 1970 (also IEEE Trans. Autom. Control, August 1971).

Mehra, R.K. (1972). Approaches to adaptive filtering. IEEE Trans. Autom. Control, October.

Mehra, R.K. (1973). Identification and control in econometric systems, similarities and differences. Second Workshop on Economic and Control Systems, Chicago, June (also Annals of Economic and Social Measurements, January 1974).

Mehra, R.K. (1976). Identification and estimation of the error-in-variables model (EVM) in structural form. RM, International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria, May 1975 (also Proceedings of the Symposium on Stochastic Systems, North-Holland, June 1976).

Mehra, R.K. (1978). Fault detection and compensation in dynamic systems. IFAC Congress, Helsinki.

Mehra, R.K., and P.S. Krishnaprasad (1974). A unified approach to the structural estimation of distributed lags and stochastic differential equations. Third NBER Conference on Stochastic Control, Washington, D.C., May.

Mehra, R.K., and D.G. Lainiotis, Eds. (1976). System Identification, Advances and Case Studies. Academic Press, New York.

Mehra, R.K., and J.S. Tyler (1973). Case studies in aircraft parameter identification. Third IFAC Conference on Identification and System Parameter Estimation, The Hague, Netherlands, June.

Mehra, R.K. and C.H. Wells (1971). Dynamic modeling and estimation of carbon in a basic oxygen furnace. Third International IFAC/IFIP Conference, Helsinki, June 2-5.

Morf, M., and T. Kailath (1975). Square root algorithms for least squares estimation. IEEE Trans. Autom. Control. 20(4): 487-97.

Olsson, G. (1976). Modeling and identification of a nuclear reactor, in R.K. Mehra and D.G. Laintiotis, Eds., System Identification, Advances and Case Studies. Academic Press, New York.

Pandya, R.N. (1972). A class of bootstrap estimates for identification of linear discrete time models. TR No. SE72-3, Carleton University, Ottawa.

Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. Ann. Math. Stat. 28: 329-48.

Parzen, E. (1976). Discussion on analysis and modeling of seasonal time series, in G.E.P. Box et al., NBER-CENSUS Conference on Seasonal Analysis of Economic Time Series. Washington, D.C., September.

Peterka, V. (1976a). Experience accumulation for decision making in multivariate time series. RM-76-70, International Institute for Applied Systems Analysis, Laxenburg, Austria, September.

Peterka, V. (1976b). Subjective probability approach to real-time identification. Fourth IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, USSR.

Phatak, A.V., R.K. Mehra, and C.N. Day (1975). Application of system identification to modeling the human controller under stress conditions. IEEE Trans. Autom. Control, 20(5).

Pleszczynska, E. (1971). Further application of the T*-test to time-series analysis. Zastosowania Matematyka, Applicationes Matematicae XIII(1).

Popov, V. (1972). Invariant description of linear time-invariant controllable systems. SIAM J. Control, 10: 252-64.

Potter, J.E. (1963). New statistical formulas. Space Guidance Analysis Memo 40, Instrumentation Laboratory, M.I.T., Cambridge, Massachusetts, April.

Quenouille, M.H. (1957). The Analysis of Multiple Time Series. Hafner, New York.

Rissanen, J. (1976). Minimax entropy estimation of models for vector processes, in R.K. Mehra and D.G. Lainiotis, Eds., System Identification, Advances and Case Studies. Academic Press, New York.

Sage, A.P. and G.W. Husa (1969). Adaptive filtering with unknown prior statistics. Proc. 1969 Joint Autom. Control Conf., 760-69.

Schmidt, S.F. (1970). Computational techniques in Kalman filtering, Theory and Application of Kalman Filtering. NATO Advisory Group for Aerospace Research and Development AGARDograph 139, Technical Editing and Reproductions Limited, Harford House, London, February.

Schweppe, F.C. (1970). Power system-static state estimation, Parts I, II, III. IEEE Trans. Power Appar. Syst. 89: 120-135.

Schweppe, F.C. (1974). Uncertain Dynamic Systems. Prentice-Hall, Englewood-Cliffs.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63(1): 117-26.

Smith, G.L. (1967). Sequential estimation of observation error variance in a trajectory estimation problem. AIAA J. 5: 1964-70, November.

Soderstrom, T. (1973). An on-line algorithm for approximate maximum likelihood identification of linear dynamic systems. Rep. 7308, Division of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Soderstrom, T. (1976). Testing of model structures in system identification. UPTEC 76 28 R, Institute of Technology, Uppsala University, Uppsala, Sweden, May.

Swanson, G.D. (1972). Dynamic Forcing in the Study of the Human Respiratory System. Ph.D. Thesis, Stanford University, August.

Szöllösi-Nagy, A. (1976). Introductory remarks on the state space modeling of water resource systems. RM 76-073, International Institute for Applied Systems Analysis, Laxenburg, Austria, September.

Theil, H. (1971). Principles of Econometrics. John Wiley and Sons, New York.

Thomas, H.A. and M.B. Fiering (1962). Mathematical synthesis of streamflow sequences

for the analysis of river basins by simulation, in A. Maas et al., Eds. Design of Water Resources Systems. Harvard University Press, Cambridge, Massachusetts, 459-93.

VandeLinde, V.D., R. Daraiswami, and H.O. Yurtseven (1972). Robust filtering for linear systems. IEEE Decision and Control Conference.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. Trans. An. Math. Soc. 54: 426-82.

Whittle, P. (1951). Hypothesis Testing in Time Series Analysis. Almquist and Wiksell, Uppsala, Sweden.

Whittle, P. (1952). Tests of fit in times series. Biometrika 39: 309-318.

Whittle, P. (1952a). The simultaneous estimation of a time series' harmonic and covariance structure. Trab.-Estad. 3: 43-57.

Whittle, P. (1954). Statistical investigation of sunspot observations. Astr. Phys. J. 120: 251-60.

Whittle, P. (1954a). On stationary processes in the plane. Biometrika 41: 434-49.

Whittle, P. (1954b). The statistical analysis of a Seiche record. J. Marine Res. 13: 76-100.

Wiggins, R.A., and E.A. Robinson (1965). Recursive solution to the multichannel filtering problem. J. Geophys. Res. 70: 1885-1981.

Willsky, A. (1977). Multiparameter Systems, Distributed Processes, and Random Fields. M.I.T. Press, Cambridge, Massachusetts.

Wishner, R.P., R.E. Larson, R.K. Mehra, and M. Athans (1971). Filters of varying complexity for radar target tracking. Joint Automatic Control Conference, St. Louis, Missouri, August.

Wold, H. (1938). The Analysis of Stationary Time Series. Almquist and Wicksell, Uppsala, Sweden.

Wold, H. (1965). Bibliography on Time Series and Stochastic Processes. Oliver and Boyd, Edinburgh.

Wong, K.Y., and E. Polak (1967). Identification of linear discrete time systems using the instrumental variable method. IEEE Trans. Autom. Control 12: 707-718.

Young, P.C. (1970). An instrument variable method for real-time identification of a noisy process. Automatica 6: 271-87.

Yule, G.U. (1927). On the method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. Philos. Trans. R. Soc. London. A 226: 267.

Zaremba, S.K. (1967). Kurtosis and determiniate components in linear processes. Rev. Inst. Internat. Statist. 33: 378-413.

Zaremba, S.K. (1972). Tests for the presence of trends in linear processes. Warszawa Państw. Wyd-wo. nauk.

Part One

Forecasting Problems of Water Resource Systems

Introduction

Hydrologic prediction or forecasting models, up until about six years ago, implied physically based models that would simulate the behavior of the hydrologic system for given (deterministic and error-free) inputs. Recently, models based on concepts from time-series analysis and the control literature have been proposed for hydrologic forecasting. The papers in this volume attest to the progress that has been made in applying these new techniques to hydrologic problems.

In this volume forecasting and prediction are used interchangeably to mean the determination of future states of a system, given the past and current state and any future, deterministic inputs. In the control theory or time-series literature, forecasting problems are referred to as "prediction" problems when the underlying model has causal or physical structure related to the process, and "forecasting" problems when the underlying model is a statistical black box model without any apparent correspondence to the process. Most of the state-space models presented here represent or reflect physical and causal relationships of the hydrologic systems. This is not a requirement. A state-space representation can be constructed for either causal models, time-series models, or a mix of the two. It is possible to develop a state-space model consisting of a causal rainfall-runoff model for stream discharges and an autoregressive time-series model for the precipitation.

The problem of applying state-estimation techniques to hydrologic forecasting can be divided into two interrelated subproblems: model development, which is usually referred to as system identification, and optimal forecasting.

SYSTEM IDENTIFICATION

System identification, as it is generally referred to in the control theory literature, denotes the iterative procedure of model specification, parameter estimation, and model verification using noisy operative input and output data. Figure 2.1 illustrated these steps. Chapter 2, by Mehra, surveys many of the current techniques. The papers in this part develop particular techniques more completely and apply them to general hydrologic forecasting problems. The purpose of this introduction is to present the framework and concepts in which the following papers can be interpreted.

While forecasting on-line, using recursive algorithms, is the aim of the operational

mode of the proposed models, system identification can be carried out either prior to or with the forecasting mode. The first method is often called off-line identification or one-shot identification and the latter is generally referred to as on-line identification.

Off-Line Identification

Off-line or one-shot identification refers to those identification techniques that do not process data as it becomes available (that is, recursively) but rather use a set of data collected prior to the identification stage. Once the model structure and parameters have been identified and verified, the model is used, in a recursive, on-line mode, for forecasting.

For identifying the order of the model, the Akaike Information Criteria (AIC) is one of the most practical and simple methods. It is used by Katayama (Chapter 4) and by Tamura and Kawaguchi (Chapter 3) who use the one-shot AIC method and modify it for use in a recursive algorithm.

Maximum likelihood identification, which uses the innovation property of the optimal filter, is also widely used for off-line identification of parameters when the structure of the model is assumed to be known. Given a linear discrete-time system

$$x(t+1) = F.x(t) + G.u(t) + \Gamma.w(t) \tag{1}$$

$$z(t) = H.x(t) + v(t) \tag{2}$$

where

$x(t)$ = nxl state vector

$u(t)$ = pxl input vector

$w(t)$ = qxl vector of random model errors

$z(t)$ = rxl output vector

$v(t)$ = rxl vector of measurement errors

with

$$E[w(t)] = E[v(t)] = 0$$

$$E[w(t).w^T It)] = Q\delta_{ij}$$

$$E[v(t).v^T(t)] = R\delta_{ij}$$

$$\text{and} \quad E[v(t).w^T(\tau)] = 0 \qquad \text{for all } t \text{ and } \tau$$

where δ_{ij} is the Kronecker delta function.

The vector of unknown parameters from F, G, Γ, H, Q and R is denoted as Θ. Assuming Θ is identifiable, the maximum likelihood (ML) estimator is given by

$$\hat{\Theta} = \text{Arg} \{\max_{\Theta} \log p(Z_N|\Theta)\} \tag{3}$$

where

$$Z_N = [z(1), z(2),...,z(N)]$$

and

$p(Z_N|\Theta)$ = conditional probability density of Z_N given Θ.

Mehra (1974) and Schweppe (1974) present more complete algorithms. Katayama (Chapter 4) uses ML for his parameter estimation.

ML estimators are widely used for estimating the error covariance matrices. In fact, the optimal steady-state Kalman gain K can be estimated in place of R and Q by using an innovation representation of the state-space equations. This is discussed by Katayama (Chapter 4) as well as by Mehra (1974).

On-Line Identification

On-line identification allows for the updating of the model structure or parameter values in real-time whenever data are received. On-line identification is done in parallel with the updating of the state estimates. The advantages of identifying models recursively are that

Identification can be done on minicomputers since the processing of large quantities of data is avoided

It is possible to identify slowly varying parameters in the system

For identifying the order of the model, Tamura and Kawaguchi (Chapter 3) used a modified form of the AIC in a recursive form. Recursive parameter estimation techniques have received far more attention in the literature than the problem of determining the structure of the model. In water resources, many models have physical cause-and-effect relationships that make the determination of the model structure quite easy. Thus, it is the parameter estimation problem that is crucial for optimal forecasting.

For on-line parameter identification, the most versatile technique uses instrumental variables. Mehra discusses the technique briefly in Chapter 2. Another approach which has been found useful when parameters can vary over time is the combination of a Gauss-Markov model for the parameters $\Theta(t)$ and an output equation that relates measured output to the parameters. This set of equations is as follows:

$$\Theta(t) = F\Theta(t-1) + \Gamma w(t)$$

$$z(t) = H(t).\Theta(t) + r(t)$$

The on-line tracking can be performed by using a Kalman filter. Examples of this formulation are given in Chapter 5 and in Wood and Szöllösi-Nagy (1978). In Chapter 5, Moore and Weiss use an extended Kalman filter to handle the nonlinear input-output function H(t).

Other recursive techniques may also be used even though the computational effort may be prohibitive. For example, recursive Bayesian techniques, utilizing the normality assumption of the error terms have been used (for example, Smith 1967). Maximum likelihood estimators, which have attractive statistical properties, can be estimated on-line by using a single iteration of the Gauss-Newton method (Mehra 1972), or can be used as a basis for developing techniques (often called "approximate maximum likelihood techniques") that are recursive and for which the computation is easy.

STATE PREDICTION AND OPTIMAL FORECASTING

System identification, parameter estimation, and state prediction can be and is viewed as one integrated effort in time series analysis. The early work in forecasting for stochastic processes and time series analysis dealt with stationary stochastic processes either for discrete-time (Kolmogorov 1941) or continuous-time (Wiener 1949) processes. Kalman (1960) and Kalman and Bucy (1961) extended these results to nonstationary processes with finite observation intervals. The Kalman filter can be applied widely, its state vector formulation allows consideration of both process and observation noise, and it is easily implemented with recursive algorithms. These are the main reasons for the popularity of the Kalman filter.

Derivation of the Kalman Filter

Consider a discrete, linear, stochastic, time-invariant dynamic system whose behavior is described by the state equation

$$x(t+1) = \phi \cdot x(t) + w(t) \tag{4}$$

where

$x(t)$ is an n-vector of state variables

ϕ is a transition matrix of size $n \times n$

$w(t)$ is an n-vector of Gaussian white noise with statistics $w(t) \sim N(0,Q)$

Also consider a linear measurement (or output) equation of the form

$$z(t) = H.x(t) + v(t) \tag{5}$$

where

$z(t)$ is an n-vector of the measured output vector

h is an $m \times n$ measurement matrix

$v(t)$ is an m-vector of measurement errors with statistics $v(t) \sim N(0,R)$; the covariance matrix R is positive definite

It is also assumed that the noise processes are uncorrelated with one another, that is

$$E[v(t)^T.w(\tau)] = 0 \quad \text{for all } t \text{ and } \tau$$

and both processes are independent of the state which has the initial statistics $x(0) = N(\hat{x}(0),P(0))$.

Let $\hat{x}(t|t)$ denote the best linear unbiased estimator of $x(t)$ based on measurements $Z(t) = \{z(1),z(2),\ldots,z(t)\}$. $\hat{x}(t|t)$ is the filtered estimator of $x(t)$ since it is the estimate of the state vector at the current time based upon all measurements

including the current one. It will be shown that the optimal estimator for the filtering problem will provide us with the optimal estimator for the prediction problem.

Kalman (1960) rigorously derived the form of a recursive filter for the above system - that is, a system with known structure, parameters, and error statistics.

The problem can be stated as follows: given a measurement sequence Z(t) observed by the measurement equation (5), estimate the state of the dynamic system (4) such that the error of estimate, $\tilde{x}(t|t) = x(t) - \hat{x}(t|t)$, will minimize the quadratic performance function

$$J = E[\tilde{x}(t|t).L.\tilde{x}^T(t|t)] \tag{6}$$

where L is any positive semidefinite matrix and, for simplicity, is chosen to be the identity matrix I.

Assume that a prior estimate $\hat{x}(t|t-1)$ of the system state x(t) at t is based upon previous measurements up to t-1. An updated estimate $\hat{x}(t|t)$ is desired which takes into account the new measurement z(t) at t. Consider this updated estimate as the linear combination of the prior estimate and the new (noisy) measurement; thus

$$\hat{x}(t|t) = \tilde{K}(t)\hat{x}(t|t-1) + K(t)z(t) \tag{7}$$

where $\tilde{K}(t)$ and K(t) are time-varying weighting matrices as yet unspecified. Introducing (5) into (7) and utilizing the statistics of the noise process, it can be seen that (7) is an unbiased estimate only if $\tilde{K}(t) = I-K(t)H$. Hence, the state estimator $\hat{x}(t|t)$, using the new measurement z(t), is of the form

$$\hat{x}(t|t) = \hat{x}(t|t-1) + K(t)[z(t) - H\hat{x}(t|t-1)] \tag{8}$$

where K(t) is still unspecified, and the initial condition at t=0 for the state estimation is $\hat{x}(0|0) = \hat{x}(0)$. A measure of accuracy for the estimate can be expressed by the covariance matrix P(·) of the prediction error defined as

$$P(t|t) = E[\tilde{x}(t|t).\tilde{x}^T(t|t)] \tag{9}$$

the initial condition of which is P(0|0) = P(0). Using (7) with the measurement noise statistics, it can be shown that the covariance matrix of $\hat{x}(t|t)$ can be projected from that of $\hat{x}(t|t-1)$ as

$$P(t|t) = (I-K(t)H)P(t|t-1)(I-K(t)H)^T + K(t)RK^T(t) \tag{10}$$

Since the loss function (6) is the trace of the error covariance matrix (9), the problem is to minimize the <u>trace norm</u> $||P(t|t)||$ of P(t|t), that is, the length of the estimation error vector. Using the properties of matrix derivatives, it can be seen that the weighting matrix K(t) can be obtained from

$$\frac{\partial}{\partial K(t)} \, ||P(t|t)|| = 0$$

as

$$K(t) = P(t|t-1)H^T[HP(t|t-1)H^T + R]^{-1} \tag{11}$$

which is referred to as the Kalman gain matrix. Equation (11) can be used to simplify (10) into

$$P(t|t) = (I - K(t).H)P(t|t-1) \tag{12}$$

The one-step-ahead prediction of the state vector, given observations up to t, is

$$\hat{x}(t+1|t) = \phi.\hat{x}(t|t) \tag{13}$$

The propagation of prediction errors $P(t|t) \rightarrow P(t+1|t)$ can be determined by computing the predicted error covariance matrix, $P(t+1|t) = E[\tilde{x}(t+1|t).\tilde{x}^T(t+1|t)]$. Using (4), (9), and the independence of the error terms, we find that

$$P(t+1|t) = \phi.P(t|t).\phi^T + Q \tag{14}$$

It can be shown that the one-step-ahead estimate of (13) will minimize (14); thus by finding the optimal estimator for the filtering problem, the optimal estimator for the prediction problem is obtained by (13). The optimal filter is summarized in Table 1.

Table 1. Summary of discrete Kalman filter equation

System model (Eq. 4)	$x(t+1) = \phi.x(t) + w(t)$
Measurements (Eq. 5)	$z(t) = H.x(t) + v(t)$
Initial condition and other assumptions	$E[x(0)]=\hat{x}(0), E[x(0).x^T(0)]=P(0)$ $w(t)\sim N(0,Q), v(t)\sim N(0,R)$ $E[w(t)v^T(\tau)] = 0$ for all t,τ
State estimate update (Eq. 8)	$\hat{x}(t\|t)=\hat{x}(t\|t-1)+K(t)[z(t)-H.\hat{x}(t\|t-1)]$
Error covariance update (Eq. 12)	$P(t\|t)=[I-K(t)H].P(t\|t-1)$
Kalman gain matrix (Eq. 11)	$K(t)=P(t\|t-1)H^T[H.P(t\|t-1).H^T+R]^{-1}$
State estimate prediction (Eq. 13)	$\hat{x}(t+1\|t)=\phi.\hat{x}(t\|t)$
Error covariance prediction (Eq. 14)	$P(t+1\|t)=\phi.P(t\|t).\phi^T+Q$

The results are exact for linear systems and only approximate for nonlinear models. It should be noted that the problem of linear dynamics with unknown parameters becomes a nonlinear estimation problem. Time-varying $\phi(t)$, $H(t)$, $R(t)$, or $Q(t)$ does not cause any problems as long as they are known.

More complete proofs, and proofs for the case of continuous linear dynamic systems can be found in many tests, for example, see Jazwinski (1970).

LITERATURE REVIEW

Modern state estimation techniques have not, as yet, found wide application in hydrology. The purpose of this review is to cite some of the work that has been done to date. For a general survey on forecasting techniques, the reader is referred to the paper by Mehra (Chapter 2), and Olsson (Chapter 6), who looks mainly at developments in water quality and wastewater facilities.

Some of the first work in hydrology to try to apply results from the early filter theory work included the work of Eagleson et al. (1966), who modeled runoff as a single-input, single-output linear system and then inverted the Wiener-Hopt equations to obtain the transfer function for the system. Similar work was done by Barrera and Perkins (1967) for multiple storms to represent time-varying systems. The identification problem was also considered by Amorocho (1967), using nonlinear transfer functions, and by Blank and Delleur (1968), using Fourier transform methods.

These and similar analyses regarded the hydrologic systems as a black box; that is, system transfer functions were derived from observed inputs and outputs without regard to the form of the functions or physical constraints that should be considered. This, along with certain computational instability problems, restricted the practical applications of the techniques.

Recent formulations, which use the state-space approach, have avoided many of the above problems. Many hydrologic processes can be described by differential or difference equations and the state-space framework provides an intuitively appealing way to relate time series analysis to deterministic modeling. Certainly, the filtering theory results that have been applied in other areas, most notably aerospace navigation and guidance, have provided motivation to test the concepts in disciplines like hydrology.

Some workers in rainfall-runoff include Hino (1970, 1973) and Wood and Szöllösi-Nagy (1978), who used the Kalman filter recursively to estimate the parameters of the rainfall-runoff response function. Hino (1973) assumed perfect observations and did not forecast future discharges, and Wood and Szöllösi-Nagy (1978) recursively estimated the error covariance matrices and forecast future discharges based on the recursively estimated response function. Duong et al. (1975) used an extended Kalman filter applied to the Prasad model and Toyoda et al. (1969) investigated the stability and adaptivity of a state-space runoff prediction model as applied to the Sagar River in Japan. Todini and Bouillot (1975) developed a Kalman filter for a linear runoff model.

Other areas of hydrology include: rain-gauge network design (Bras and Rodriguez-Iturbe 1975, 1976), typhoon forecasting (Takeuchi 1976), groundwater modeling (McLaughlin 1975) and water supply consumption (Fallside and Perry 1975).

There have also been a number of applications in water quality and wastewater treatment. Olsson (Chapter 4) surveys many of these applications. Also, in water quality monitoring, there have been applications by Moore (1973), Lettenmaier (1975) and De Guida et al. (1977). In water quality identification for parameter

estimation of dissolved oxygen/biochemical oxygen demand (DO/BOD) models, there is the work by Koivo and Phillips (1971, 1972, 1976) as well as work by Beck (1976) and Beck and Young (1975, 1976).

The papers in this part of the volume are structured into two groups. The first group deals primarily with the system identification parameter estimation problem. The first four papers structure the problem within a state-space formulation and present detailed discussions and results of many of the techniques surveyed by Mehra in Chapter 2.

The second group of papers deals primarily with the forecasting or prediction of future states. These papers cover Chapters 7 to 10; the first two focus on a Kalman filter algorithm for the predictions. The remaining papers present other related techniques. These include a self-tuning predictor (Chapter 9) and a heuristic self-organization method (Chapter 10).

REFERENCES

Amorocho, I. (1967). The non-linear prediction problem in the study of the runoff cycle. Water Resour. Res. 3(3).

Barrera, A., and F.E. Perkins (1967). An extension of the role of linear systems analysis in hydrologic theory. Hydrodynamics Lab. Rep. No. 106, M.I.T. Dept. of Civil Engineering, Cambridge, Massachusetts, September.

Beck, M.B. (1976). Identification and parameter estimation of biological process models, in G.C. Vansteenkiste, Ed., System Simulation in Water Resources. North-Holland, Amsterdam.

Beck, M.B., and P.C. Young (1975). A dynamic model for DO-BOD relationships in a non-tidal stream. Water Res 9: 769-776.

Beck, M.B., and P.C. Young (1976). Systematic identification of DO-BOD model structure. Proc. ASCE, J. Env. Eng. Div. 102 (EE5): 909-927.

Blank, D., and J.W. Delleur (1968). A program for estimating runoff from Indiana watersheds, Part I: Linear system analysis in surface water hydrology and its applications to Indiana water-sheds. Technical Report No. 4, Purdue University, Water Resources Research Center, Lafayette, Indiana, August.

Bras, R., and I. Rodriguez-Iturbe (1975). Rainfall-runoff as a spatial stochastic process: Data collection and synthesis. Hydrodynamics Lab Report No. 196, M.I.T. Dept. of Civil Engineering, Cambridge, Massachusetts.

Bras, R., and I. Rodriguez-Iturbe (1976). Rainfall network design for runoff prediction. Water Resour. Res. 12(6): 1197-1208.

De Guida, R.N., J.J. Connor, and B.R. Pearce (1977). Application of estimation theory to design of sampling programs for verification of coastal dispersion prediction, in W.G. Gray, G.F. Pinder, and C.A. Brebbia, Eds., Finite Elements in Water Resources. Pentech Press, London.

Duong, N., C.B. Winn, and G.R. Johnson (1975). Modern control concepts in hydrology. IEEE Trans. Syst. Sci. Cybern. 6(4): 322-329.

Eagleson, P.S., R. Mejia, and F. March (1966). The computation of optimum realizable unit hydrographs. Water Resour. Res. 2(4): 755-764.

Fallside, F. and P.F. Perry (1975). On-line prediction of consumption for water supply network control. Preprints, 6th Triennial World IFAC Congress, Boston, Massachusetts.

Hino, M. (1970). Runoff forecasts by linear predictive filter. Proc. ASCE, J. Hyd. Div. 96(HY3): 681-701.

Hino, M. (1973). On-line prediction of a hydrologic system, presented at XV Congress of IAHR, Istanbul, September.

Jazwinski, A.H. (1970). Stochastic Processes and Filter Theory. Academic Press, New York.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Trans. ASME, J. Basic Eng. 82: 35-45.

Kalman, R.E., and R.S. Bucy (1961). New results in linear filtering and prediction theory. Trans. ASME, J. Basic Eng. 83: 95-107.

Koivo, A.J., and G.R. Phillips (1971). Identification of mathematical models for DO and BOD concentration in polluted streams from noise corrupted measurements. Water Resour. Res. 7(4): 853-862.

Koivo, A.J. and G.R. Phillips (1972). On determination of BOD- and DO-parameters in polluted stream models from DO measurements only. Water Resour. Res. 8(2): 478-486.

Koivo, A.J. and G.R. Phillips (1976). Optimal estimation of DO, BOD, and stream parameters using a dynamic discrete time model. Water Resour. Res. 12(4): 705-711.

Kolmogorov, A.N. (1941). Interpolation and extrapolation of stationary random sequences. Bull. Moscow University, Ser. Math. 5, Moscow.

Lettenmaier, D.P. (1975). Design of monitoring systems for detection of trends in stream quality. Technical Report No. 39, University of Washington, Hydraulic Laboratory, Seattle, Washington.

McLaughlin, D.B. (1975). Investigation of alternative procedures for estimating groundwater basin parameters. Report to Office of Water Research and Technology, U.S. Dept. of Interior, Washington, D.C.

Mehra, R.K. (1972). Approaches to adaptive filtering. IEEE, Trans. Autom. Control, Oct.

Mehra, R.K. (1974). Topics in stochastic control theory: Identification in control and econometrics; Similarities and differences. Annals of Economic and Social Measurements 3(1): 21-47.

Moore, S.F. (1973). Estimation theory application to design of water quality monitoring systems. Proc. ASCE, J. Hyd. Div. HY5: 815-831.

Schweppe, F.C. (1974). Uncertain Dynamic Systems. Prentice-Hall, Englewood Cliffs, New Jersey.

Smith, G.L. (1967). Sequential estimation of observation error variance in a trajectory estimation problem. AIAA. J. 5: 1961-1970, Nov.

Takeuchi, K. (1976). Application of the Kalman filter to cyclone forecasting.

RM-76-9, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Todini, E., and D. Bouillot (1975). A rainfall-runoff Kalman filter model, in G.C. Vanteenkiste, Ed., System Simulation in Water Resources. North-Holland, Amsterdam.

Toyoda, J., N. Toriumi, and Y. Inoue (1969). An adaptive predictor of river flow for on-line control of water resource systems. Automatica 5: 175-181.

Wiener, N. (1949). The Extrapolation, Interpolation and Smoothing Stationary Time Series. Wiley, New York.

Wood, E.F. and A. Szöllösi-Nagy (1978). An adaptive algorithm for analyzing short-term structural and parameter changes in hydrologic prediction models. Water Resour. Res. 14(4): 577-581.

3 Real-Time Estimation of Orders and Parameters of Distributed-Lag Models of River Quality

Hiroyuki Tamura and Tatsuro Kawaguchi

The Streeter-Phelps model (1925) and its modifications have been widely used for analysis and control of river quality, where it is assumed that concentration of biochemical oxygen demand (BOD) and dissolved oxygen (DO) suffice to describe the biochemical processes in rivers. As an accurate and simple BOD-DO model of river quality, distributed-lag models (Tamura 1974, Rinaldi et al. 1978) have been proposed. The discrete time, distributed-lag model of river quality, which can be interpreted as an approximate Streeter-Phelps dispersion model (Rinaldi et al. 1978) is described by a multidimensional linear difference equation of high order, and can take into account diffusion and dispersion. However, we need to estimate the order and parameters of the model on the basis of observation data of BOD and DO (or, when possible, of DO only).

Among the various methods of statistical system identification (Box and Jenkins 1970; Parzen 1974, 1975; Akaike 1970, 1971, 1974), Akaike's method is one of the most practical and simple. However, the method is a one-shot procedure, which uses a fixed set of time series data. Hence, the method is not suited for on-line, real-time use.

In this paper, Akaike's method of statistical system identification is modified and used recursively for real-time river quality modeling; the modified method can update the model in real time whenever observation data are obtained. The advantages of identifying the river quality system recursively, instead of using a one-shot procedure, are the following:

- The convergence profile of the identification can be followed in real time so that the procedure can be stopped when the identified results do not vary significantly.
- It is possible to identify slowly varying orders and parameters of the system.
- There is no need to store large amounts of old data on the computer.

This work was supported in part by the Ministry of Education in Japan for the science research program, "Monitoring and Control of Environmental Pollution", under Grant No. 111219.

A DISTRIBUTED-LAG MODEL OF RIVER QUALITY

Considering a stretch of a river composed of a sequence of N reaches, a continuous-time, distributed-lag model of river quality can be derived from the Streeter-Phelps dispersion model and can be written as (Rinaldi et al. 1978)

$$\dot{b}_i = -(k_{1i} + \frac{Q_i}{V_i})b_i + \frac{Q_{i-1}}{V_i}b^*_{i-1}(t) + \frac{Q_{Ei}}{V_i}E_i(t) \tag{3.1a}$$

$$\dot{d}_i = k_{1i}b_i - (k_{2i} + \frac{Q_i}{V_i})d_i + \frac{Q_{i-1}}{V_i}d^*_{i-1}(t) + \frac{Q_{Ei}}{V_i}p_i(t) + \frac{r_i}{V_i}$$

where

$$b^*_{i-1}(t) = \int_0^t \phi_{i-1}(\tau)b_{i-1}(t-\tau)d\tau \tag{3.1b}$$

$$d^*_{i-1}(t) = \int_0^t \phi'_{i-1}(\tau)b_{i-1}(t-\tau)d\tau + \int_0^t \psi_{i-1}(\tau)d_{i-1}(t-\tau)d\tau$$

$$i = 1,2,\ldots,N$$

$$b_0(t)\ ,\quad d_0(t) \qquad \text{given.}$$

In (3.1) variables and parameters with subscript i denote variables and parameters for the reach i, and

- b = BOD concentration (mg/l)
- d = DO deficit concentration (mg/l)
- Q = river flow rate (m^3/day)
- V = volume of water in the reach (m^3)
- k_1 = BOD decay rate and deoxygenation rate (day^{-1})
- k_2 = reoxygenation rate (day^{-1})
- E = BOD concentration of effluent discharge (mg/l)
- p = DO deficit concentration of effluent discharge (mg/l)
- Q_E = flow rate of effluent discharge (m^3/day)
- r = net addition of DO deficit by the combined effects of photosynthesis, respiration, and bottom sludge (mg/day)
- ϕ, ϕ', ψ = parameters for distributed lag

Figure 3.1 shows the schematic diagram of a stretch of a river composed of N reaches. Derivation of (3.1) from the Streeter-Phelps dispersion model can be found in Rinaldi et al. (1978). Equation (3.1a), which represents the mass balance of BOD in reach i, can be interpreted as follows. The evolution of BOD is described

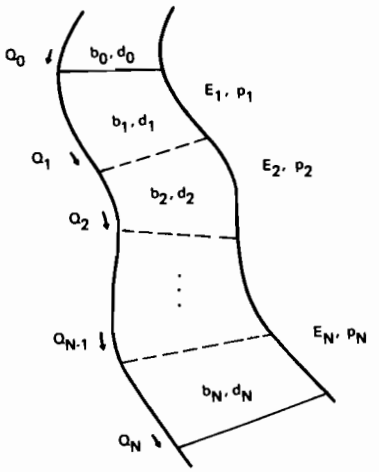


Fig. 3.1 A stretch of a river composed of N reaches.

as the sum of BOD removal by self-purification in reach i and by the transport from reach i to reach (i+1), BOD addition from reach (i-1) to reach i, and BOD addition by the effluent discharge into reach i. Between the adjacent reaches we assume a hypothetical channel that represents distributed time lag, whose mathematical model can be written as (3.1b). By suitably selecting $\phi(\tau)$, $\phi'(\tau)$, and $\psi(\tau)$, we could obtain dynamics of river quality close to those of the dispersion model.

Only discrete-time measurements every Δt days are assumed to be available, so we first need to make model (3.1) discrete with respect to time. By integrating (3.1a) from time $k\Delta t$ to $(k+1)\Delta t$ and writing, for simplicity, $b_i(k)$ and $d_i(k)$ instead of $b_i(k\Delta t)$ and $d_i(k\Delta t)$, we obtain

$$b_i(k+1) = e^{-(k_{1i} + \frac{Q_i}{V_i})\Delta t} b_i(k) + \frac{Q_{i-1}\Delta t}{V_i} b^*_{i-1}(k) + \frac{Q_{Ei}\Delta t}{V_i} E_i(k)$$

$$d_i(k+1) = \frac{k_{1i}}{k_{1i}-k_{2i}}\left(e^{-(k_{2i} + \frac{Q_i}{V_i})\Delta t} - e^{-(k_{1i} + \frac{Q_i}{V_i})\Delta t}\right) b_i(k)$$

$$+ e^{-(k_{2i} + \frac{Q_i}{V_i})\Delta t} d_i(k) + \frac{Q_{i-1}\Delta t}{V_i} d^*_{i-1}(k) + \frac{Q_{Ei}\Delta t}{V_i} p_i + \frac{r_i \Delta t}{V_i} \qquad (3.2)$$

$$i = 1,2,\ldots,N \qquad k = 0,1,\ldots,T-1$$

The discrete version of (3.1b) can be written as

$$b^*_{i-1}(k) = \sum_{j=0}^{\Theta} \phi_{i-1}(j) b_{i-1}(k-j)$$

$$d^*_{i-1}(k) = \sum_{j=0}^{\Theta} \phi'_{i-1}(j) b_{i-1}(k-j) + \sum_{j=0}^{\Theta} \psi_{i-1}(j) d_{i-1}(k-j) \tag{3.3}$$

Eqs. (3.2) and (3.3) can be combined as

$$b_i(k+1) = \alpha_i b_i(k) + \frac{Q_{i-1}}{V_i} \sum_{j=0}^{\Theta} \phi_{i-1}(j) b_{i-1}(k-j) + \frac{Q_{Ei}}{V_i} E_i(k) \tag{3.4a}$$

$$d_i(k+1) = \alpha'_i b_i(k) + \beta_i d_i(k) + \frac{Q_{i-1}}{V_i} \sum_{j=0}^{\Theta} \phi'_{i-1}(j) b_{i-1}(k-j)$$

$$+ \frac{Q_{i-1}}{V_i} \sum_{j=0}^{\Theta} \psi_{i-1}(j) d_{i-1}(k-j) + \frac{Q_{Ei}}{V_i} p_i(k) + \frac{r_i}{V_i} \tag{3.4b}$$

$$i = 1,2,\ldots,N \qquad k = 0,1,\ldots,T-1$$

where

$$\alpha_i = e^{-(k_{1i} + \frac{Q_i}{V_i})\Delta_t}$$

$$\alpha'_i = \frac{k_{1i}}{k_{1i}-k_{2i}} \left(e^{-(k_{2i} + \frac{Q_i}{V_i})\Delta t} - e^{-(k_{1i} + \frac{Q_i}{V_i})\Delta t} \right)$$

$$\beta_i = e^{-(k_{2i} + \frac{Q_i}{V_i})\Delta t}$$

and Q_i, Q_{i-1}, Q_{Ei}, and r_i stand for $Q_i\Delta t$, $Q_{i-1}\Delta t$, $Q_{Ei}\Delta t$, and $r_i\Delta t$, i.e., Δt is the unit of time. The boundary conditions $b_0(t)$, $d_0(t)$, and the initial conditions are assumed to be given. Eq. (3.4) is the discrete-time, distributed lag model for river quality, which was proposed by Tamura (1974).

Eq. (3.4) can be written in vector-matrix form as

$$\underline{x}(k+1) = A_0\underline{x}(k) + A_1\underline{x}(k-1) + \ldots + A_\Theta\underline{x}(k-\Theta) + \underline{u}(k) \tag{3.5}$$

where

$$\underline{x}(k) = [b_1(k), d_1(k), \ldots, b_N(k), d_N(k)]^T$$

$$A_0 = \begin{bmatrix} A_{011} & 0 & \cdots & 0 & 0 \\ A_{021} & A_{022} & \cdots & 0 & 0 \\ 0 & A_{032} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & A_{0,N,N-1} & A_{0NN} \end{bmatrix}$$

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ A_{j21} & 0 & \cdots & 0 & 0 \\ 0 & A_{j32} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & A_{j,N,N-1} & 0 \end{bmatrix} \qquad j = 1,2,\ldots,\Theta$$

$$A_{0ii} = \begin{bmatrix} \alpha_i & 0 \\ \alpha_i' & \beta_i \end{bmatrix} \qquad i = 1,2,\ldots,N$$

$$A_{j,i,i-1} = \frac{Q_{i-1}}{V_i} \begin{bmatrix} \phi_{i-1}(j) & 0 \\ \phi_{i-1}'(j) & \psi_{i-1}(j) \end{bmatrix} \qquad \begin{matrix} i = 2,3,\ldots,N \\ j = 0,1,\ldots,\Theta \end{matrix}$$

$$\underline{u}(k) = [\underline{u}_1(k)^T, \underline{u}_2(k)^T, \ldots, \underline{u}_N(k)^T]^T$$

$$\underline{u}_1(k) = \left[\frac{Q_0}{V_1} b_0(k) + \frac{Q_{E1}}{V_1} E_1(k), \frac{Q_0}{V_1} d_0(k) + \frac{Q_{E1}}{V_1} p_1 + \frac{r_1}{V_1}\right]^T$$

$$\underline{u}_i(k) = \frac{Q_{Ei}}{V_i} E_i(k), \frac{Q_{Ei}}{V_i} p_i + \frac{r_i}{V_i}^T, \qquad i = 2,3,\ldots,N$$

Eq. (3.5) can be thought of as a general discrete-time, distributed-lag model, for which we need to estimate the order Θ and the parameters $A_0, A_1, \ldots, A_\Theta$ of the model by using the time series data $\underline{x}(k)$ and $\underline{u}(k)$.

THE MODIFIED AKAIKE'S METHOD

Akaike's Method

Akaike's method fits n-dimensional time series data $\underline{x}(k)$, $k = 1,2,\ldots,N$, to an autoregressive (AR) model of finite order. This method can estimate the order and the parameters of the model. For simplicity, let the mean of $\underline{x}(k)$ be 0. The order M of the AR model should be chosen such that the residual series

$$\underline{e}(k) = \underline{x}(k) - \sum_{m=1}^{M} A_M(m)\underline{x}(k-m) \tag{3.6}$$

of the AR model can be considered as a white noise process. Akaike (1970) proposed a criterion for estimating the order M such that the multivariate final prediction error (MFPE),

$$\mathrm{MFPE}(M) = \left(1 + \frac{Mn + 1}{N}\right)^n \left(1 - \frac{Mn + 1}{N}\right)^n \det(D_M) \tag{3.7}$$

is minimized with respect to M, D_M denotes the estimates of covariance matrix for $\underline{e}(k)$, and $\det(\cdot)$ denotes determinant. The MFPE given by (3.7) gives the final prediction error for the multivariate AR model. The important thing to note is that the predictive quantities of a model do not improve monotonically with the increase of M even though the fitting error $\det(D_M)$ decreases monotonically with the increase of M. In other words, Akaike's criterion takes into account the error covariance and the amount of data N. The estimates of the model parameters $A_M(m)$ can be obtained with the least squares method. The normal equation for estimating parameters can be written with matrices as

$$\begin{bmatrix} C^N(0) & C^N(1) & \cdots & C^N(M-1) \\ C^N(1)^T & C^N(0) & \cdots & C^N(M-2) \\ \vdots & \vdots & & \vdots \\ C^N(M-1)^T & C^N(M-2)^T & \cdots & C^N(0) \end{bmatrix} \begin{bmatrix} \hat{A}_M(1)^T \\ \hat{A}_M(2)^T \\ \\ \hat{A}_M(M)^T \end{bmatrix} = \begin{bmatrix} C^N(1)^T \\ C^N(2)^T \\ \vdots \\ C^N(M)^t \end{bmatrix} \tag{3.8}$$

where

$$C^N(\ell) = \frac{1}{N} \sum_{k=1}^{N-\ell} \underline{x}(k+\ell)\underline{x}(k)^T , \quad \ell = 0,1,\ldots,L \tag{3.9}$$

is the estimate of the covariance matrix for lag ℓ, and L is a prescribed maximum number of the order. For simplicity, this normal equation is written as

$$R_M^N \hat{A}_M^N = \Theta_M^N . \tag{3.10}$$

In the following, N or M will sometimes be omitted. For each M, this equation can be solved using the recursive relations with respect to M. That is, in the process of obtaining $\hat{A}_L(m)$, $m = 1,2,\ldots,L$, the parameter estimates $\hat{A}_M(m)$, $m = 1,2,\ldots,M$ and D_M for each M up to L can be obtained. Computing MFPE(M) for all M, the order and the parameters that minimize MFPE(M) can be estimated.

A brief summary follows of Akaike's algorithm for time series $\underline{x}^*(k)$, $k = 1,2,\ldots,N$.

Step 1 : Delete the mean from the data as

$$\bar{\underline{x}} = \frac{1}{N} \sum_{k=1}^{N} \underline{x}^*(k)$$

$$\underline{x}(k) = \underline{x}^*(k) - \bar{\underline{x}} , \qquad k = 1,2,\ldots,N .$$

Step 2 : Calculate $C^N(\ell)$, $\ell = 0,1,\ldots,L$.

Step 3 : Set $\hat{A}_0(m) = B_0(m) = [0]$, $m = 1,2,\ldots,L$.

Step 4 : For $M = 0,1,\ldots,L$ calculate

$$D_M = C^N(0) - \sum_{m=1}^{M} \hat{A}_M(m) C^N(m)^T$$

$$E_M = C^N(M+1) - \sum_{m=1}^{M} \hat{A}_M(m) C^N(M+1-m)$$

$$F_M = C^N(0) - \sum_{m=1}^{M} B_M(m) C^N(m)$$

$$G_M = E_M F_M^{-1}$$

$$H_M = E_M^T D_M^{-1}$$

$$\hat{A}_{M+1}(m) = \begin{cases} \hat{A}_M(m) - G_M B_M(M+1-m) , & m = 1,2,\ldots,M \\ G_M , & m = M+1 \end{cases}$$

$$B_{M+1}(m) = \begin{cases} B_M(m) - H_M \hat{A}_M(M+1-m) , & m = 1,2,\ldots,M \\ H_M , & m = M+1 \end{cases}$$

Step 5 : Find the AR order M and the parameter estimates $\hat{A}_M(m)$, $m = 1,2,\ldots,M$ that minimize MFPE(M).

See Chapter 2 for more information about Akaike's information criteria.

Modified Akaike's Method

Akaike's method gives a recursive algorithm with respect to the order of the model, however, it gives only a one-shot procedure with respect to time, that is, the method uses a fixed set of time series data. In this section, a recursive version of Akaike's method which can be used in real time is proposed. It can update the model whenever time series data are added.

In the real time recursive situation, the recursive relationship between $\hat{A}_M(m)$ and $\hat{A}_{M+1}(m)$ cannot be obtained, since the estimates of covariance matrices in the normal

equation (3.8) are evolving with the increase of data; furthermore, the estimate of the order may change with the increase of data. But the inverse of matrix R in (3.10) can be used for estimating the parameters, and updated recursively with the increase of data. The real time recursive algorithm proposed here can be divided into three parts as follows.

- Evolve the estimates of the covariance matrices
- Evolve R^{-1} without changing the order
- Calculate R^{-1} for various orders, and determine the order and the parameters of the model.

The AR model obtained by this method is expected to be updated with the increase of data as shown in Fig. 3.2. From the evolution and the convergence profile of the identified AR model, one can easily find whether the model represents the system satisfactorily or not.

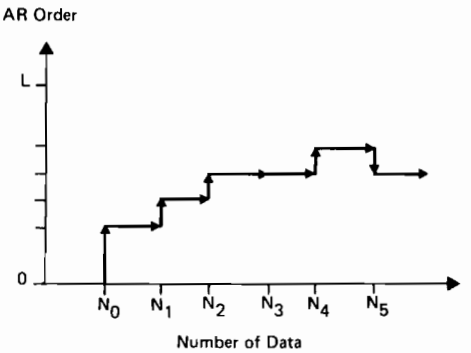


Fig. 3.2 Evolution of the AR model with increasing data.

Evolution of the covariance matrices. When some data are added, the estimates of the covariance matrices $C(\ell)$ are re-estimated. With subtraction of the mean, (3.9) can be written as follows for the initial N data.

$$\underline{x}^{\Sigma N} = \sum_{k=1}^{N} \underline{x}^*(k)$$

$$\bar{\underline{x}}^N = \frac{1}{N} \underline{x}^{\Sigma N}$$

$$\underline{x}^N(k) = \underline{x}^*(k) - \bar{\underline{x}}^N \qquad (3.11)$$

$$S^N(\ell) = \sum_{k=1}^{N-\ell} \underline{x}^N(k+\ell)\underline{x}^N(k)^T, \qquad \ell = 0,1,\ldots,L$$

$$C^N(\ell) = \frac{1}{N} S^N(\ell), \qquad \ell = 0,1,\ldots,L$$

and for the next step we calculate

$$\underline{p}^N(\ell) = \sum_{k=1}^{N-\ell} \underline{x}^N(k) , \qquad \ell = 0,1,\ldots,L$$

$$\underline{q}^N(\ell) = \sum_{k=1}^{N-\ell} \underline{x}^N(k+\ell), \qquad \ell = 0,1,\ldots,L \tag{3.12}$$

Next, we consider the case when K data $\underline{x}^*(k)$, k = N+1, N+2,...,N+K, are added. Deleting the mean is performed as

$$\underline{x}^{\Sigma(N+K)} = \underline{x}^{\Sigma N} + \sum_{k=N+1}^{N+K} \underline{x}^*(k)$$

$$\bar{\underline{x}}^{N+K} = \frac{1}{N+K} \underline{x}^{\Sigma(N+K)}$$

$$\underline{x}^{N+K}(k) = \begin{cases} \underline{x}^*(k) - \bar{\underline{x}}^{N+K} , & k = N+1,N+2,\ldots,N+K \\ \underline{x}^N(k) - \Delta\bar{\underline{x}} , & k = N-L+1,N-L+2,\ldots,N \end{cases} \tag{3.13}$$

where

$$\Delta\bar{\underline{x}} = \bar{\underline{x}}^{N+K} - \bar{\underline{x}}^N .$$

The estimates of the covariance matrices based on N+K data can be obtained from

$$S^{N+K}(\ell) = S^N(\ell) + (N-\ell)\Delta\bar{\underline{x}}\,\Delta\bar{\underline{x}}^T - \Delta\bar{\underline{x}}\,\underline{p}^N(\ell)^T$$

$$- \underline{q}^N(\ell)\Delta\bar{\underline{x}}^T + \sum_{k=N+1}^{N+K} \underline{x}^{N+K}(k)\underline{x}^{N+K}(k-\ell)^T \tag{3.14}$$

$$C^{N+K}(\ell) = \frac{1}{N+K} S^{N+K}(\ell) , \qquad \ell = 0,1,\ldots,L .$$

For the next step we calculate

$$\underline{p}^{N+K}(\ell) = \underline{p}^N(\ell) - (N-\ell)\Delta\bar{\underline{x}} + \sum_{k=N+1}^{N+K} \underline{x}^{N+K}(k-\ell)$$

$$\underline{q}^{N+K}(\ell) = \underline{q}^N(\ell) - (N-\ell)\Delta\bar{\underline{x}} + \sum_{k=N+1}^{N+K} \underline{x}^{N+K}(k) . \tag{3.15}$$

Equations (3.11) to (3.15) give the recursive relationship between $C^N(\ell)$ and $C^{N+K}(\ell)$.

Evolution of inverse matrix. If $(R_M^N)^{-1}$ is known, we can calculate $(R_M^{N+K})^{-1}$ for the unchanged AR order M as follows. Using R_M^{N+K} which is composed of $C^{N+K}(\ell)$, $(R_M^{N+K})^{-1}$ can be obtained by using the iterative formula of second-order convergence

$$X_{k+1} = x_k(2I - R_M^{N+K} X_k) \tag{3.16}$$

Setting $X_0 = (R_M^N)^{-1}$, X_k converges to $(R_M^{N+K})^{-1}$ with sufficient accuracy after a few iteration steps. This iteration is terminated when the matrix in the parentheses converges to a unit matrix as accurate as expected. Hence, the new parameter estimates for unchanged AR order are obtained as

$$\hat{A}_M^{N+K} = (R_M^{N+K})^{-1} \Theta_M^{N+K} . \tag{3.17}$$

These computations correspond to the revision of the system identification along the horizontal arrows in Fig. 3.2.

When the AR order is changed from M to (M+1), $(R_{M+1})^{-1}$ can be obtained from $(R_M)^{-1}$ by means of the matrix inversion identity (Fortmann 1970)

$$(R_{M+1})^{-1} = \begin{bmatrix} C(0) & \vdots & \Theta_M^T \\ \cdots & \vdots & \cdots \\ \Theta_M & \vdots & R_M \end{bmatrix}^{-1} = \begin{bmatrix} D_M^{-1} & \vdots & -D_M^{-1}\Theta_M^T R_M^{-1} \\ \cdots & \vdots & \cdots \\ -R_M^{-1}\Theta_M D_M^{-1} & \vdots & R_M^{-1} + R_M^{-1}\Theta_M D_M^{-1}\Theta_M^T R_M^{-1} \end{bmatrix} \tag{3.18}$$

where

$$D_M = C(0) - \Theta_M^T R_M^{-1} \Theta_M .$$

On the other hand, when the AR order is changed from M to (M-1), $(R_{M-1})^{-1}$ can be obtained as

$$(R_{M-1})^{-1} = T_{M-1} - U_{M-1} D_{M-1} U_{M-1}^T \tag{3.19}$$

where each term in the right-hand side can be obtained from

$$(R_M)^{-1} = \begin{bmatrix} D_{M-1}^{-1} & \vdots & U_{M-1}^T \\ \cdots & \vdots & \cdots \\ U_{M-1} & \vdots & T_{M-1} \end{bmatrix} .$$

Using (3.18) or (3.19), or both, we can obtain $(R_M)^{-1}$ for different orders around the one that was estimated in the previous step, and at the same time MFPE(M) for each M can be evaluated using D_M. Now, let M^N be the AR order estimated by using N data. When K data are added, MFPE(M)s are calculated for each M around M^N, and M^{N+K}, which minimizes MFPE(M) locally, can be obtained. These computations correspond to the revision of the system identification along the vertical arrows in Fig. 3.2. This M^N is converging to an integer which minimizes MFPE(·) as N increases.

The algorithm of the modified Akaike's method. The algorithm of the modified Akaike's method can be summarized as follows:

Step 1 : Delete the mean from the initial N data and calculate $S^N(\ell)$, $C^N(\ell)$, $\underline{p}^N(\ell)$, and $\underline{q}^N(\ell)$, for $\ell = 0,1,\ldots,L$, with (3.11) and (3.12).

Step 2 : Find M^N that minimizes MFPE(M) and at the same time calculate $(R^N_{M^N})^{-1}$ and $\hat{A}^N_{M^N}$.

Step 3 : If K = 0, the computation is terminated. Otherwise, delete the mean from the added data and calculate $S^{N+K}(\ell)$, $C^{N+K}(\ell)$, $\underline{p}^{N+K}(\ell)$, and $\underline{q}^{N+K}(\ell)$, for $\ell = 0,1,\ldots,L$, with (3.13) to (3.15).

Step 4 : Calculate $(R^{N+K}_{M^N})^{-1}$ with (3.16).

Step 5 : Calculate $(R^{N+K}_{M^N})^{-1}$ for M around M^N with (3.18) and (3.19), and find M^{M+K} that minimizes MFPE(M) locally. At the same time, calculate $\hat{A}^{N+K}_{M^{N+K}}$ with (3.17). Set N+K → N and go back to Step 3.

In this algorithm, the number of data that have to be stored in the computer memory is only K+L, as seen from (3.13) and (3.14).

Modified Akaike's Method with Observed Random Input

When an r-dimensional, stationary Gaussian input $\underline{u}(k)$ exists with an n-dimensional output $\underline{x}(k)$, the method described above can also be applied. A random input $\underline{u}(k)$ should be uncorrelated with the noise experienced by $\underline{x}(k)$. An AR model for an augmented (n+r) dimensional vector $\underline{z}(k) = \underline{x}(k)^T$, u

$$\underline{z}(k) = \sum_{m=1}^{M} \begin{bmatrix} A_M(m) & \vdots & B_M(m) \\ \cdots & \vdots & \cdots \\ * & \vdots & * \end{bmatrix} \underline{z}(k-m) + \begin{bmatrix} \underline{e}(k) \\ \cdots \\ * \end{bmatrix} \tag{3.20}$$

can be fitted in a way similar to that in the preceding section. Instead of the MFPE of the final prediction error criterion

$$FPEC(M) = \left(1 + \frac{M(n+r) + 1}{N}\right)^n \left(1 - \frac{M(n+r) + 1}{N}\right)^{-n} \det(D_{M,n}) \tag{3.21}$$

can be used (Akaike 1971). In (3.21), $D_{M,n}$ denotes the nxn upper left part of D_M, the estimate of the covariance matrix of the residuals.

NUMERICAL EXAMPLE FROM THE YOMO RIVER

A stretch of four km of the Yomo river is divided into four reaches as shown in Fig. 3.3. The values of the parameters k_1 and k_2 for all four reaches are

$$k_{1i} = 0.53 \ (\text{days}^{-1}) , \qquad k_{2i} = 0.10 \ (\text{days}^{-1})$$

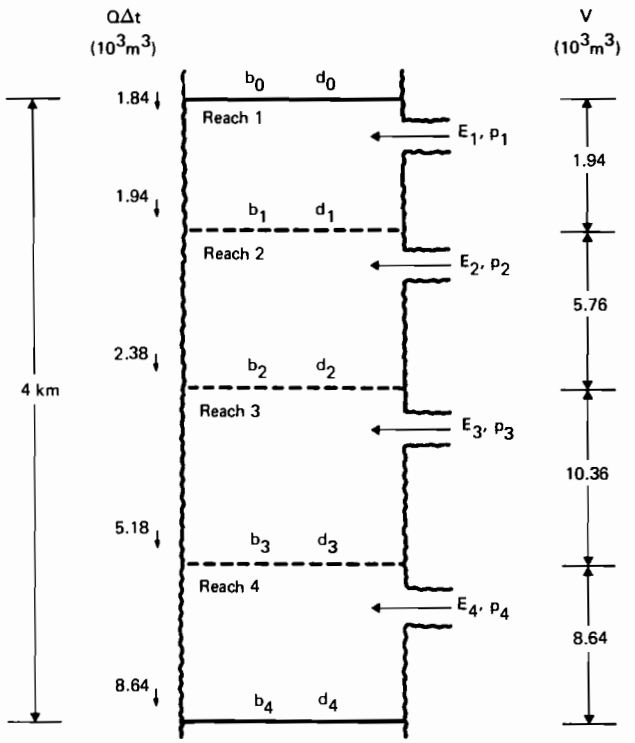


Fig. 3.3 The four-reach river basin of the Yomo river.

The values of the other parameters are listed in Table 3.1; $\Delta t = 1/8$ days. For simplicity, the unknown order and parameters of only the BOD model, (3.4a) are estimated. Since only steady state data of this river were available, realistic time series data were generated by computer simulation, and these "synthetic" data were used for the identification study. The simulation model used for obtaining time series data is

$$\underline{x}(k) = \begin{bmatrix} -0.064 & 0 & 0 & 0 \\ 0.246 & 0.436 & 0 & 0 \\ 0 & 0.046 & 0.436 & 0 \\ 0 & 0 & 0.120 & -0.064 \end{bmatrix} \underline{x}(k-1) + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.103 & 0 & 0 & 0 \\ 0 & 0.115 & 0 & 0 \\ 0 & 0 & 0.300 & 0 \end{bmatrix} \underline{x}(k-2)$$

$$+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0.062 & 0 & 0 & 0 \\ 0 & 0.046 & 0 & 0 \\ 0 & 0 & 0.120 & 0 \end{bmatrix} \underline{x}(k-3) \quad \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0.023 & 0 & 0 \\ 0 & 0 & 0.060 & 0 \end{bmatrix} \underline{x}(k-4) + \underline{u}(k) + \underline{v}(k)$$

where input $\underline{u}(k) = [0.85, 5.25, 41.1, 4.67]^T$, and $\underline{v}(k)$ is a vector of white Gaussian noise with zero mean and variance 0.1.

The identification was recursively performed for every 100 data ($K = 100$). The evolution of the estimated results for the order is shown in Fig. 3.4. The

Table 3.1. Parameter values for the study of the Yomo River.

Reach i	Steady State b_i (mg/ℓ)	Steady State d_i (mg/ℓ)	Q_i ($10^3 m^3/\Delta t$)	V_i ($10^3 m^3$)	Q_{Ei} ($10^3 m^3/\Delta t$)	E_i (mg/ℓ)	π_i^+ (mg/ℓ)	$\phi_{i-1}(0)$	$\phi_{i-1}(1)$	$\phi_{i-1}(2)$	$\phi_{i-1}(3)$
0	12.0	6.13	1.84								
1	11.4	6.13	1.94	1.94	0.11	15.3	0.02	1.00	0	0	0
2	17.6	7.79	2.38	5.76	0.43	57.8	1.24	0.60	0.25	0.15	0
3	80.0	9.50	5.18	10.36	2.81	151.0	4.32	0.20	0.50	0.20	0.10
4	49.5	9.10	8.64	8.64	3.46	11.7	3.20	0.20	0.50	0.20	0.10

$$\pi_i^+ = \frac{Q_{Ei}}{V_i} p_i + \frac{r_i}{V_i}$$

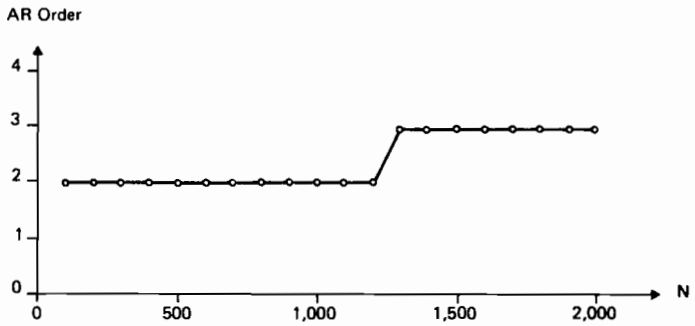


Fig. 3.4 The estimated results for the order of the BOD model in the Yomo River study with increasing data.

parameter estimates were, with 1,000 data:

$$\hat{A}_2(1) = \begin{bmatrix} -0.057 & 0.038 & 0.020 & -0.015 \\ 0.198 & 0.443 & 0.030 & -0.013 \\ 0.005 & 0.011 & 0.420 & -0.018 \\ -0.020 & 0.020 & 0.098 & -0.005 \end{bmatrix} \quad \hat{A}_2(2) = \begin{bmatrix} -0.036 & -0.011 & -0.009 & 0.046 \\ 0.138 & 0.020 & 0.007 & -0.004 \\ 0.009 & 0.165 & 0.013 & 0.022 \\ 0.018 & 0.029 & 0.378 & 0.019 \end{bmatrix}$$

and with 2,000 data:

$$\hat{A}_3(1) = \begin{bmatrix} -0.042 & -0.027 & -0.009 & -0.019 \\ 0.220 & 0.448 & 0.026 & 0.007 \\ -0.007 & 0.026 & 0.411 & -0.023 \\ -0.004 & 0.001 & 0.120 & -0.040 \end{bmatrix} \quad \hat{A}_3(2) = \begin{bmatrix} -0.041 & -0.016 & 0.009 & 0.029 \\ 0.131 & -0.029 & -0.029 & -0.018 \\ 0.001 & 0.117 & 0.023 & 0.026 \\ 0.026 & 0.026 & 0.284 & 0.019 \end{bmatrix}$$

$$\hat{A}_3(3) = \begin{bmatrix} -0.006 & -0.002 & -0.011 & -0.010 \\ 0.018 & 0.030 & 0.018 & 0.011 \\ 0.011 & 0.058 & 0.002 & -0.002 \\ -0.009 & -0.026 & 0.139 & 0.001 \end{bmatrix}$$

To test the accuracy and convergence of the identification algorithm, we evaluate

$$C(K,N) = \frac{1}{nK} \sum_{k=N-K+1}^{N} (\underline{y}(k) - \underline{\hat{y}}(k))^T (\underline{y}(k) - \underline{\hat{y}}(k))$$

The development of C(K,N) is plotted in Figure 3.5, where

$$\underline{y}(k) = \sum_{m=1}^{M} A(m)\underline{y}(k-m) , \qquad \underline{\hat{y}}(k) = \sum_{m=1}^{M} \hat{A}(m)\underline{\hat{y}}(k-m)$$

and $\hat{M}$ and $\hat{A}(m)$ are the estimates of M and A(m). For our case, n = 4, K = 100, and M = 4. C(K,N) shows the difference between the real system and the identified

Fig. 3.5 The correspondence between the real system and the BOD model in the Yomo River study with increasing data.

model. Although the estimated order is lower than that of the real system (i.e., simulation model), the similarity of the parameter values and the convergence profile of C(K,N) show that the real-time recursive algorithm of the modified Akaike's method is quite satisfactory.

CONCLUDING REMARKS

In the numerical example, we estimated only the BOD part of the model for four reaches at the same time (n = 4). However, since the river quality models in the different reaches are not affected by each other, we could have identified both the BOD and DO part of the model reach by reach; the size of each identification problem would have been small (n = 2). Furthermore, since we can analyze data recursively in the algorithm and since the time increments between data collections are sufficiently long (1/8 day), it is possible to carry out system identification for the multiple reaches using parallel processing.

REFERENCES

Akaike, H. (1970). Statistical predictor identification. Annals of the Institute of Statistical Mathematics 22: 203-217.

Akaike, H. (1971). Autoregressive model fitting for control. Annals of the Institute of Statistical Mathematics 23: 163-180.

Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. Annals of the Institute of Statistical Mathematics 26: 363-387.

Box, G.E., and G.M. Jenkins (1970). Time Series Analysis - Forecasting and Control. Holden-Day, San Francisco.

Fortmann, T.E. (1970). A matrix inversion identity. IEEE Trans. Autom. Control 15: 599.

Parzen, E. (1974). Some solutions to the time series modelling and prediction problem. Technical Report No. 5, Statistical Science Division, State University of New York at Buffalo.

Parzen, E. (1975). Multiple time series, determining the order of approximating

autoregressive schemes. Technical Report No. 23, Statistical Science Division, State University of New York at Buffalo.

Rinaldi, S., R. Soncini-Sessa, H. Stehfest, and H. Tamura (1978). Modelling and Control of River Quality. McGraw-Hill, Dusseldorf.

Streeter, H.W., and E.B. Phelps (1925). A study of the pollution and natural purification of the Ohio River. Public Health Bulletin, No. 146, U.S. Public Health Service.

Tamura, H. (1974). A discrete dynamic model with distributed transport delays and its hierarchical optimization for preserving stream quality. IEEE Trans. Syst. Man. Cybern. 4: 424-431.

4 Application of Maximum Likelihood Identification to River Flow Prediction

Tohru Katayama

Planning and control of water resource systems have recently become the subject of research of systems and control engineers. To predict river flow, one must know the behavior of the river basin. There are two different approaches for the development of dynamic models of the river basin; one is based on hydrologic theory and the other on statistical estimation theory.

This paper is concerned with an application of the maximum likelihood (ML) identification method to rainfall/runoff modeling. Both linear and multiplicative models are used. The parameter identification was carried out by using the daily rainfall/runoff data of the Karasu river in the Kanto district of Japan. The derived models are used for one-day-ahead prediction of the river flow.

THE LINEAR MODEL

In general, a discrete-time, single-input, single-output (SISO) linear system is expressed as

$$\begin{aligned} x(t+1) &= Fx(t) + Gu(t) + Lw(t) \\ y(t) &= Hx(t) + v(t) \end{aligned} \tag{4.1}$$

where $x(t)$ is an n-dimensional state vector at time t, and $u(t)$ and $y(t)$ are a scalar input and a scalar output. F, G, H, and L are parameter matrices with appropriate dimensions. The system and observation noises $w(t)$ and $v(t)$ are Gaussian white noise sequences with zero means. In Chapter 2 of this volume, Mehra shows that only the innovation representation of (4.1) is suitable for parameter identification (see also, Kailath 1968 and Mehra 1971). In other words, (4.1) is redundant with respect to noise and system parameters, so that not all the parameters in (4.1) are identifiable (Tse and Weinert 1975). In order to derive an identifiable form, we define the one-step predicted estimate $\hat{x}(t)$ of the state $x(t)$ based on the observations $y^{t-1} = \{y(1),\ldots,y(t-1)\}$ and $u^{t-1} = \{u(1),\ldots,u(t-1)\}$

This work was supported by the Ministry of Education in Japan under the special research project, "Evaluation and Control of Environmental Systems."

as

$$\hat{x}(t) = E\{x(t) | \sigma(y^{t-1}, u^{t-1})\} \tag{4.2}$$

where $\sigma(y^{t-1}, u^{t-1})$ is the σ-field generated by y^{t-1} and u^{t-1}, and $E\{\cdot|\cdot\}$ denotes the conditional expectation. By using Kalman filter theory (Bucy and Joseph 1968), we have an innovation representation of system (4.1):

$$\begin{aligned} \hat{x}(t+1) &= F\hat{x}(t) + Gu(t) + K\nu(t) \\ y(t) &= H\hat{x}(t) + \nu(t) \end{aligned} \tag{4.3}$$

where $\nu(t)$ is the innovation process representing Gaussian white noise with mean zero and variance σ^2. The innovation process $\nu(t)$ expresses the one-step prediction error for $y(t)$ based on y^{t-1} and u^{t-1}, namely,

$$\nu(t) = y(t) - \hat{y}(t) = y(t) - H\hat{x}(t) \tag{4.4}$$

The K is an n-dimensional vector that denotes the steady-state Kalman filter gain. It should be noted that in (4.3) the state vector is the one-step predicted estimate of the original state vector in (4.1) and that (4.3) is the mean value of the process represented by (4.1). When the parameters are identified, it readily follows from (4.3) that the steady state Kalman filter is given by

$$\hat{x}(t+1) = F\hat{x}(t) + Gu(t) + K[y(t) - H\hat{x}(t)] \tag{4.5}$$

Hence, if the parameters in (4.3) are known, it is not necessary to solve a Riccati matrix equation to obtain the steady-state Kalman filter gain. In the following, we consider the system (4.3) as a state-space representation of the process to be identified.

It is well known that an SISO system has canonical representations (Kalman 1963). In this paper, we use an observable canonical form, namely,

$$F = \begin{bmatrix} 0 & 0 & \cdots & a_n \\ 1 & 0 & \cdots & a_{n-1} \\ & 1 & & \\ & & \ddots & \\ 0 & & & a_1 \end{bmatrix} \quad G = \begin{bmatrix} b_n \\ b_{n-1} \\ \vdots \\ b_1 \end{bmatrix} \quad K = \begin{bmatrix} k_n \\ k_{n-1} \\ \vdots \\ k_1 \end{bmatrix} \tag{4.6}$$

$$H = [0 \ \cdots \ 0 \ 1]$$

For convenience, we define

$$\underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \qquad \underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \qquad \underline{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

where $c_i = k_i - a_i$ $(i = 1,\ldots, n)$. We see from (4.6) that the pair (H, F) is observable. However, the pairs (F, G) and (F, K) are not always controllable. Therefore, to obtain a minimal realization that contains the least possible number of parameters, the uncontrollable parts of the state vector $\hat{x}(t)$ should be eliminated. There exist several algorithms for the minimal realization (Ho and Kalman 1966, Mayne 1968, Furuta 1973).

By using (4.6), (4.3) can be transformed into an autoregressive moving average (ARMA) model:

$$y(t) = \sum_{i=1}^{n} a_i y(t-i) + \sum_{i=1}^{n} b_i u(t-i) + \nu(t) + \sum_{i=1}^{n} c_i \nu(t-i) \tag{4.7}$$

where it is assumed that the initial conditions

$$\begin{aligned} &u(0),\ u(-1),\ldots,\ u(-n+1) \\ &y(0),\ y(-1),\ldots,\ y(-n+1) \end{aligned} \tag{4.8}$$

are known. Taking the z-transform of (4.7) yields

$$A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})\nu(t) \tag{4.9}$$

where

$$A(z^{-1}) = 1 - a_1 z^{-1} - \ldots - a_n z^{-n}$$

$$B(z^{-1}) = \quad b_1 z^{-1} + \ldots + b_n z^{-n}$$

$$C(z^{-1}) = 1 + c_1 z^{-1} + \ldots + c_n z^{-n}$$

and z^{-m} is the backward shift operator defined as $z^{-m}y(t) = y(t-m)$. In the following, we take the ARMA model (4.7) as our model for identification.

For the parameter identification, it is assumed that

- All the roots of $A(z^{-1}) = 0$ lie in the unit circle of the z-plane.
- All the roots of $C(z^{-1}) = 0$ lie in the unit circle of the z-plane.

The first assumption implies that the homogeneous equation corresponding to (4.7) is asymptotically stable, and the second implies that $C(z^{-1})$ is invertible.

The problem approached by this paper is to find an optimal order n_0 of the system

and to identify the parameters $\underline{a}$, $\underline{b}$, and $\underline{c}$ based on the given input and output data. To find the optimal order of the system is much more difficult than the parameter identification for a fixed n. Appendix B is devoted to the discussion of this problem.

If the parameter identification is done, it follows from (4.5) and (4.6) that the one-step predicted estimate $\hat{y}(t)$ of the output $y(t)$ is recursively given by

$$\hat{y}(t) + \sum_{i=1}^{n} c_i \hat{y}(t-i) = \sum_{i=1}^{n} b_i u(t-i) + \sum_{i=1}^{n} k_i y(t-i) \tag{4.10}$$

with the initial conditions

$$\hat{y}(0) = \hat{y}(-1) = \ldots = \hat{y}(-n+1) = 0$$

It should be noted here that the second assumption is necessary for ensuring the asymptotic stability of the homogeneous part of (4.10).

THE MULTIPLICATIVE MODEL

Since the output observation of a river flow, say, $y(t)$, takes positive values only, we can assume the following multiplicative form

$$y(t) = \prod_{j=1}^{n} y(t-j)^{a_j} \prod_{j=1}^{n} u(t-j)^{b_j} \prod_{j=0}^{n} e^{c_j \nu(t-j)} \tag{4.11}$$

where $\underline{a}$, $\underline{b}$, and $\underline{c}$, with $c_0 = 1$, are unknown parameters to be identified, and $\nu(t)$ is a Gaussian white noise innovation process with mean zero and variance σ^2, as in the previous section. Taking the logarithm of both sides of (4.11) gives

$$\eta(t) = \sum_{j=1}^{n} a_j \eta(t-j) + \sum_{j=1}^{n} b_j \xi(t-j) + \nu(t) + \sum_{j=1}^{n} c_j \nu(t-j) \tag{4.12}$$

where

$$\eta(t) = \log y(t), \qquad \xi(t) = \log u(t)$$

Since (4.12) is the same ARMA model as (4.7), we will not repeat the necessary assumptions and related arguments in the previous section. By using the ML method in the next section, the parameters $\underline{a}$, $\underline{b}$, and $\underline{c}$ and σ^2 are identified.

Let $\hat{\eta}(t)$ be the conditional mean estimate of $\eta(t)$ based on η^{t-1} and ξ^{t-1}. Since the logarithm function gives a one-to-one, monotonic mapping, $\sigma(\eta^{t-1}, \xi^{t-1})$ is equivalent to $\sigma(y^{t-1}, u^{t-1})$, and

$$\begin{aligned} \hat{\eta}(t) &= E\{\eta(t) | \sigma(\eta^{t-1}, \xi^{t-1})\} \\ &= E\{\eta(t) | \sigma(y^{t-1}, u^{t-1})\} \end{aligned} \tag{4.13}$$

Referring to (4.10), the conditional mean estimate $\hat{\eta}(t)$ satisfies the following equation

$$\hat{\eta}(t) + \sum_{j=1}^{n} c_j \hat{\eta}(t-j) = \sum_{j=1}^{n} b_j \xi(t-j) + \sum_{j=1}^{n} k_j \eta(t-j) \tag{4.14}$$

where $k_j = c_j + a_j$ ($j = 1,\ldots, n$). The one-step predicted estimate of $y(t)$ based on y^{t-1} and u^{t-1} is given by

$$\hat{y}(t) = E\{y(t) | \sigma(y^{t-1})\} = \exp\{\hat{\eta}(t) + \sigma^2/2\} \tag{4.15}$$

and its variance is

$$\hat{V}(t) = E\{(y(t) - \hat{y}(t))^2\} = (\hat{y}(t))^2 [e^{\sigma^2} - 1], \tag{4.16}$$

as shown in Appendix A. Since we see from (4.15) that $\hat{y}(t)$ is a function of σ^2, the estimation of σ^2 is very important for a multiplicative, lognormal system.

An obvious disadvantage of using the multiplicative system (4.12) for rainfall/runoff modeling is that if $u(t) = 0$ for some t, that is, if there is no rainfall at t, then the outputs of the system (4.12), $y(s)$, are zero for all $s > t$. This cannot be true. A simple way of alleviating the difficulty is to introduce a positive constant β and to define

$$\xi(t) = \log(\beta + u(t))$$

We will discuss the choices of the constant β in a later section.

MAXIMUM LIKELIHOOD IDENTIFICATION

Given the conditional probability density function (pdf) of the output $y(t)$ conditional on $\underline{\Theta}$, we define the likelihood function of $\underline{\Theta}$ as

$$L(\underline{\Theta}) = p(\underline{y} | \underline{\Theta}) \tag{4.17}$$

where $p(\underline{y}|\underline{\Theta})$ is the conditional pdf. For a given sample value of $\underline{y}$, $L(\underline{\Theta})$ is only a function of the parameter Θ. A maximum likelihood estimate (MLE) of $\underline{\Theta}$, which is denoted by $\hat{\underline{\Theta}}_{ML}$, is the value of $\underline{\Theta}$ that maximizes $L(\underline{\Theta})$. In other words the MLE $\hat{\underline{\Theta}}_{ML}$ is determined so that the sample value of $\underline{y}$ is most likely. It is well known that the MLE has the following important statistical properties (Aström and Bohlin 1965, Anderson 1958).

- Asymptotic normality: the convergence of $\sqrt{N}(\hat{\underline{\Theta}}_{ML}(N) - \underline{\Theta})$ to a random vector with $N(0, I)$, where $\hat{\underline{\Theta}}_{ML}(N)$ is the MLE based on N observations
- Asymptotic efficiency: the covariance matrix of $\hat{\underline{\Theta}}_{ML}(N) - \underline{\Theta}$ attains the Cramér-Rao lower bounds for $N \to \infty$

- Consistency: $\hat{\underline{\theta}}_{ML}(N)$ converges in probability to $\underline{\theta}$ for $N \to \infty$.

In the following, $\underline{\theta}$ denotes the aggregation of the unknown parameters $\underline{a}$, $\underline{b}$, and $\underline{c}$. Since $\nu(t)$ in (4.7) (or in (4.12)) is Gaussian white noise with mean zero and variance σ^2, the conditional pdf of y^N given u^{N-1}, $\underline{\theta}$, σ^2, and initial conditions (IC) (4.8) becomes

$$p(y^N|u^{N-1}, \underline{\theta}, \sigma^2, IC) = \prod_{t=1}^{N} p(y(t)|y(i), u(i), i < t-1, \underline{\theta}, \sigma^2)$$

$$= \prod_{t=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2} \nu^2(t)\} \tag{4.18}$$

where $\nu(t)$ $(t = 1,\ldots, N)$ are computed from (4.7), or from (4.12) for given input and output data and system parameters. Therefore the conditional log-likelihood function is given by

$$L_N(\underline{\theta},\sigma^2) = \log p\ (y^N|u^{N-1}, \underline{\theta}, \sigma^2, IC)$$

$$= -\frac{N}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^{N} \nu^2(t) \tag{4.19}$$

Since the logarithm is monotonic increasing the MLEs $\hat{\underline{\theta}}$ and $\hat{\sigma}^2$ maximize the log-likelihood function (4.19) with respect to $\underline{\theta}$ and σ^2, respectively. Thus, from the likelihood equation

$$\frac{\partial}{\partial\sigma^2} L_N(\underline{\theta}, \sigma^2)\Big|_{\sigma^2=\hat{\sigma}^2} = 0 \tag{4.20}$$

the MLE of σ^2 becomes

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^{N} \nu^2(t) \tag{4.21}$$

Substituting (4.21) into (4.19), we see that the MLE of $\underline{\theta}$ is obtained by minimizing

$$L_N^*(\underline{\theta}) = \frac{1}{N} \sum_{t=1}^{N} \nu^2(t) \tag{4.22}$$

subject to equality constraints (4.7), or (4.12). Hence, the ML identification is reduced to a parametric optimization problem that minimizes the innovation process. Although the development is heavily dependent on the assumption that the innovation is a Gaussian white noise process, the resultant minimization problem is quite reasonable from an engineering point of view.

It should be noted from (4.7) or (4.12) that the innovation process $\nu(t)$ is linear with respect to parameters $\underline{a}$ and $\underline{b}$ for fixed input/output data. Therefore if $\underline{c} = 0$ in (4.7) or (4.12), then $L_N^*(\underline{\theta})$ is reduced to a quadratic form in $\underline{a}$ and $\underline{b}$, so that the unique minimum point can easily be obtained by solving a so-called normal equation by least squares (LS). But the LS estimates may be biased, because the correlated noise $\{\nu(t) + \Sigma_1^n c_i \nu(t-i),\ t = 1,\ldots,N\}$ was replaced by uncorrelated

noise $\{v(t), t = 1,\ldots,N\}$. For $\underline{c} \neq 0$, however, $L_N^*(\underline{\theta})$ becomes a complicated nonlinear function. Hence, for the numerical solution to the minimizing problem, we have employed Davidon's conjugate gradient method (Kowalik and Osborne 1968). For the algorithm of optimization, see the presentation by Katayama, Akimoto, and Sawaragi (1976). Appendix B discusses the order of the model, and the overall identification procedure is briefly summarized in Appendix C.

APPLICATION TO RIVER FLOW PREDICTION

In this section, the ML identification method is applied to rainfall/runoff modeling of the Karasu river in Gumma, Japan. The Karasu river with a basin area of 156 km^2 and length of 27 km, is one of the most important tributaries of the Tone River (see Fig. 4.1). Daily input/output data from 1964 to 1965 (731 days) are available. The rainfall inputs u(t) (mm/day) in the catchment area were obtained as an average of the measurements of four rain gauges shown in Fig. 4.1, and the outputs y(t) (m^3/sec), the runoff, were estimated from the daily average of the water levels at the gauging station. Tables 4.1 and 4.2 show the frequency distributions of the input and output data. The normalized correlation functions are depicted in Fig. 4.2.

Table 4.1. Frequency distribution of input data: Rainfall from January 1964 to December 1965

Range (mm/day)	Number of Days	Frequency (%)
0 - 2.5	540	73.87
2.5 - 5.0	54	7.39
5.0 - 7.5	27	3.69
7.5 - 10.0	19	2.60
10.0 - 12.5	18	2.46
12.5 - 15.0	16	2.19
15.0 - 17.5	7	0.96
17.5 - 20.0	4	0.55
20.0 - 25.0	15	2.05
25.0 - 30.0	5	0.68
30.0 - 35.0	3	0.41
35.0 - 40.0	6	0.82
40.0 - 50.0	6	0.82
50.0 - 60.0	6	0.82
60.0 - 70.0	1	0.14
70.0 -	4	0.55
Total	731	100.00

The identification was carried out by using 1964 data (366 days). The results are shown in Table 4.3, where $\hat{\sigma}^2_n$ denotes the minimum value of $L_N^*(\underline{\theta})$ defined by (4.22). Akaike's information criterion (AIC) and the final prediction error (FPE) are also computed (Akaike 1974b). Comparing the variances of estimation errors with those from some conceptual hydrologic models given in Table 4.4, we see that the parameter identification has been done very well. Numerical studies show that the "whiteness" of the innovation process is improved for larger values of n and that for $n > 3$ the autocorrelation functions are within the 95 percent confidence intervals. Thus we see that the innovation process (or residuals) is, for practical purposes,

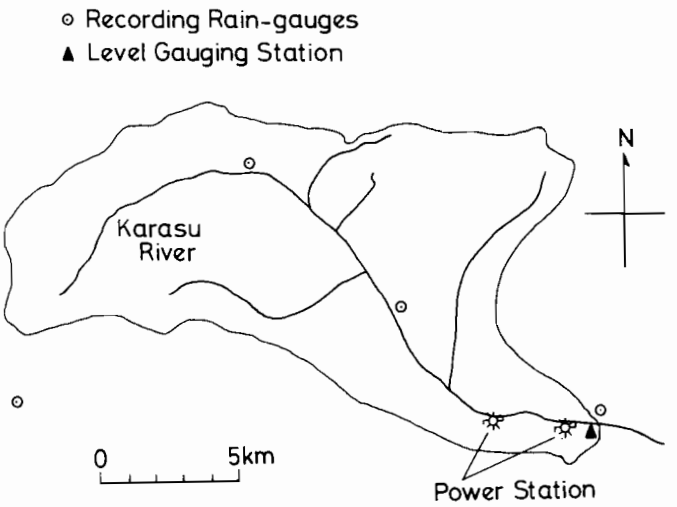


Fig. 4.1 The Karasu river catchment area.

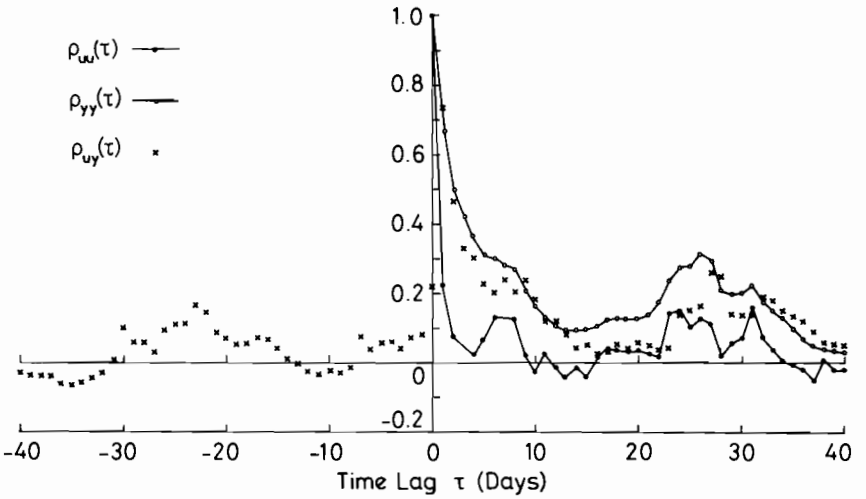


Fig. 4.2 Correlation functions of input/output data.

whitened for $n \geqslant 3$ (see Figs. 4.3 and 4.4). As shown in Fig. 4.5, we observe that the impulse responses for $n \geqslant 3$ oscillate around the curve for $n = 2$. This oscillatory nature may not be a true reflection of the actual dynamics of the river basin, but may have resulted from the fitting process of linear models to the input/output data disturbed by various nuisance variances inherent in complex hydrologic systems. The models so obtained are applied to the one-day-ahead prediction of the runoff from May 1 to September 10, 1965 (133 days); the results are also shown in Table 4.3. The best performance was obtained by the second-order model among those with $n \leqslant 10$; the mean square error (MSE) is comparable to that of Sugawara (Table 4.4), who has developed one of the best tank models (see Sugawara 1972). The minimum of AIC(3n) was obtained at $n = 9$, and a local minimum at $n = 3$, where $n_L = 10$. Based on the above numerical studies, we may conclude that low order models with $n = 2$ or 3 will be sufficient for the one-day-ahead prediction of river flow. Moreover, the parameters in these models can be renewed

Table 4.2. Frequency distribution of output data: Runoff from January 1964 to December 1965

Range (m^3/sec)	Number of Days	Frequency (%)
0 - 2.5	72	9.85
2.5 - 5.0	396	54.17
5.0 - 7.5	148	20.25
7.5 - 10.0	43	5.88
10.0 - 12.5	27	3.69
12.5 - 15.0	14	1.91
15.0 - 17.5	9	1.23
17.5 - 20.0	9	1.23
20.0 - 30.0	10	1.37
30.0 - 40.0	1	0.14
40.0 - 50.0	1	0.14
50.0 - 60.0	0	
60.0 - 70.0	0	
70.0 - 80.0	1	0.14
Total	731	100.00

Table 4.3. Results of linear model identification

n	$\hat{\sigma}^2_n$: LS	FPE	$\hat{\sigma}^2_n$: ML	AIC	Prediction
1	1.54000	1.55696	1.53464	0.444733	8.37474
2	1.38978	1.42066	1.32478	0.314313	6.80336[a]
3	1.33280	1.37759	1.27594	0.293271	7.02037
4	1.33432	1.39462	1.28021	0.313319	7.23136
5	1.28463	1.35781	1.25482	0.310094	7.05388
6	1.26286	1.34992[a]	1.24077	0.315728	7.41441
7	1.26594	1.36863	1.21708	0.313446	8.29196
8	1.26790	1.38645	1.19631	0.313320	8.00905
9	1.26348	1.39753	1.12646	0.270337[a]	9.75737
10	1.26561	1.41161	1.20975	0.358952	7.99043

[a] Minimum values.

as soon as new data are available. Figure 4.6 depicts the actual flow y(t), the estimates from the linear model with n = 3, and the estimates from Sugawara's tank model (Tone River Dams Control Office 1967).

The multiplicative models described in the third section are also applied; the results are summarized in Table 4.5, where $\beta = 10$. Numerical studies show that

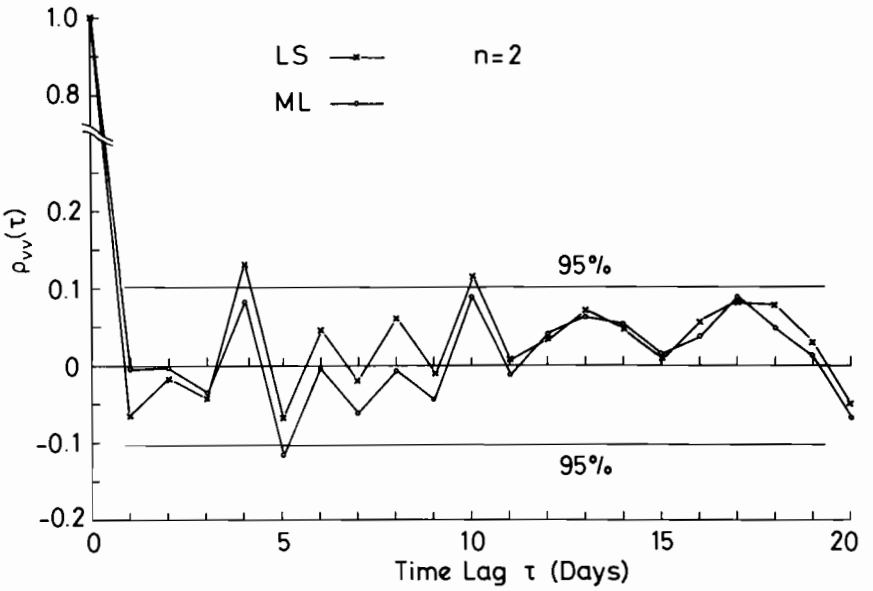


Fig. 4.3 Autocorrelation function of the innovation process (n = 2).

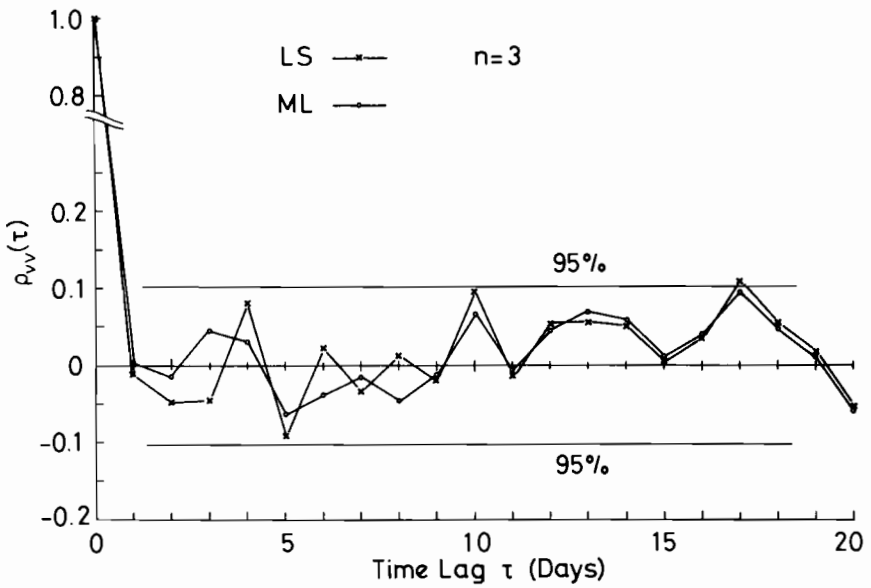


Fig. 4.4 Autocorrelation function of the innovation process (n = 3).

the larger the constant β, the better the identification. This suggests that a logarithmic transform of input data u(t) may not be relevant in this situation. In view of the identification, we see that the parameter identification is successful. Comparing Tables 4.4 and 4.5, however, we observe that prediction was very poor. Although relatively few computer studies have been done until now, we notice that the multiplicative models are sensitive to the waveform of the time series. Therefore, for the use of the multiplicative model, extensive studies remain to be done.

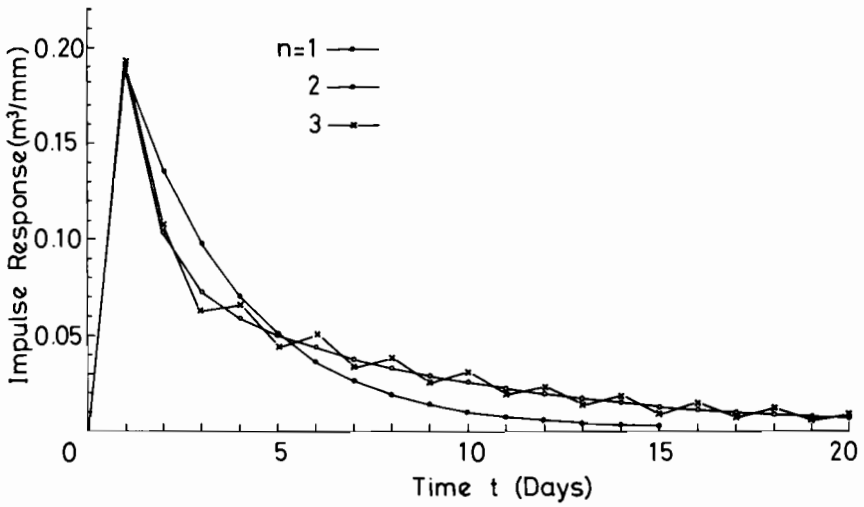


Fig. 4.5 Impulse response functions.

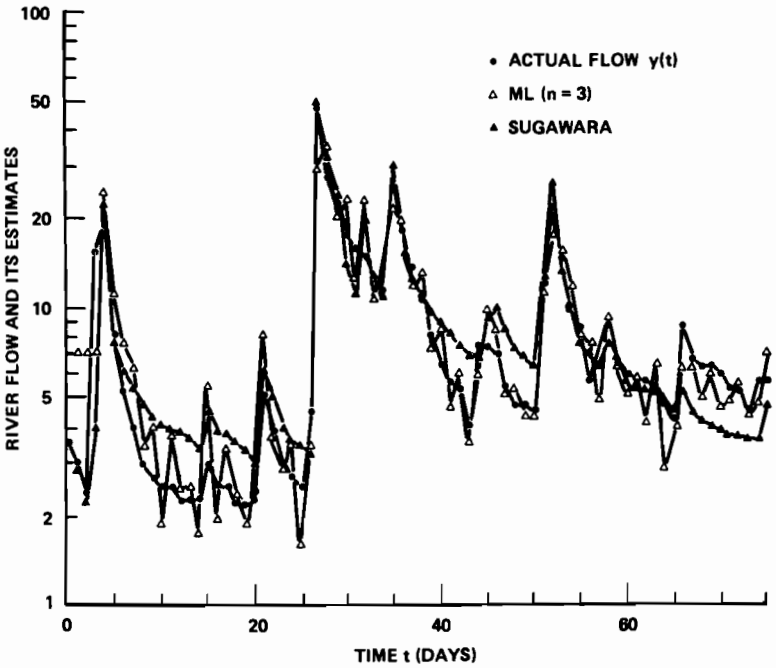


Fig. 4.6 Actual river flow and estimates with the linear and tank models.

Table 4.4. MSE performance of river flow estimation by tank models

Model	Identification[a]	Prediction[b]
Sygawara	2.41324	6.54934
Nishihara	3.29951	13.08607
Storage function	9.22446	15.16746

Source: Tone River Dams Control Office, 1967.

[a] Data from January 1 to December 31, 1964 (366 days).

[b] Data from May 1 to September 10, 1965 (133 days).

Table 4.5. Results of multiplicative model identification: MSE performance with $\beta = 10$

n	Identification[a]		Prediction[b]
	LS	ML	
1	1.57057	1.56544	19.28503
2	1.28237	1.22938	15.87144
3	1.23342	1.21886	15.75892
4	1.23361	1.23291	16.01611
5	1.19985	1.19949	16.00065
6	1.18723	1.18345	16.86354
7	1.20358	1.18203	16.84989
8	1.20258	1.18969	16.67419
9	1.19101	1.09863	19.61833
10	1.15186	1.12629	17.24032

[a] Data from January 1 to December 31, 1964 (366 days).

[b] Data from May 1 to September 10, 1965 (133 days).

ACKNOWLEDGEMENT

The author would like to express sincere thanks to Professor Y. Sawaragi for his guidance and encouragement during this study.

REFERENCES

Akaike, H. (1972). Information theory and an extension of the maximum likelihood principle. Proc. 2nd Int. Symp. Information Theory, 267-281, House of Hungarian Academy of Sciences.

Akaike, H. (1974a). Stochastic theory of minimal realization. IEEE Trans. Autom. Control 19: 667-674, December.

Akaike, H. (1974b). A new look at the statistical model identification. IEEE Trans. Autom. Control 19: 716-726, December.

Anderson, T.W. (1958). An Introduction to Multivariate Statistical Analysis. John Wiley, New York.

Aström, K.J., and T. Bohlin (1965). Numerical identification of linear dynamic systems from normal operating records. Proc. 2nd IFAC Symp. on Self-Adaptive Systems, 96-111.

Bucy, R.S., and P.D. Joseph (1968). Filtering and Stochastic Processes with Applications to Guidance. Interscience, New York.

Chow, J.C. (1972). On estimating the order of an autoregressive moving average process with uncertain observations. IEEE Trans. Autom. Control 17: 707-709, December.

Furuta, K. (1973). An application of realization theory to identification of a multivariable process. Preprints 3rd IFAC Symp. on Identification and Parameter Estimation, TS-3, 939-942, Delft, The Netherlands.

Gersch, W. (1970). Estimation of the autoregressive parameters of a mixed autoregressive moving average time series. IEEE Trans. Autom. Control 15: 583, 588, October.

Ho, B.L., and R.E. Kalman (1966). Effective construction of linear state variable models from input/output functions. Regelungstechnik 14: 545-548, December.

Johnson, C., and E.B. Stear (1974). Optimal filtering in the presence of multiplicative noise. 5th Symp. Nonlinear Estimation and its Applications, 124-134, September.

Kailath, T. (1968). An innovations approach to least square estimation, Part I: Linear filtering in additive noise. IEEE Trans. Autom. Control 13: 646-655.

Kalman, R.E. (1963). Mathematical description of linear dynamical systems. SIAM J. Control A1 (2): 152-192.

Katayama, T., K. Akimoto, and Y. Sawaragi (1976). Application of maximum likelihood method to boiler system identification.

Kowalik, J., and M.R. Osborne (1968). Methods for Unconstrained Optimization Problems. Elsevier, New York.

Mayne, D.Q. (1968). Computational procedure for the minimal realization of transfer function matrices. Proc. IEE 115: 1363-1368, September.

Mehra, R.K. (1971). On-line identification of linear dynamic systems with applications to Kalman filtering. IEEE Trans. Autom. Control 16: 21, February.

Rao, R.A., and R.L. Kashyap (1974). Stochastic modeling of river flows. IEEE Trans. Autom. Control 19: 874-881, December.

Sugawara, M. (1972). Method of Runoff Analysis, Kyoritsu Pub. Co., Tokyo.

Tone River Dams Control Office, Kanto regional construction bureau, Japan (1967). Report on the Runoff Analysis of the Upper Reaches of the Tone River, July (in Japanese).

Tse, E., and H.L. Weinert (1975). Structure determination and parameter identification for multivariable linear systems. IEEE Trans. Autom. Control 20: 603-613, October.

Unbehauen, H., and B. Göhring (1974). Tests for determining model order in parameter estimation. Automatica 10 (3): 233-244.

Woodside, C.M. (1971). Estimation of the order of linear systems. Automatica 7: 727-733, November.

APPENDIX A

A KALMAN FILTER FOR A MULTIPLICATIVE SYSTEM (Johnson and Stear 1974)

Consider a multiplicative, lognormal system described by

$$x_t = \prod_{j=1}^{n} (x_{t-j})^{f_j} \prod_{j=1}^{n} e^{g_j w_{t-j}}, \quad t = 1, 2, \ldots \tag{4A.1}$$

where x_t is a state that takes positive values only, and w_t is Gaussian white noise with mean zero and variance Q. The output observation is expressed as

$$y_t = x_t \, e^{v_t} \tag{4A.2}$$

where y_t is an output taking positive values only, and v_t is Gaussian white noise with mean zero and variance R. It is assumed that the initial conditions x_j, $j = -n+1, \ldots, -1, 0$ are lognormally distributed, that is, log x_j are normally distributed, and that the system and observation noises w_t and v_t, and the initial conditions are mutually independent.

The purpose of this section is to derive the conditional mean estimate $\hat{x}_t$ of the state x_t based on $y^{t-1} = \{y_1, \ldots, y_{t-1}\}$, namely,

$$\hat{x}_t = E\{x_t | \sigma(y^{t-1})\} \tag{4A.3}$$

where $\sigma(y^{t-1})$ is the σ-field generated by y^{t-1}.

Now we define

$$\begin{aligned} \eta(t) &= \log y_t \\ \underline{\mu}(t) &= [\log x_{t-n+1}, \ldots, \log x_{t+1}, \log x_t]^T \end{aligned} \tag{4A.4}$$

where $\eta(t)$ and $\underline{\mu}(t)$ are the transformed output and an n-dimensional state vector, and the superscript T denotes the transpose of a vector or a matrix. We also define the following matrix and vectors:

$$F = \begin{bmatrix} 0 & 0 & \cdots & f_n \\ 1 & 0 & \cdots & f_{n-1} \\ & 1 & & \\ & & \ddots & \\ 0 & & 1 & f_1 \end{bmatrix} \qquad G = \begin{bmatrix} g_n \\ g_{n-1} \\ \vdots \\ g_1 \end{bmatrix} \tag{4A.5}$$

$$H = [0 \ \ldots \ 0 \ 1]$$

Then we have

$$\underline{\mu}(t+1) = F\underline{\mu}(t) + Gw_t \tag{4A.6}$$

$$\eta(t) = H\underline{\mu}(t) + v_t \tag{4A.7}$$

This is a state-space representation of an SISO, linear discrete-time system encountered in the second section of the paper.

Let $\sigma(\eta^{t-1})$ be the σ-field generated by η^{t-1}. Since (4A.6) and (4A.7) are linear and are subjected to Gaussian white noises, the conditional pdf $p(\underline{\mu}(t)|\sigma(\eta^{t-1}))$ of the state vector $\underline{\mu}(t)$ given η^{t-1} is Gaussian. Thus, by using Kalman filter theory (Bucy and Joseph 1968), the conditional mean $\underline{\hat{\mu}}(t)$ and covariance matrix P(t) can be expressed as

$$\underline{\hat{\mu}}(t+1) = F\underline{\hat{\mu}}(t) + K(t)[\eta(t) - H\underline{\hat{\mu}}(t)] \tag{4A.8}$$

and

$$P(t+1) = F[P(t) - P(t)H^T[HP(t)H^T + R]^{-1}HP(t)]F^T + GQG^T \tag{4A.9}$$

where the optimal gain K(t) is

$$K(t) = P(t)H^T[HP(t)H^T + R]^{-1} \tag{4A.10}$$

and

$$\underline{\hat{\mu}}(t) = E\{\underline{\mu}(t)|\sigma(\eta^{t-1})\} \tag{4A.11}$$

$$P(t) = E\{[\underline{\mu}(t) - \underline{\mu}(t)][\underline{\mu}(t) - \underline{\hat{\mu}}(t)]^T\} \tag{4A.12}$$

Since the logarithm is a monotonic, one-to-one mapping, the σ-field $\sigma(\eta^{t-1})$ is identical to the σ-field $\sigma(y^{t-1})$, generated by the original output observation up to time t-1. Therefore, the conditional pdf $p(\underline{\mu}(t)|\sigma(y^{t-1}))$ of $\underline{\mu}(t)$ given y^{t-1} is equal to the conditional pdf $p(\underline{\mu}(t)|\sigma(\eta^{t-1}))$, that is,

$$p(\underline{\mu}(t)|\sigma(y^{t-1})) = p(\underline{\mu}(t)|\sigma(\eta^{t-1})) = \text{Gaussian} \tag{4A.13}$$

so that the conditional pdf $p(\underline{x}(t)|\sigma(y^{t-1}))$ of the state vector $\underline{x}(t) = [x_{t-n+1} \cdots x_{t+1}\ x_t]^T$ based on y^{t-1} is lognormally distributed. From

$$\underline{\mu}_j(t) = \log x_{t-n+j} = \log \underline{x}_j(t) \qquad (j = 1,\ldots, n) \tag{4A.14}$$

it follows from the property of lognormal distribution that

$$\begin{aligned} \underline{\hat{x}}_j(t) &= E\{\underline{x}_j(t)|\sigma(y^{t-1})\} \\ &= \exp\{\underline{\mu}_j(t) + \tfrac{1}{2} P_{jj}(t) \qquad (j = 1,\ldots, n) \end{aligned} \tag{4A.15}$$

and

$$\begin{aligned} \hat{V}_{jk}(t) &= E\{(\underline{x}_j(t) - \underline{\hat{x}}_j(t))(\underline{x}_k(t) - \underline{\hat{x}}_k(t))\} \\ &= \underline{\hat{x}}_j(t)\underline{\hat{x}}_k(t)\left[e^{P_{jk}(t)} - 1\right] \qquad (j, k = 1,\ldots, n) \end{aligned} \tag{4A.16}$$

where $P_{jk}(t)$ is the j,k-component of $P(t)$. Since $x_t = \underline{x}_n(t)$, the optimal estimate $\hat{x}_t$ and its variance $\hat{V}_t$ are given by

$$\hat{x}_t = \exp\{\underline{\hat{\mu}}_n(t) + \tfrac{1}{2} P_{nn}(t)\} \tag{4A.17}$$

and

$$\hat{V}_t = E\{(x_t - \hat{x}_t)^2\} = (\hat{x}_t)^2\left[e^{P_{nn}(t)} - 1\right] \tag{4A.18}$$

APPENDIX B

ESTIMATION OF ORDER OF ASSUMED MODEL

The problem of estimating the system order has received great attention in recent years. For deterministic systems, the order determination problem has already been solved as the minimal realization algorithms (Ho and Kalman 1966, Mayne 1968). For stochastic systems, however, the problem remains to be solved and requires extensive study and numerical evidence (Akaike 1974a). Many methods for determining the order of stochastic systems have been developed (Aström and Bohlin 1965, Gersch 1970, Woodside 1971, Chow 1972, Unbehauen and Göhring 1974). The algorithms in these papers have shown very good results for digital simulation studies. However, for real processes, such as power plants, chemical processes, and environmental studies, these methods do not necessarily yield reasonable results because real systems are subject to disturbances and uncertainties. In other words, it is very difficult to determine the system order uniquely from available input/output data, because the real data will be contaminated by measurement errors and external disturbances. Thus when applied to actual data, all such methods require, at least in part, trial and error and value judgements.

A practical approach to statistical identification with order determination is the following: (a) assign the order of the model and estimate the unknown parameters from the data; (b) evaluate appropriate error measures, response characteristics, such as impulse and step responses, and locations of poles and zeros; (c) compare the results for $n = 1, 2, \ldots, n_L$, and take the n_0 as a possible order of the model for which reasonable error measures and response characteristics are obtained. If there is information about the structure of the system, it must also be taken into account.

In this paper, we have used the following methods:

- AIC criterion The idea that the optimal order or dimension of the model is a parameter to be estimated was introduced by Akaike (1972, 1974b) who proposed an information theoretic criterion. For any maximum likelihood situation, AIC(k) is defined as

$$AIC(k) = -2\log(\text{maximum likelihood}) + 2k \qquad (4B.1)$$

where k is the number of arbitrarily adjustable parameters in the model. The order selected is the value of k for which AIC(k(is minimized. Some good results with the use of AIC are presented for both artificial and real data (Akaike 1974b).

- Test for independence of the innovation process. This method, after Mehra (1971), is used to test the assumption that the innovation process $\nu(t)$, $(t = 1,\ldots, N)$, is a white noise. The autocorrelation function

$$p_{\nu\nu}(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} \nu(t)\nu(t+\tau), \qquad \tau = 0, 1,\ldots \tag{4B.2}$$

must be zero for $\tau \neq 0$, if the model is acceptable. In fact, if $\nu(t)$ is a Gaussian white noise, then $\rho_{\nu\nu}(\tau)/\rho_{\nu\nu}(0)$ is asymptotically a normal distribution with mean zero and variance $1/N$ (Anderson 1958), so that the 95 percent confidence interval is $(-1.96/\sqrt{N})$. Thus by using (4B.2), we can apply a statistical test to evaluate the accuracy of the model.

- Data checking. A simple but effective method for testing the structure of a model is to use part of the given data as checking data. First we divide the data into training data D_1 and checking data D_2. Then the model is identified by using D_1 and checked by D_2. In general, as the order n of the model increases, the variance of the innovation process for D_1 decreases, and the accuracy of the estimates of the parameters decreases because the training data D_1 is finite. Various numerical results show that for data D_2, the variance of the prediction error decreases in relation to n for small n and then gradually increases in relation to n because of the uncertainty associated with the estimates of unknown parameters. A plausible order of the model is the value n_0 for which the variance D_2 is minimized. In a sense, AIC(k) is a theoretical estimate of the logarithm of the variance of the prediction error for the checking data.

APPENDIX C

OVERALL IDENTIFICATION ALGORITHM

We summarize the identification algorithm used in this paper.

Step 1: Divide the data into $D_1 = \{u(t), y(t), t = 1,\ldots,N_1\}$ and $D_2 = \{u(t), y(t), t = N_1+1,\ldots, N\}$.

Step 2: Set $n = 1$, and go to next step.

Step 3: Let $\underline{c} = 0$, and find the estimates $\hat{\underline{a}}$ and $\hat{\underline{b}}$.

Step 4: Using Davidon's algorithm, find the ML estimates $\hat{\underline{\theta}}_n$ and $\hat{\sigma}^2{}_n$.

Step 5: Using the ML estimates $\hat{\underline{\theta}}_n$ obtained in step 4, compute impulse and step responses, and autocorrelation function $\rho_{vv}(\tau)$.

Step 6: Compute AIC(3n) by

$$\mathrm{AIC}(3n) = \log(\hat{\sigma}^2{}_n) + 2(3n)/N_1, \qquad \dim(\underline{\theta}) = 3n \tag{4C.1}$$

where it should be noted that under the Gaussian assumption, (4B.1) is equivalent to $N_1\mathrm{AIC}(3n)$ up to a constant (Akaike 1974b).

Step 7: For checking data D_2, compute the variance of prediction error.

Step 8: If $n < n_L$, set $n \to n+1$, and go to step 3; if $n = n_L$, then go to next step.

Step 9: Based on the idea in the last section, find a possible order of the model.

5 Real-Time Parameter Estimation of a Nonlinear Catchment Model using Extended Kalman Filters

R.J. Moore and G. Weiss

Forecasts of flow at critical sites of a river network are required for optimal short-term operation of a water resource system. They provide information for use of the system to supply water, alleviate floods, dilute effluents, or generate power.

In this paper, an existing deterministic-conceptual, nonlinear catchment model is developed in a stochastic framework where its parameters are estimated and updated in real-time as new data become available. The conceptual quality of the deterministic model is maintained so as to preserve the prior knowledge of the system being modeled. The stochastic-conceptual model is presented as a system formulation that can be used with two extensions of the Kalman filter algorithm.

THE DETERMINISTIC-CONCEPTUAL MODEL

A simple four-parameter catchment response model for simulating isolated storm events has been developed and shown to yield realistic flood hydrographs for design purposes (Natural Environment Research Council 1975, Mandeville 1975). This model has been taken (with one generalization) as the basic element of our real-time flow forecasting algorithm.

The conceptual routing elements of this deterministic model are the time-invariant, nonlinear reservoir and the linear channel. The magnitude of flow in a river, $q \equiv q(t)$, is considered proportional to some power of the catchment storage, $S \equiv S(t)$, as defined by the relation

$$q = \kappa S^n \tag{5.1}$$

Combining (5.1) with the equation of continuity

$$\frac{dS}{dt} = r - q \ , \tag{5.2}$$

where $r \equiv r(t)$ is the effective rainfall, gives

$$\frac{dq}{dt} = n\kappa^{\frac{1}{n}} (r - q)\, q^{\frac{n-1}{n}}$$

Making the parameter substitutions $a = n\kappa^{\frac{1}{n}}$, $b = \frac{n-1}{n}$, we obtain

$$\frac{dq}{dt} = a\,(r - q)\, q^{b} \tag{5.3}$$

To account for the delay between rainfall and catchment response, a linear channel element is introduced. Thus, in (5.3) r and q should be replaced by q(t) and r(t - τ), where τ is a pure time delay, but for convenience the fact that rainfall is to be shifted in time with respect to flow is assumed without added notation.

A procedure is now required for conversion of measured rainfall, u ≡ u(t), into effective rainfall, r. The applied volume conversion procedure is based on the contributing area concept in which the area of catchment contributing to runoff, and consequently the runoff magnitude, is considered a function of the initial catchment wetness. One measure of catchment wetness that is used operationally is the soil moisture deficit (Grindley 1967) and is computed routinely by the United Kingdom Meteorological Office. Essentially, this is an accounting procedure in which the soil moisture storage is depleted by actual evaporation, computed as a function of a Penman estimate of potential evaporation, vegetation type and the current soil moisture deficit, and added to by rainfall. An approximate technique for obtaining half-hourly estimates of soil moisture deficit from half-hourly rainfall estimates and daily Penman evaporation estimates has been developed, assuming evaporation to be sinusoidally distributed over the hours of daylight.

Mandeville (1975) found empirically that the ratio of effective to measured rainfall, $\frac{r}{u}$, may be represented by an exponential function of the catchment soil moisture deficit with a 75-mm root constant, S_{75}, so that effective rainfall may be obtained simply from

$$r = c\, u\, e^{-d\, S_{75}}$$

where c and d are two parameters to be estimated. The complete model, representing the dynamic response of a catchment to rainfall, may now be formulated as

$$\frac{dq}{dt} = a\,(c\, u\, e^{-d\, S_{75}} - q)\, q^{b} \tag{5.4}$$

In the work of Mandeville, n in the discharge-storage relationship of (5.1) was restricted to 2 to permit an analytic solution of (5.4) and, in part, to reduce the number of parameters to be estimated. This restriction was found too restrictive in catchments where observed recessions were more concave than n = 2 could represent. We include it as a parameter but restrict the number of parameters to be estimated by the extended Kalman filter to four by estimating the pure time delay, τ, prior to estimation of a, b, c, and d. We estimate τ by obtaining the approximate system impulse response function in the manner described by Box and Jenkins (1970). Heuristic adjustment of τ about this estimate, to ascertain whether the model's predictive ability may be improved, may be made when applying the filter algorithm. It is worth noting however, that in the practical real-time situation, no measured rainfall is available for making forecasts for lead times in excess of the pure time delay. Thus, an optimal lag, in the conventional model

fitting sense where future rainfalls are assumed known, may not be optimal for real-time flow forecasting and may be somewhat larger. This observation supports our belief that the estimation of the pure time delay be considered separately from that of a, b, c, and d.

STOCHASTIC MODEL FORMULATION

We now proceed to take account of the fact that our dynamic model cannot be a perfect representation of the real situation, but is only an approximation to it. Let the flow at time t_k be q_k, and let h_k be a function satisfying

$$\frac{dh_k}{dt} = a\,(c\,u\,e^{-dS_{75}} - h_k)h_k^b \qquad , t_k < t < t_{k+1} \tag{5.5}$$

with

$$h_k(t_k) = q_k \ . \tag{5.6}$$

Then a stochastic model describing the catchment response is

$$q_{k+1} = h_k(t_{k+1}) + v_{k+1} \tag{5.7}$$

where v_{k+1} is a random error term accounting for imperfections in the model dynamics and noise entering the model. We will assume for t_k equally spaced that v_k are independent and identically distributed Gaussian terms with mean 0 and variance R.

The stochastic dynamic equation (5.7) describes a Markovian nonlinear model. When the parameters are known this model can be used to give a prediction of the flow at time t_{k+1} by

$$\hat{q}_{k+1} = h_k(t_{k+1}) \tag{5.8}$$

The practical problem is one of parameter estimation which, once overcome, will permit predictions of future flows via solution of the differential equation for $h_k(t_{k+1})$.

SYSTEM DESCRIPTION

The system formulation for this estimation problem is set up as follows. The state of the system at any time t_k is described by the four parameters a, b, c, and d, which we will denote by the state vector

$$x_k \triangleq (a,\ b,\ c,\ d)^T \tag{5.9}$$

This state vector is assumed to be normally distributed with expected value x_o, and variance P_0 at the initial time t_o.

Since the parameters are assumed to be time-invariant, the system transition is

described by

$$x_{k+1} = x_k \tag{5.10}$$

As part of the system description we assume that river flow is measured without error. This assumption has been fundamental in developing the system equations as outlined in this section and is justifiable because observations of streamflow are spatially integrated measurements of water volume; they are considerably more noise free than point-sampled measurements of rainfall volume. Rainfall measurement errors will add to errors attributed to the inadequacy of the dynamic model itself, and will be compensated for by the stochastic disturbance term. Filtered estimates of river flow, and parameter estimates, using an augmented state vector have proved unrewarding and have supported our belief that measured flow values are best considered error free.

In order to conform with conventional notation we denote the measured flow at time t_k by y_k, so

$$y_k = q_k \cdot$$

Then (5.7) yields the observation equation

$$y_{k+1} = h_k(t_{k+1}) + v_{k+1} \tag{5.11}$$

In order to stress the dependence of h_k on the value of the parameters x and on the initial condition $h_k(t_k) = y_k = q_k$, we shall replace it by $h(t, x, y_k)$. The observation equation is then

$$y_{k+1} = h(t_{k+1}, x_{k+1}, y_k) + v_{k+1} \tag{5.12}$$

with x_o assumed independent of $\{v_k\}$.

This approach of defining the state vector in terms of the model parameters and the quantity (model output) to be predicted as the observation is not new. Mayne (1963), Graupe (1972), and Szöllösi-Nagy (1975) have adopted it in parameter estimation, but only for linear models.

EXTENDED KALMAN FILTER SOLUTION

An approximate solution to this stochastic nonlinear estimation and prediction problem is provided by the extended Kalman filter algorithm. In the present context where nonlinearities occur only in the observation equation, the solution is derived easily from the Kalman filter algorithm for linear systems (Kalman 1960, Kalman and Bucy 1961) when the measurement equation has been linearized about the current estimate of x as a first order Taylor series. For further details the reader is referred to Jazwinski (1970).

We denote $\hat{x}_{k|k}$ and $\hat{x}_{k+1|k}$ as the expected value of x_k and x_{k+1} respectively, given observations up to and including t_k, with $\hat{x}_{o|o} \triangleq x_o$. Similarly $P_{k|k}$ and $P_{k+1|k}$ denote the conditional (error) covariance matrix of $(x_k - \hat{x}_{k|k})$ and $(x_{k+1} - \hat{x}_{k+1|k})$ respectively, conditional on observations up to and including t_k, and with $P_{o|o} \triangleq P_o$. We also denote $\hat{y}_{k+1|k}$ and $Y_{k+1|k}$ as the expected value of y_{k+1} and the

conditional (residual) variance of $(y_{k+1} - \hat{y}_{k+1|k})$, respectively, conditional on observations up to and including t_k.

The extended Kalman filter is an algorithm that gives approximate values of $\hat{x}_{k|k}$, $P_{k|k}$, $\hat{y}_{k+1|k}$ and $Y_{k+1|k}$. For our nonlinear continuous-discrete stochastic dynamic system described by the equations (5.10) and (5.12) the extended Kalman filter solution is as follows.

At prediction:

State
$$\hat{x}_{k+1|k} = \hat{x}_{k|k} \tag{5.13}$$

Output
$$\hat{y}_{k+1|k} = h\,(t_{k+1},\ \hat{x}_{k+1|k},\ y_k) \tag{5.14}$$

Error covariance matrix
$$P_{k+1|k} = P_{k|k} \tag{5.15}$$

At an observation:

State
$$\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}\,(y_{k+1} - \hat{y}_{k+1|k}) \tag{5.16}$$

Error covariance matrix
$$P_{k+1|k+1} = (I - K_{k+1}\,M_{k+1})\,P_{k+1|k} \tag{5.17}$$

Kalman gain
$$K_{k+1} = P_{k+1|k}\,M_{k+1}^T\,Y_{k+1|k}^{-1} \tag{5.18}$$

Residual variance
$$Y_{k+1|k} = M_{k+1}\,P_{k+1|k}\,M_{k+1}^T + R \tag{5.19}$$

The row vector, M, here of dimension four, is that of the linearized observation equation and contains the first partial derivatives of h with respect to the four parameters. Element i of M at t_{k+1} is

$$(M_{k+1})_i \triangleq \frac{\partial h(t_{k+1},\ \hat{x}_{k+1|k},\ y_k)}{\partial x_i} \tag{5.20}$$

We consider, in addition, an extension of the Kalman filter where system nonlinearities are approximated to include second-order terms. The filter has been derived and called the modified Gaussian second-order filter by Jazwinski (1970), and was also developed independently by Athans et al. (1968). It is identical to the extended Kalman filter algorithm given above, except in the addition of second order terms to the predicted output and residual variance:

Predicted output
$$\hat{y}_{k+1|k} = h\,(t_{k+1},\ \hat{x}_{k+1|k},\ y_k) + \tfrac{1}{2}\,P\partial^2 h \tag{5.21}$$

Residual variance
$$y_{k+1|k} = M_{k+1}\,P_{k+1|k}\,M_{k+1}^T + R + \tfrac{1}{2}\,\partial^2 h P^2 \partial^2 h \tag{5.22}$$

If we let element (k,j) of the (4 x 4) matrix of second partial derivatives be

$$(h_{xx})_{ij} \triangleq \frac{\partial^2 h(t_{k+1},\ \hat{x}_{k+1|k},\ y_k)}{\partial x_i \partial x_j} \tag{5.23}$$

then the two scalar quantities $P\partial^2 h$ and $\partial^2 h P^2 \partial^2 h$ may be defined as

$$P\partial^2 h \triangleq \text{tr}(P_{k+1|k}\, h_{xx}) \tag{5.24}$$

$$\partial^2 h P^2 \partial^2 h \triangleq \text{tr}(P_{k+1|k}\, h_{xx})^2 \tag{5.25}$$

CALCULATION OF FILTER COMPONENTS

The two filters we propose to employ, as an approximate solution to our parameter estimation problem, have now been defined in terms of our particular stochastic system formulation. We now must evaluate the predicted flow $h(t_{k+1}, \hat{x}_{k+1|k}, y_k)$, the vector of first partial derivatives of this function with respect to the parameters, and the matrix formed by its second partial derivatives, h_{xx}. To do this we use the conceptual rainfall-runoff model, that is;

$$\frac{\partial h}{\partial t} = f(x,h) \tag{5.26}$$

where

$$f(x,h) = a(cue^{-dS_{75}} - h)h^b \tag{5.27}$$

with initial condition

$$h(0,x) = q \tag{5.28}$$

The above set of equations are solved to obtain the value $h = h(t,x)$. This may be done numerically, using the Runge-Kutta method for example, or an analytical solution can be derived (Ding 1967) for the case of no forcing function, i.e., no rain, so that

$$h = (abt + q^{-b})^{-\frac{1}{b}} \tag{5.29}$$

To form a general expression for the row vector of first partial derivatives, M, we differentiate

$$t = \int_q^h \frac{dh_1}{f(h_1,x)}$$

to obtain the result

$$\frac{\partial h}{\partial x_i} = -f(h,x) \int_q^h \frac{\partial}{\partial x_i}\left[\frac{1}{f(h_1,x)}\right] dh_1 \tag{5.30}$$

Differentiating this result leads to the following general expression for the matrix of second partial derivatives

$$\frac{\partial^2 h}{\partial x_i \partial x_j} = \frac{\partial f}{\partial h}\frac{1}{f}\frac{\partial h}{\partial x_i}\frac{\partial h}{\partial x_j} + \frac{f_{x_i}}{f}\frac{\partial h}{\partial x_j} + \frac{f_{x_j}}{f}\frac{\partial h}{\partial x_i} - f\int_q^h \frac{\partial^2}{\partial x_i \partial x_j}\left[\frac{1}{f(h_1,x)}\right] dh_1 \tag{5.31}$$

where $f(h,x)$ is now replaced by f for brevity, and f_{x_i}, f_{x_j} denote the first partial derivatives with respect to x_i, x_j. Applying these two general expressions to our particular function $f(h,x)$ given by (5.27), we obtain the elements of M and h_{xx} as given below:

$$\frac{\partial h}{\partial a} = \frac{t}{a} f$$

$$\frac{\partial h}{\partial b} = f \int_q^h \frac{\ln h_1}{f(h_1,x)} \, dh_1$$

$$\frac{\partial h}{\partial c} = f \int_q^h \frac{aue^{-dS_{75}h_1^b}}{[f(h_1,x)]^2} \, dh_1$$

$$\frac{\partial h}{\partial d} = - S_{75} \, c \, \frac{\partial h}{\partial c}$$

$$\frac{\partial^2 h}{\partial a^2} = \frac{t^2}{a^2} \frac{\partial f}{\partial h} f$$

$$\frac{\partial^2 h}{\partial a \partial b} = \frac{t}{a} \left(f \ln h + \frac{\partial f}{\partial h} \frac{\partial h}{\partial b}\right)$$

$$\frac{\partial^2 h}{\partial a \partial c} = \frac{t}{a} \left(aue^{-dS_{75}} h^b + \frac{\partial f}{\partial h} \frac{\partial h}{\partial c}\right)$$

$$\frac{\partial^2 h}{\partial a \partial d} = - S_{75} \, c \, \frac{\partial^2 h}{\partial a \partial c}$$

$$\frac{\partial^2 h}{\partial b^2} = \frac{\partial f}{\partial h} \frac{1}{f} \left[\frac{\partial h}{\partial b}\right]^2 + 2\frac{\partial h}{\partial b} \ln h - f \int_q^h \frac{(\ln h_1)^2}{f(h_1,x)} \, dh_1$$

$$\frac{\partial^2 h}{\partial b \partial c} = \frac{\partial f}{\partial h} \frac{1}{f} \frac{\partial h}{\partial c} \frac{\partial h}{\partial b} + \frac{aue^{-dS_{75}}}{f} \cdot \frac{\partial h}{\partial b} + \ln h \frac{\partial h}{\partial c} - f \int_q^h \frac{aue^{-dS_{75}} \ln h_1}{[f(h_1,x)]^2} \, dh_1$$

$$\frac{\partial^2 h}{\partial b \partial d} = - S_{75} \, c \, \frac{\partial^2 h}{\partial b \partial c}$$

$$\frac{\partial^2 h}{\partial c^2} = \frac{\partial f}{\partial h} \cdot \frac{1}{f} \left[\frac{\partial h}{\partial c}\right]^2 + \frac{2aue^{-dS_{75}h_1^b}}{f} \frac{\partial h}{\partial c} - 2f \int_q^h \frac{(aue^{-dS_{75}h_1^b})^2}{[f(h_1,x)]^3} \, dh_1$$

$$\frac{\partial^2 h}{\partial c \partial d} = - \left[S_{75} \frac{\partial h}{\partial c} + S_{75}\, c \frac{\partial^2 h}{\partial c^2} \right]$$

$$\frac{\partial^2 h}{\partial d^2} = - S_{75}\, c \frac{\partial^2 h}{\partial c \partial d}$$

where $$\frac{\partial f}{\partial h} = - a h^b + \frac{b}{h} f$$

Note that for the special case of no forcing function, i.e., no rain, all partial derivatives are zero except for $\frac{\partial h}{\partial a}$, $\frac{\partial h}{\partial b}$, $\frac{\partial^2 h}{\partial a^2}$, $\frac{\partial^2 h}{\partial a \partial b}$ ($\frac{\partial^2 h}{\partial b \partial a}$), and $\frac{\partial^2 h}{\partial b^2}$, and the following simplifications hold

$$\frac{\partial h}{\partial b} = \frac{h}{b} \left[\left(\frac{h}{q} \right)^b \left(\ln q + \frac{1}{b} \right) - \left(\ln h + \frac{1}{b} \right) \right]$$

$$\frac{\partial^2 h}{\partial b^2} = \frac{\partial f}{\partial h} \frac{1}{f} \left[\frac{\partial h}{\partial b} \right]^2 + 2 \frac{\partial h}{\partial b} \ln q - \frac{h}{b} \left[\left(\frac{h}{q} \right)^b \left(\ln q + \frac{2}{b} + \frac{2}{b^2} \right) - \left(\ln h + \frac{2}{b} + \frac{2}{b^2} \right) \right]$$

Having first found the solution for h, the evaluation of the expressions for the first and second partial derivatives given above is a relatively straightforward task. Note also that numerical calculation of the five integrals is not required for the case of no rain.

REAL-TIME APPLICATION

In the real-time situation when new measurements of streamflow may be received every half hour, the error covariance P and consequently the Kalman gain will soon become very small. Information contained in new measurements will be ignored, and the parameter state vector x will stabilize. This may cause the filter estimates of x and predictions of y to diverge from the reality of the present. A means is required of limiting the memory of the filter so our model remains responsive to the changing dynamic system.

A convenient method is the exponentially age-weighted filter of Fagin (1964), in which observations are weighted exponentially into the past. For the extended Kalman filter described by equations (5.13) through (5.19), the following two modifications are required

Residual variance $$Y_{k+1|k} = M_{k+1} P_{k+1|k} M_{k+1}^T + e^{-(t_{k+1}-t_k)/T} R \qquad (5.32)$$

Error covariance matrix $$P_{k+1|k+1} = e^{(t_{k+1}-t_k)/T} \left[I - K_{k+1} M_{k+1} \right] P_{k+1|k} \qquad (5.33)$$

To put greater weight on more recent observations, one simply decreases the value of T; $P_{k+1|k+1}$ increases, $Y_{k+1|k}$ decreases, and consequently the Kalman gain increases. Note that by introducing a fading memory filter we have relaxed the

assumption of parametric time-invariance made in the system formulation. By choosing a suitable T, the model can adapt to system nonstationarity as a result of seasonal effects on the catchment response, for example.

VALIDATION OF FILTER

A preliminary application of the model to an isolated storm on the 18.6 km^2 River Ray catchment at Grendon Underwood has been encouraging. Using as initial parameter values those derived from a Rosenbrock optimization (1960), a 64 percent reduction in the variance of the one-step-ahead prediction error has been obtained relative to the fixed parameter model prediction. The residual errors were also unbiased. However, this comparison is hardly a fair one since the Rosenbrock derived parameters were not optimized to minimize the one-step-ahead prediction error.

REFERENCES

Athans, M., R.P. Wishner, and A. Bertolini (1968). Suboptimal state estimation for continuous-time nonlinear systems from discrete noisy measurements. 1968 Joint Automatic Control Conference, Ann Arbor, Michigan, 364-382, June.

Box, G.E.P., and G.M. Jenkins (1970). Time Series Analysis Forecasting and Control. Holden-Day, San Francisco.

Clarke, R.T. (1974). The representation of a short period of experimental catchment data by a linear stochastic difference equation. Mathematical Models in Hydrology, Proc. Warsaw Symp. July 1971, IAHS Publ. No. 100, 1, 3-15.

Ding, J.Y. (1967). Flow routine by direct integration method. Proc. Int. Hydrology Symp., Fort Collins 1: 113-120.

Fagin, S.L. (1964). Recursive linear regression theory, optimal filter theory, and error analysis of optimal systems. IEEE Intern. Conv. Rec. 12: 216-240.

Graupe, D. (1972). Identification of Systems. Van Nostrand Reinhold, New York.

Grindley, J. (1967). The estimation of soil moisture deficits. Meteorol. Mag. 96: 97-108.

Ibbitt, R.P. (1970). Systematic Parameter Fitting for Conceptual Models of Catchment Hydrology. Ph.D. thesis, Department of Civil Engineering, Imperial College, London.

Jamieson, D.G., J.C. Wilkinson, and R.P. Ibbitt (1972). Hydrologic forecasting with sequential deterministic and stochastic stages. Proc. Int. Symp. on Uncertainties in Hydrologic and Water Resource Systems, Tucson, Arizona, 1: 177-187.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Kalman, R.E. (1960). A new approach to linear filtering theory and prediction problems. Trans. ASME, Series D, J. Basic Eng. 82: 35-45.

Kalman, R.E., and R.S. Bucy (1961). New results in linear filtering and prediction theory. Trans. ASME, J. Basic Eng. 83: 95-108.

Mandeville, A.N. (1975). Nonlinear conceptual catchment modelling of isolated storm events, Ph.D. thesis. University of Lancaster

Mayne, D.Q. (1963). Optimal non-stationary estimation of the parameters of a linear system with Gaussian inputs. J. Electron. Control 14: 101-112.

Natural Environment Research Council (1975). Flood Studies Report, Vol. 1, 513-531.

Rosenbrock, H.H. (1960). An automatic method of finding the greatest or lowest value of a function. Comput. J. 3: 175-184.

Szöllösi-Nagy, A. (1975). An adaptive identification and prediction algorithm for the real-time forecasting of hydrologic time series, RM-75-22, International Institute for Applied Systems Analysis, Laxenburg, Austria.

6 Estimation and Identification Problems in Wastewater Treatment

Gustaf Olsson

Concern about environmental quality has contributed to the increasing interest in and strict regulation of wastewater treatment processes. The investment costs of wastewater collection and treatment systems are enormous, and it is therefore desirable to use the system efficiently. The operational costs have also risen rapidly as a result of the increasing cost of power, chemicals, and labor.

It should be emphasized that the wastewater treatment system should be considered as part of a larger water quality system; the sewer system is connected to the treatment plant and further to the receiving water. There is a trade-off between the operation of the collection system and the treatment plant.

This paper will present unsolved problems as well as some recent advances in modeling, identification, and estimation of treatment systems. Some developments are illustrated by case studies at a full-scale activated sludge plant.

SEWER AND PLANT INFLUENT FLOW RATE

A sewer network should be operated in such a way that flow rate and water quality changes can be predicted and the influent flow rate to the plant controlled within certain limits. Most of the disturbances in a wastewater treatment plant are related to changes in the influent flow rate, composition, or concentration. The flow rate of the feed stream can vary significantly, compared with many other aspects of process control. In a small plant with a concentrated sewer network, the ratio of peak to minimum flow rate can be as high as ten to one. In a larger plant with a distributed network of sewers, this ratio is lower, but it is still large enough to create significant flow rate disturbances in the operation. For these reasons, linear dynamic models are seldom adequate.

Flow variations can appear on vastly different time scales. A rainstorm or melting snow can create a shock load for a plant within hours. These disturbances must be controlled quite differently from daily or seasonal flow variations.

Measurements

It is not a trivial task to measure flow rate with precision and reliability. In particular, storm water flow rate measurements pose special problems. An acceptable stormwater in-sewer flowmeter must overcome many obstacles, like high transient

flows, large operating ranges, high levels of suspended solids, and large pieces of debris. Better flowmeters need to be developed (see Molvar, 1975).

Some contributions from control theory can, however, improve the situation. The possibilities of systematically using filter theory to achieve better flow rate measurements have not been fully explored. For example, flow rates are often measured by taking level measurements in a sewer tunnel. Since the variations in level may be quite small, but noisy, the resulting flow rate value is unreliable if the signals are not filtered properly.

Predictions

Flow rate prediction in a sewer network is important for two reasons:

- The flow into the sewer network must be predicted in order to make the proper routing of flows across the network.

- The flow from the sewer network must be predicted with a lead time for adequate control of the flow in the wastewater treatment plant.

The problem of sewer network modeling and control is similar to that of a river system for hydroelectric power generation. A sewer network has a large number of input flow sources and is generally distributed over a large area. The flows are collected into a few or maybe only one large trunk sewer that enters the treatment plant.

Rainstorms, melting snow, and infiltration can create major changes in the sewer network flow rate. Also, in a large system the local variations in rainfall must be taken into account. In order to predict the sewer inflow, several models have been developed during the last few years. Most of these models are extremely complex and mostly deterministic, for example Chen and Shubinski (1971), Papadakis and Preul (1973), and Offner (1973). Applications are found for the cities of Seattle (Leiser 1974) and Cleveland (Anderson and Pew 1974 and Pew et al. 1973) in the United States.

It is not self-evident that the urban runoff or stormwater models should be as complex as they are in the references cited above. The modeling of the stochastic nature of the systems has not been fully explored. Still, the structure of the deterministic part of the model has to be retained. A good example of a mixture of deterministic and stochastic elements in such a model for a river system is found in Lorent and Gevers (1976).

The prediction of rainfall or urban runoff into a sewer system has to be improved. There is an increasing interest in radar measurements to predict the direction and the intensity of rainfall (see Chapter 16). Typically, short-term predictions (1-2 hours) are desirable for sewer operation. Long term predictions are desirable for operation of a treatment plant. If the influent flow could be predicted 8-10 hours ahead, then the flows could be rerouted within the plant. In an activated sludge plant, step-feed control, discussed in a later section, could be used.

Control

Different wastewater collection systems operate differently. Most of them are not operated or controlled at all, except at some local pump control stations.

In an advanced sewer system there are several objectives of control that are mainly related to water flow:

- The storage capacity of trunk and interceptor lines within the network should

be utilized so that overflows caused by storm inflow are reduced or eliminated.

- The daily flow should be regulated for the best operation of the treatment plant.

Other objectives have to do with water quality:

- Overflow points should be selected to cause the least harm to receiving waters, when overflows cannot be avoided.

- Early warning of changes in water quality or toxic materials should be given to the plant operator.

The water quality aspects are further discussed in the next section.

In order to prepare a treatment plant for flow disturbances, different control actions should be taken depending on the magnitude of the disturbance, the length of it, and the preparation time.

If a shock load caused by a rainstorm is entering an activated sludge plant, three control actions are possible. If there is an equalization tank, it should be used even if the disturbance could be predicted several hours in advance. If there is no equalization tank (which is an expensive device), then some plants are equipped with step-feed control facilities. This control, however, demands that the disturbance is known well ahead (longer than the hold-up time of the process) so that the process flows can be redirected. The step-feed control can damp both the detrimental effects of an increasing flow rate and concentration load to the plant. If, however, only short-term predictions of the flow rate are obtained, very little can be done. The flow rate must be accepted and the increasing load to the sedimentation tanks will create severe increases in the effluent suspended solids concentrations.

Figure 6.1 gives an example of how the secondary clarifier effluent concentration of a full-scale plant will change owing to disturbances in the influent water flow rate.

PLANT INFLUENT WATER QUALITY

Wastewater is by no means a well-defined liquid. There are many contaminants, so many, in fact, whose concentrations are low, that only a few substances exist at a measurable level.

If a significant part of the influent water contains industrial effluents, then the composition and concentration can change rapidly. Toxic substances may appear without any warning. Many heavy metals or other toxic substances have concentrations so low that it is neither technically nor economically feasible to measure them continuously. Their effect is not observed until the water quality deteriorates noticeably.

It is not possible to measure all the different contaminants in the wastewater. Therefore, the effect of different contaminants on the treatment operation and on the receiving water must be evaluated.

Most of the water quality measurements are performed by analytical instruments. They can be divided into two main groups:

- In-stream sensors or electrodes
- Automated wet chemistry, in which a sample of the flow stream is taken

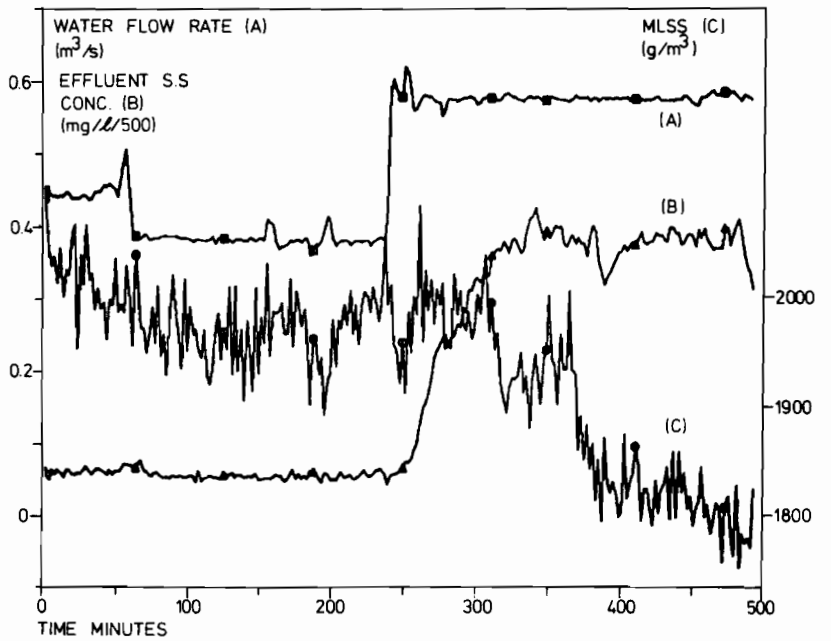


Fig. 6.1 Input and output signals of a water flow experiment in a wastewater treatment plant. The water flow rate (A) has been manipulated, causing the effluent suspended solids (SS) concentration to vary (B). The mixed liquor suspended solids (MLSS) concentration (C) is also affected by the water flow change.

It is clear that the cost of the instruments must be taken into account, and the contaminant must be worth measuring. Moreover, particularly for automated wet chemistry, each measurement will cost money. The sample is subjected to an automated laboratory procedure requiring reagents. The problem of allocating resources for measurement and control has been considered by several authors. In most cases it is assumed that there is no measurement time lag; see Cooper and Nahi (1971), Kramer and Athans (1972), Aoki and Li (1969) and Meier et al. (1967). The measurement lag usually cannot be overlooked. Even if the automatic laboratory procedure is fast, a certain amount of time will elapse before the measured value is recorded.

Olgac et al. (1976) have considered the problem of river re-aeration. Biological oxygen demand (BOD) measurements are assumed to be costly and time consuming. The question is therefore whether resources should be expended on the aerator or on BOD measurements in order to achieve the greatest improvement of the water quality.

Early detection of toxic materials is crucial. The most reliable detection method is to take measurements at each potential influent point in the sewer network; see Andrews and Olsson (1976) for more details. This is, however, not always economically feasible. Because of the dilution of the contaminants in a combined sewer tunnel, measurement is a great problem in itself, as mentioned before.

The sampling rate for measurement of toxic materials is particularly interesting. There is always a trade-off between the measurement costs and the cost of contaminants in the process and ultimately in the receiving water. Problems falling in

this category have been investigated by Brewer and Moore (1974) and Pimentel (1975). Therefore, the total performance index should take both these costs into consideration. Such an analysis has been treated extensively in theory; see, e.g., Kaplan and Haimes (1975), but hardly any of these theories have been applied to treatment systems.

The concentration changes of contaminants in the influent stream can have quite different time behaviors. There are rapid shock disturbances as well as daily, weekly, and seasonal variations. In order to achieve a suitable sampling rate, the stochastic nature of the concentration changes must be taken into consideration, and deterministic models are by no means sufficient. Again, the necessary lead time from the warning to the process depends upon the character of the contaminant as well as the magnitude of the concentration.

Different control actions in a wastewater treatment plant have quite different time responses as discussed in the next section. Depending on which control action is necessary for a certain load disturbance (expressed in terms of influent flow rate and concentration) different lead times are necessary. Dissolved oxygen (DO) control needs a lead time of just a few minutes, while step-feed control needs a lead time of several hours. To change the total mass of organisms in order to meet higher loads takes several days.

BIOLOGICAL PROCESS MODELING

Dynamic behavior of biological sewage treatment processes is still not completely understood. It is not unusual to find significant variations in process efficiency in a plant or among different plants. There are of course several reasons for this, such as insufficient knowledge of the basic phenomena, inadequate instrumentation, lack of competence of the personnel, and significant variations in the process dynamics.

During the last few years, extensive work has been done to develop dynamic models of biological treatment processes, in particular the activated sludge process. This is by far the most important process for wastewater treatment in large systems. It has several qualities that are common to all biological systems. It is also a challenging process for control engineers.

The Activated Sludge Process

In the activated sludge process, microorganisms react with organic pollutants in the wastewater and with oxygen dissolved in the water to produce more cell mass, carbon dioxide, and water. The aerobic environment is achieved by the use of diffused air, mechanical aeration, or even pure oxygen. In the last case, the tanks are covered. The effluent of the reactor flows to a sedimentation basin, where the activated sludge is separated from the liquid phase (Fig. 6.2). A portion of the concentrated sludge is recycled in order to maintain the mass of organisms in the system and a reasonable food-to-mass ratio. Part of the settled sludge is thrown away. The process effluent consists of the clarified overflow from the separator.

Because of the recycling of sludge, the aerator is inseparable from the separator. The recycling both increases the concentration of microorganisms in the aerator and maintains the organisms in a condition such that they will readily flocculate. However, recycling also results in difficulties in understanding and modeling the process since it is a feedback loop and results in a strong interaction between the aerator and the separator.

Sophisticated models of the biological part of the process exist. Reviews can be

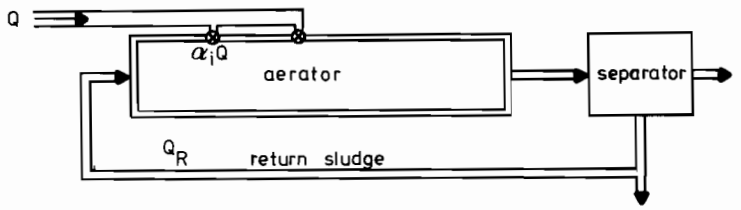


Fig. 6.2 The activated sludge process.

found in Andrews (1974), Buhr et al. (1975) and Olsson (1976). The separator models have not yet attained the same degree of sophistication and lack some fundamental properties due to the coupling forces between the settling particles or flocs in the thickener. Further details can be found in Keinath (1975), Fitch (1975), and Stenstrom (1975).

Generally, the activated sludge models are very complex and describe the dynamics in a deterministic way from basic physical laws, and essentially use mass balances for each component. The reactor design can be a mix of several subreactors in series, or a long channel aerator. Consequently, the hydraulics have to be modeled to fit the aerator design. The models still have not been verified in pilot plant or full-scale operation. At best, the available models are semi-quantitative, i.e., their results have the right order of magnitude.

In a structured model it is assumed that the substrate (or pollutant) is degraded in the following steps:

- The substrate penetrates the cell membrane by a purely physicochemical process. The removed substrate is thus "stored" in the floc phase. This procedure can take place in 15 to 30 min if the cells are in the right condition.

- The pollutants stored in the sludge (cells) are metabolized by the organisms over a period of days.

- The organisms are degraded to inert mass through endogeneous respiration and decay. Typical duration for this process is several days.

Model Structure and Complexity

A mathematical model retains only the essential elements of a system to allow prediction of the behavior of the system for a particular purpose. How complex a model is depends on its purpose. Model complexity is discussed here by using the characteristics of a model as a framework:

- State variables

- Time scale

- Nonlinearities

- Spatial distribution

- Adaptivity

State vector size. The state vector of a biological system should contain both pollutant and organism concentrations. It is quite clear that different organisms feed on different pollutants. Moreover, only certain parts of the substrate are

biodegradable. The representation of the substrate must be complex enough to reflect this fact. The living organisms must generally be represented by two or more state variables. Sometimes there are new species, such as filamentous organisms, dispersed bacteria, or ciliated Protozoa. All of them have different settling properties. Even if their synthesis rates are weighted together, in a model their relative concentrations will vary owing to the coupling between the biological reactor and the separator.

The size of the state vector may vary with time. Organisms may be killed by toxic pollutants or certain species or organisms may overtake other types of organisms. This state vector change is generally extremely difficult to predict and is intimately related to the feed water composition.

Still another problem is connected to the changes in the feed water composition. Not only the concentration but also the physiological state of the microorganisms in the process may be time varying. Their response to changes may reflect past conditions as well as the present one. In other words, the state vector of the system must include not only present organism concentrations but also variables that represent how these concentrations have been reached - the history of the organisms. This fact has been recognised for several years by biochemists, e.g., Powell (1967).

In some control schemes a detailed process description is not necessary. For DO control or mixed liquor suspended solids control, much simpler models will suffice. However, control of bulking or rising sludge in the separator is more complex.

When the settling rate is poor, the sludge is said to be bulking. One important cause of bulking is a change in sludge composition, i.e., new organisms, such as filamentous organisms, with poor settling properties may have been growing too much. This in turn depends on the feed composition, DO concentration, and other physical and chemical conditions.

Sludge may have good settling properties, but may occasionally rise to the surface after a settling period. This is sometimes caused by denitrification, in which nitrates and nitrites are converted into nitrogen gas. If enough gas is trapped in the sludge mass, the sludge rises to the surface.

Nitrification in itself is a desirable process, because it reduces the nitrogen content of the pollutant. It is made possible by the interaction of Nitrosomonas and Nitrobacter organisms. The nitrification process occurs only for long aerator retention times. If, however, nitrification is desirable, then the process must be modified so that denitrification can occur without causing rising sludge.

Time response. There is an immense difference between the smallest and largest operating times for the units of a biological system. The response times for pumps and blower systems are less than a minute, and chemical precipitation is also very fast. Oxygen dissolves in about 15 minutes. Typical flow response times for an activated sludge system are several hours. The rapid absorption of substrate by organisms takes from 15 to 30 min, but metabolism takes days. Endogeneous respiration is even slower; anaerobic decomposition of sludge takes almost a month. On top of this, there may be strong seasonal variations. The organisms are temperature sensitive, and therefore the control problem in many countries is not the same in summer and winter.

Nonlinearities. Nonlinear characteristics of the dynamics appear frequently. Because of the magnitude of disturbances, linear models are seldom adequate. In biological reactions, the rate coefficient is a nonlinear function of the substrate concentration. The separator is also a typical nonlinear process. The nonlinear behavior of the secondary clarifier is illustrated in the effluent concentration

curve B in Fig. 6.1. Here, the effluent concentration does not decrease for the negative step change of the flow rate, but increases significantly for the positive step input.

Bilinear approximation of biological reactors can easily be derived. Because of the recirculation of activated sludge the product of the return sludge flow rate and the concentration of organisms appears in the state equation. In the DO balance equation, the oxygen transfer mechanism also behaves bilinearly. It can be described by

$$k_1 u = u(c_s - c)$$

where

c = DO concentration

c_s = saturation value of DO concentration

$k_1 u$ = overall oxygen transfer coefficient

u = air flow from compressors

If it is assumed, as here, that the oxygen transfer coefficient is proportional to the air flow, then the bilinear character is obvious.

The analysis of a bilinear environmental system with respect to state observation has been made by Williamson (1975). His approach may be relevant for a laboratory process but is generally oversimplified for most full-scale plants.

Spatial distribution. The concentrations of substrate, sludge, and dissolved oxygen are in general space dependent in the biological reactor. In an activated sludge system with a dispersed plug flow reactor, the profile of substrate or dissolved oxygen varies significantly along the aerator. Fig. 6.3 is a typical example. Such a profile allows better control than conventional methods. This is further analyzed in Olsson and Andrews (1976).

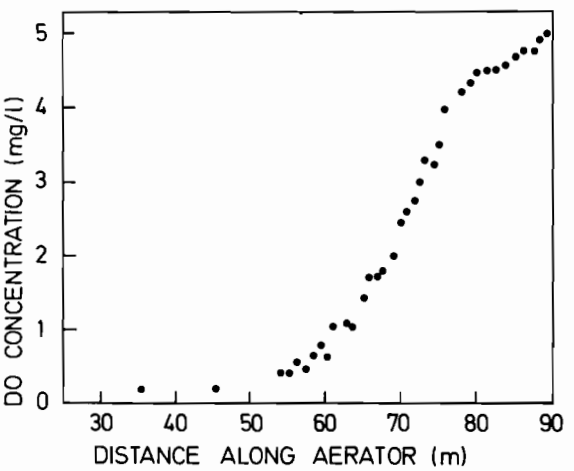


Fig. 6.3 Typical profile of dissolved oxygen in the activated sludge process of the Käppala wastewater treatment plant, Stockholm.

Because of the spatial concentration differences, the measurement locations are crucial. Analysis of sensor location has been made by Bar-Shalom and Cohen (1976) for an environmental surveillance system. For an activated sludge system, some preliminary analysis of instrument location has been made by Olsson (1975).

The concentration of sludge in the thickening part of the separator is space dependent. It is crucial to know this spatial distribution if the buffer capacity of the return sludge has to be calculated. This buffer capacity determines and limits the return flow of the sludge.

Adaptivity. A biological system is generally much more complex than other physiochemical systems because of its adaptive properties. It can adjust to certain disturbances, and its dynamic properties may change on the order of days. A large amount of industrial pollutant quickly entering a plant may kill some organisms. If the same amount of pollutant enters the plant gradually the organisms may adapt and use the substrate for their metabolism.

IDENTIFICATION OF BIOLOGICAL WASTEWATER PROCESSES

The field of identification and parameter estimation has developed rapidly over the past decade. Four specialized symposia have been arranged by the International Federation of Automatic Control (IFAC) since 1967 and numerous papers on methods and applications have been published. The review and survey paper by Aström and Eykhoff (1971) and the textbook by Eykhoff (1974) give a comprehensive description of the state of the art. The applicability of parameter identification to water quality systems has been demonstrated in several publications. Recent reviews and surveys are given in Beck (1975), Olsson (1976), and Sawaragi and Ikeda (1976). Here, some general problems of identification in biological wastewater treatment systems will be given and then some recent results of activated sludge system identification will be presented.

Special Problems in Biological Systems

If a wastewater treatment plant is compared with other physiochemical processes there are some major differences in the applicability of identification techniques. In a wastewater treatment plant:

- The influent flow rate, composition, or concentration can seldom or never be manipulated.
- Accurate measurement is difficult.
- The underlying phenomena are not well understood.

Unfortunately, because of these restrictions, often only natural variations - particularly of concentration or composition - can be observed. General and artificial disturbances are most often not noticeable. This means that normal operational records tend to exhibit just one particular mode of process behavior. There is a low signal-to-noise ratio. Routine measurement sampling rates may obscure the most important time constants for the system. It is, however, easier to perturb a wastewater treatment system than a river system, as will be shown in the next section.

Data collection is an awkward problem. It is labor intensive, requiring manual or automatic sampling followed by laboratory analysis. Because of the spatial distributions, a large amount of instrument measurement is needed to get representative data. Still - because of poor initial knowledge of the model - the data acquisition may be inadequate.

DO Identification - A case study

The dynamics of DO in an activated sludge plant have been studied in a full-scale wastewater treatment plant at Käppala in Stockholm, Sweden. The plant serves 300,000-400,000 people in the northern suburbs of Stockholm. It was completed in 1969. The average dry weather flow rate is about 1.3 m^3/sec. The details of the results are presented in Olsson and Hansson (1976a, 1976b).

Dissolved oxygen dynamics are interesting from an economic viewpoint and because of the relation between DO and the biological activity of the organisms. The DO dynamics contain four basic terms:

- The hydraulic dispersion and transportation of DO in the reactor
- The transfer mechanism by which a free gas goes into solution
- The oxygen demand of cell metabolism
- The oxygen uptake due to endogenous respiration

Three of these phenomena have been studied in the identification stage. The hydraulic properties were identified by manipulating the influent flow stream or the sludge recycling flow rate. The oxygen transfer could be studied by artificially disturbing the air flow from the compressors into the reactor tank. The oxygen demand of cell metabolism could be manipulated by dilution of the mixed liquor. Then the suspended solids concentration and consequently the organism concentration was disturbed. This manipulation, however, was difficult to make and the input variation was too limited to give accurate results.

The fact that the time scales are significantly different for the different phenomena can be systematically used in the design of the experiment. The oxygen transfer mechanism has a response time on the order of 15 min; therefore, the metabolism or endogeneous respiration rate can be neglected during an air flow experiment. On the other hand, oxygen dissolves instantaneously in a flow rate experiment.

In order to estimate the model parameters the maximum likelihood (ML) identification technique has been used. By this technique, both deterministic and stochastic parts of the models have been identified. This gives not only a meaningful physical interpretation of the parameters but also a measure of the model accuracy as well as the parameter accuracy.

In the model there are several time-varying parameters, and on-line parameter identification gives a possibility to track them. As an example, it is crucial to update the overall oxygen transfer coefficient - probably on a daily basis - if a good estimate of the biological activity or the oxygen utilization rate is to be obtained.

Due to the measurement quality and the number of sensors, the model complexity is in general quite low. As already mentioned, it is extremely difficult to verify highly structured models. Such verification requires a lot of data; more characteristics have to be measured in more places. With the available low-order models, only relatively simple control laws could be derived.

A couple of examples may illustrate the results. Figures 6.4 and 6.5 show the results of an experiment in which the influent flow rate was manipulated according to Fig. 6.4. The flow rate changes caused the mixed liquor suspended solids concentration to vary (Fig. 6.4), and these together affected the DO concentration. The identified model output (the deterministic part) is compared with the

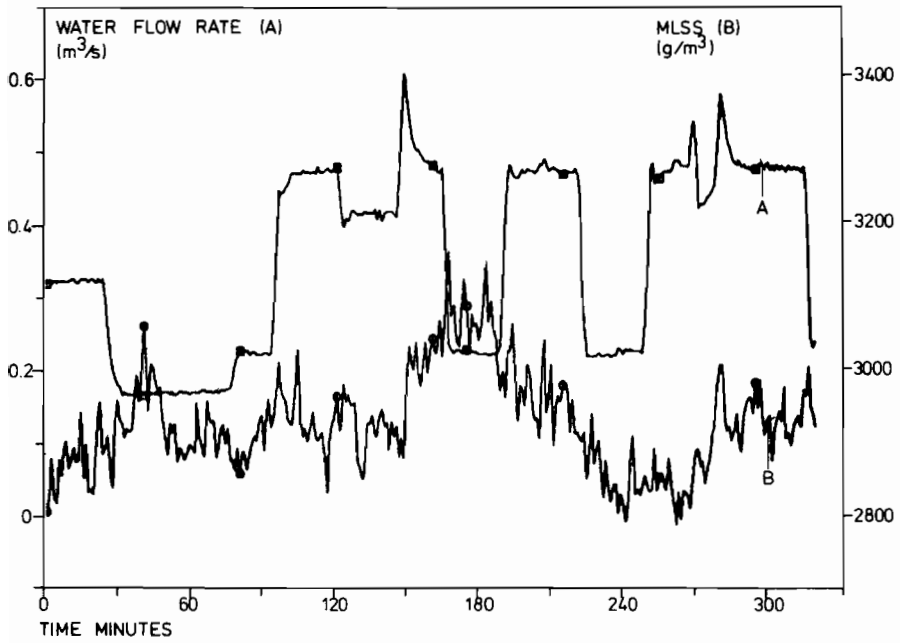


Fig. 6.4 Disturbance signals of a hydraulic experiment in the Käppala wastewater treatment plant. The water flow rate has been manipulated, causing the mixed liquor suspended solids concentration (MLSS) to vary.

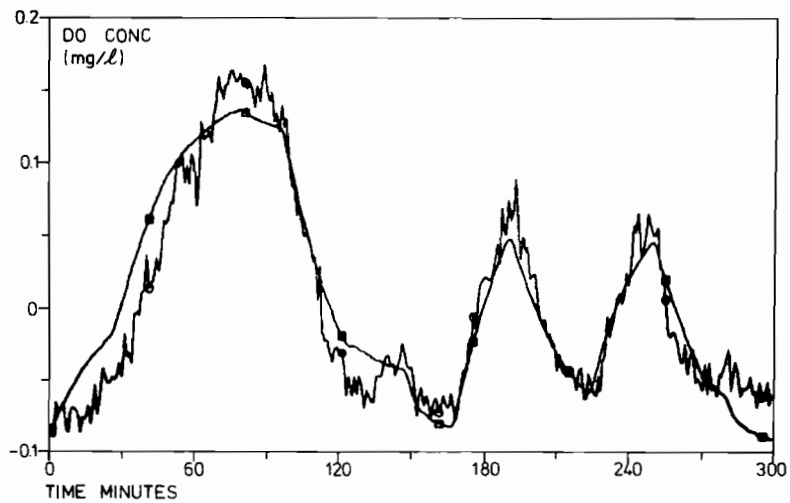


Fig. 6.5 Dissolved oxygen levels in an experiment in the Käppala wastewater treatment plant. The experimental DO output is compared with a first-order model output. The water flow rate and the MLSS concentration (see Fig. 6.4) are the inputs.

experimental DO in Fig. 6.5. The hydraulic dispersion in the aerator can be determined from a series of such experiments.

The identification of secondary thickener dynamics can illustrate another advantage of the ML identification procedures. The underflow concentration changes may be caused by several input variables, influent flow rate, mixed liquor suspended solids concentration of the aerator, as well as underflow rate. By systematic tests of models of different complexity with different inputs, cause-effect relationships can be established and verified. A systematic test of parametric models with different inputs gives more accurate and reliable results than simple correlation analysis. In particular, it was found in the Käppala experiments that the influence of the influent water flow rate on the underflow concentration was negligible. The experimental result is illustrated by Fig. 6.6.

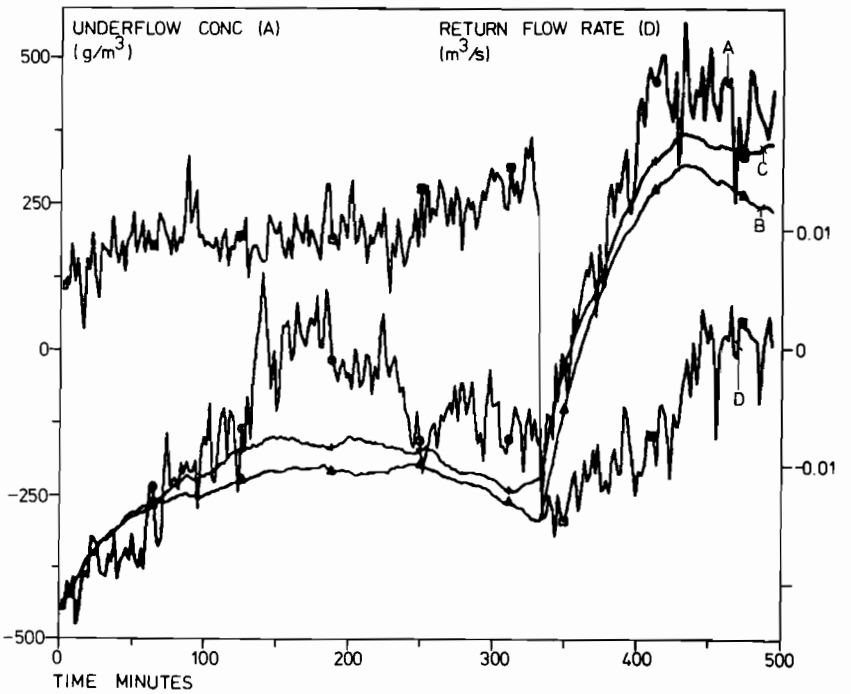


Fig. 6.6 Input and output signals of a hydraulic experiment for the thickener identification. The return sludge flow rate (D) has been manipulated. The underflow concentration level (A) is compared with two model outputs: (B) flow rate used as input; (C) flow rate and mixed liquor suspended solids concentration as input.

MEASUREMENT AND ESTIMATION

As already mentioned, the instrumentation problems in wastewater treatment can be formidable. Therefore, the possibilities for on-line parameter and state estimation should be thoroughly explored in order to overcome some of the sensor limitations. The papers by Brewer and Moore (1974) and Pimentel (1975) emphasize this point for more general environmental systems.

No sensor can replace the judgement of a good human operator and his ability for

pattern recognition. It would be very difficult to detect settling problems, rising or bulking sludge, the presence of certain organism species, or certain odors with sensors. It is probably not even desirable to replace the human operator in all his functions. Instead, the dynamic models should be constructed in such a way that inputs from the operator could be inserted easily in a continuous operation.

The organic activity in a biological wastewater treatment plant is crucial for the control and operation. It is a good example to show how estimation can complement measurements. There has been considerable work devoted to establishing measurement techniques of ATP and DNA concentrations as measures of organic activity. There is, however, considerable disagreement in the literature over their value as a measure of sludge activity. But there seems to be almost universal agreement that the specific oxygen utilization rate (SCOUR) would be an indicator of biological activity in aerobic systems. The SCOUR parameter has been studied for use in control operations by Andrews et al. (1974). It is intimately related to the growth rate of the organisms. The SCOUR parameter can be related to the DO concentration, and estimation theory allows one to get the biological activity from the relatively easy DO measurements.

Estimation procedure can also be used in a more systematic way in toxic pollutant warning systems. Toxic contaminants could be noticed in the early part of the process.

For the systems analyst, another problem should be emphasized. There have been too few sensor requirement studies in parallel with modeling, simulation, and control system design investigations. There is too little knowledge about instrumentation and the measurements really needed in order to control a plant.

CONCLUSIONS

Dynamic modeling, real-time prediction, estimation, and control are still fairly new concepts for many sanitary engineers. Interest in newly available tools is, however, rapidly increasing.

The areas of prediction and estimation are probably the most profitable areas for the application of control theory in wastewater treatment systems. Instrumentation is a serious problem and will probably be for a long time. State variable and parameter estimation can help to overcome this dilemma. The stochastic nature of sewer network flow rates and water quality variables should complement deterministic modeling.

ACKNOWLEDGEMENTS

This work was partially supported by the Swedish Board for Technical Development. The case study was performed in cooperation with Datema AB, Nynäshamn, and the Käppala wastewater treatment plant, Lidingö. In particular, Mr. O. Hansson and Mr. T. Gillblad of Datema AB, and Dr. K.I. Dahlqvist, Käppala, have contributed significantly to the results.

REFERENCES

Anderson, J.J. and K. Pew (1974). Interfacing the sewer network with the treatment plant, in Andrews et al., Eds., Progress in Water Technology Vol. 6: 532-439. Pergamon Press, Oxford.

Andrews, J.F. (1974). Dynamic models and control strategies for wastewater treat-

ment processes. Water Res. 8 (5): 261-289.

Andrews, J.F., R. Briggs, and S.H. Jenkins, Eds. (1974). Instrumentation control and automation for waste-water treatment systems. Progress in Water Technology Vol. 6. Pergamon Press, Oxford.

Andrews, J.F., H.O. Buhr, and M. Stenstrom (1974). Control systems for the reduction of effluent variability from the activated sludge process. Int. Conf. on Effluent Variability from Wastewater Treatment Processes and Its Control, IAWPR, New Orleans.

Andrews, J.F., and G. Olsson (1976). A computer-based operational strategy for the joint treatment of municipal and industrial wastewaters. Third National Conference on Complete Water Reuse, American Inst. of Chemical Engineers, Cincinnati, Ohio, June.

Aoki, M., and M.T. Li (1969). Optimal discrete-time control system with cost for observation. IEEE Trans. Autom. Control, 14: 165-175.

Aström, K.J., and P. Eykhoff (1971). System identification - A survey. Automatica 7: 123-162.

Bar-Shalom, Y., and A. Cohen (1976). Optimal resource allocation for an environmental surveillance system. IEEE Trans. Syst. Man Cybern. 6: 391-400.

Beck, M.B. (1975). Identification and parameter estimation of biological process models, in G.C. Vansteenkiste, Ed., System Simulation in Water Resources. North-Holland, Amsterdam.

Brewer, J.W., and S.F. Moore (1974). Monitoring: An environmental state estimation problem. ASME J. Dyn. Syst. Meas. Control: 363-365.

Buhr, H.O., J.F. Anderson, and T.M. Keinath, Eds. (1975). Research Needs for Automation of Wastewater Treatment Systems. Proc. U.S. Environmental Protection Agency/Clemson University Workshop, Clemson, South Carolina.

Chen, C.W., and R.P. Shubinski (1971). Computer simulation of urban storm water runoff. Proc. ASCE, J. Hydr. Div. 97, HY2: 289-301.

Cooper, C.A., and N.E. Nahi (1971). An optimal stochastic control problem with observation cost. IEEE Trans. Autom. Control 16: 185-189.

Eykhoff, P. (1974). System Identification. Wiley, New York.

Fitch, B. (1975). Current theory and thickener design. Dorr-Oliver Inc., Technical reprint No. 2027.

Kaplan, M.A., and Y. Haimes (1975). Multiobjective analysis in water resource planning and management, a framework for policy. Proc. IEEE Conf. on Decision and Control, Houston, Texas, pp. 145-147.

Keinath, T.M. (1975). Automation of physical and chemical processes, in H.O. Buhr et al. Eds., Research Needs for Automation of Wastewater Treatment Systems, pp. 58-63.

Kramer, L.C., and M. Athans (1972). On simultaneous choice of dynamic control and measurement strategies for stochastic systems. Proc. JACC, Stanford, Calif., pp. 176-182.

Leiser, C.P. (1974). Computer Management of a Combined Sewer System, Environ. Protection Techn. Series, EPA-670/2-74-022. U.S. Environmental Protection Agency, Cincinnati.

Lorent, B., and M. Gevers (1976). Identification of a rainfall-runoff process, application to the Semois River, Belgium. Proc. 4th IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, USSR, September.

Meier, L., J. Peschon, and R.W. Dressler (1967). Optimal control of measurement subsystems. IEEE Trans. Autom. Control 12: 528-536.

Molvar, A.E. (1975). Field evaluation of the effectiveness of automation, in H.O. Buhr et al. Eds., Research Needs for Automation of Wastewater Treatment Systems, pp. 118-127.

Offner, E.F. (1973). Computer simulation of storm water runoff. Proc. ASCE, J. Hydr. Div. 99, HY12: pp. 2185-2194.

Olgac, N.M., C.A. Cooper, and R.W. Longman (1976). Optimal allocation of measurement and control resources with application to river depollution. IEEE Trans. Syst. Man Cybern 6: 377-384.

Olsson, G. (1975). Activated Sludge Dynamics: Static Analysis, Dept. of Civil Engineering, Univ. of Houston, Houston, Texas.

Olsson, G. (1976). State of the art in sewage treatment plant control. Engineering Foundation Conference on Chemical Process Control, Pacific Grove, California, January 18-23.

Olsson, G., and J.F. Andrews (1976). Analysis of dissolved oxygen spatial properties in activated sludge systems. Dept. of Automatic Control, Lund Inst. of Technology, Lund, Sweden, and Dept. of Civil Engineering, Univ. of Houston, Houston, Texas.

Olsson, G., and O. Hansson (1976a). Modeling and identification of an activated sludge process. Proc. 5th IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, USSR, September, pp. 134-146.

Olsson, G., and O. Hansson (1976b). Stochastic modeling and computer control of a full scale wastewater treatment plant. Symp. on Systems and Models in Air and Water Pollution, The Institute of Measurement and Control, London, September.

Papadakis, C.N., and H.C. Preul (1973). Testing of methods for determination of urban run-off. Proc. ASCE, J. Hydr. Div. 99, HY9: 1319-1335.

Pew, K.A., R.L. Callery, A. Brandsetter, and J.J. Anderson (1973). Data acquisition and combined sewer controls in Cleveland. J. Water Pollut. Control Fed. 45: 2276-2289.

Pimentel, K.D. (1975). Toward a Mathematical Theory of Environmental Monitoring: The Infrequent Sampling Problem. Ph.D. Thesis, Univ. of California, Davis (Lawrence Livermore Lab. Rep. UCRL-51837), June 1975.

Powell, E.O. (1967). The growth rate of microorganisms as a function of substrate concentration, in E.O. Powell et al., Eds., Proc. 3rd Int. Symp. in Microbial Physiology and Continuous Culture, pp. 34-55. Her Majesty's Stationery Office, London.

Sawaragi, Y., and S. Ikeda (1976). Identification methods in environmental pollution problems. A survey of the researches in Japan. Proc. 4th IFAC Symp. on Identification and System Parameter Estimation, Tbilisi, USSR, pp. 169-189.

Stenstrom, M. (1975). A Dynamic Model and Computer Compatible Control Strategies for Wastewater Treatment Plants. Ph.D. Thesis, Dept. of Environmental Systems Engineering, Clemson Univ., Clemson, South Carolina.

Williamson, D. (1975). Modelling and observation of microbial growth processes. IFAC World Congress, Paper 55.2, Boston, Massachusetts.

7 Application of Kalman Filtering to Groundwater Basin Modeling and Prediction

Dennis B. McLaughlin

Statistical estimation and filtering techniques have received much attention from water resource engineers. Interest has been especially keen among surface water hydrologists, undoubtedly in part because Kalman filtering concepts provide an intuitively appealing way to relate time series analysis to deterministic modeling. The spatially distributed nature of environmental systems typically leads to large numbers of state variables regardless of the discretization scheme used. This high dimensionality imposes severe computational burdens, particularly when the system of interest is nonlinear. Another problem with environmental filtering applications is the limited availability of information about the statistics of important noise processes (measurement and process noise). The various adaptive techniques that have been developed to deal with this problem usually increase the already large computational burden, and are not always effective. Finally, measurements of environmental processes are often very limited. It seems to be a general rule, in fact, that in environmental applications there is never enough data of the right kind.

Even with all these difficulties, it is becoming apparent that statistical estimation techniques can make important contributions to environmental forecasting if the traditional approach to filtering is appropriately modified. The modifications required are reviewed in a general way in this paper. Most of the comments and conclusions presented can be applied as well to hydrology, water quality managements, oceanography, and meteorology.

This paper focuses on some of the implementation problems of state estimation techniques in hydrology - high dimensionality, uncertain statistics, and limited data. These problems will be discussed using the results of a case study of a groundwater forecasting problem. Some possible solutions to these problems are summarized and a few recommendations for future research are suggested.

The research described in this paper was supported in part by funds provided by the United States Department of Interior as authorized under the Water Resources Research Act of 1964, Public Law 88-379, as amended.

THE GROUNDWATER BASIN FILTERING PROBLEM

The Groundwater Flow Model

A wide variety of groundwater flow models have been proposed in the literature and applied to problems. It is convenient to consider a relatively simple model of unsteady saturated flow in a nonuniform, isotropic medium (the terms used here are defined by Greenkorn and Kessler 1969). If the aquifer being modeled is confined, but its depth is large compared with variations in the water surface elevation, the following equations may be used to describe the flow (Eagleson 1970):

Continuity
$$-\nabla \cdot q = S \frac{\partial h}{\partial t} + \sum_s Q_s \, \delta(x-x_s) \tag{7.1}$$

Momentum (Darcy's law)
$$q = K \nabla h \tag{7.2}$$

where

h = hydraulic head (meters),

q = apparent fluid velocity vector (meters per second),

t = time (seconds),

S = specific storage coefficient dependent on the compressibility and porosity of the medium and the compressibility and specific gravity of the water (liters per meter),

K = soil permeability or conductivity coefficient (meters per second),

$Q_s \delta(x-x_s)$ = known source flow of magnitude Q_s at vector position x_s.

Since the medium is nonuniform, both the specific storage and permeability coefficient depend on position.

The continuity and momentum equations may be combined to give the usual partial differential equation for hydraulic head:

$$S \frac{\partial h}{\partial t} = \nabla \cdot (K \nabla h) + \sum_s Q_s \, \delta(x-x_s) \tag{7.3}$$

The equation may be adapted to one-, two-, or three-dimensional flow, provided that all boundary conditions and spatial averaging operations are properly accounted for.

Equation (7.3) must be made discrete in both time and space in order to be solved in most practical applications. Many different techniques have been proposed and, to some extent, the choice of a technique depends on the nature of the problem being considered. Some common spatial procedures (link-node, finite difference, and finite element) are illustrated for two dimensions in Fig. 7.1. All of these procedures transform (7.3) into a set of ordinary differential equations whose state variables are the hydraulic heads at discrete node points. The resulting equation set can be written in matrix form as:

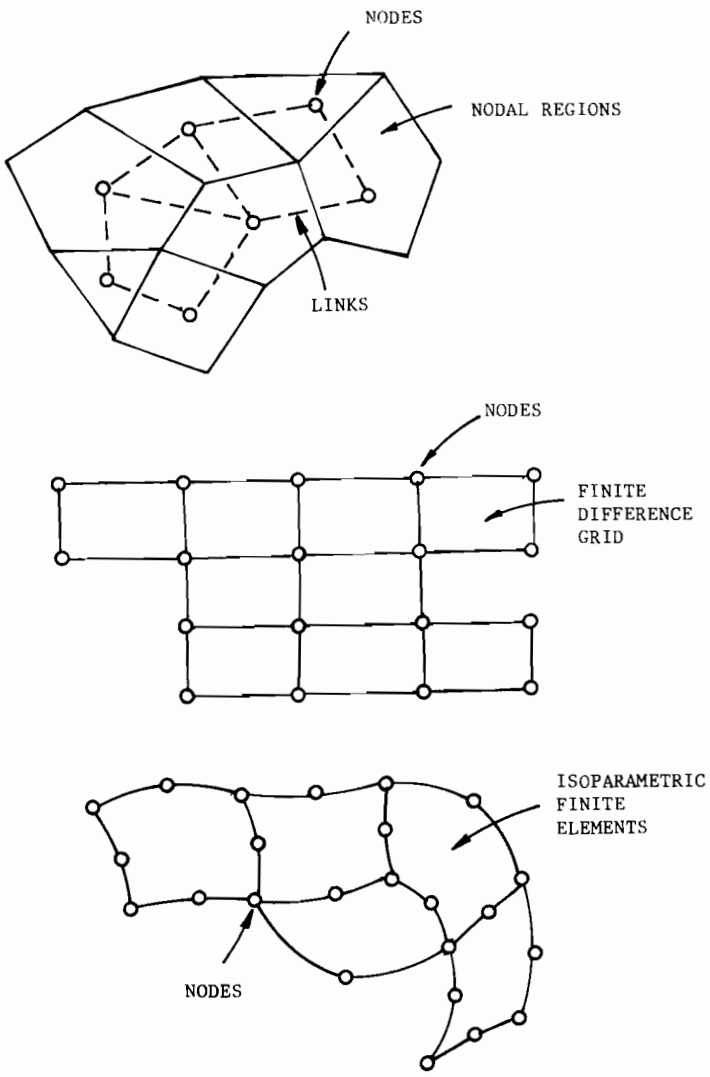


Fig. 7.1 Common spatial techniques for making equations discrete.

$$A(S)\frac{dy}{dt} + B(K)y + Cu = 0 \qquad (7.4)$$

where

y = vector of simulated hydraulic heads at N_y node points,

u = vector of known source terms at N_u node points ($N_u \leq N_y$),

$A(S)$ = $N_y \times N_y$ "mass" matrix,

$B(K)$ = $N_y \times N_y$ "stiffness" matrix,

$C = N_y \times N_u$ input matrix,

S = vector of specific storage coefficients at N_s node points ($N_s \leq N_y$), and

K = vector of permeability coefficients at N_k node points ($N_k \leq N_y$).

Note that since the model coefficients and inputs vary with location they must also be made discrete, although not necessarily as finely discrete as the hydraulic head. In general, different levels of detail will be required for S, K, and u, as reflected in the differing numbers of discrete nodal values used (N_s, N_k, and N_u). The structures of the matrices A(S), B(K), and C depend on the properties of the procedure selected.

Equation (7.4) may be solved by the usual time integration techniques (explicit, implicit, or mixed explicit-implicit schemes, again depending on the particular problem being considered). The resulting vector difference equation may be written in familiar transition matrix format:

$$y_k = \phi(S, K)\, y_{k-1} + r_{k-1} \tag{7.5}$$

where

$\phi(S,K) = N_y \times N_y$ state transition matrix related to A(S) and B(K), and

r_{k-1} = an input vector related to A(S), B(K), C and values of u(t) for $t \leq t_k$.

The subscripts k-1 and k refer to discrete times. As might be expected, both $\phi(S,K)$ and r_{k-1} depend on the properties of the time integration procedure selected.

The groundwater simulation model given in (7.5) has some important properties that are worth noting for future reference:

- The model is linear in the state variables and source terms, although the parameters enter nonlinearly. Since the model parameters are constant, the state transition matrix need only be computed once.

- The number of state variables included depends on the number of node points used when the model is made spatially discrete. This implies that multi-dimensional models possessing even a modest degree of spatial detail will have very large state vectors (hundreds of nodes for typical groundwater applications).

- The number of parameters incorporated in the model depends on the technique used to make the parameters discrete. This number may become large if nonuniformities in the medium are modeled in much detail. Obviously, the modeler may have trouble specifying large numbers of soil parameters that are difficult to measure or estimate.

- The transition matrix in (7.5) tends to be diagonally dominant because the hydraulic head at any node is affected most by conditions in the immediate vicinity of that node. Proper consideration of the diagonal nature of the system equations can provide some guidelines for model simplification and decomposition.

Formulation of the Optimal Linear Filter

Many investigators, including Freeze (1975) and McLaughlin (1975), have observed that groundwater model predictions are affected by a number of error sources and uncertainties. These may be conveniently grouped into three broad categories:

- Model simplification and idealizations (for example, assumptions regarding isotropy or the validity of Darcy's law),
- Input and parameter uncertainties, and
- Computational errors caused by making the model spatially and temporally discrete.

Statistical filtering provides a systematic means for minimizing the detrimental effects of these error sources when hydrologic measurements are available. As discussed in the Introduction to Part One, the Kalman filter is a particularly useful recursive optimal linear filtering algorithm that combines information from a deterministic model, such as (7.5) with uncertain measurements. The general form of this algorithm is described by:

$$\hat{y}_{k|k} = \hat{y}_{k|j} + \epsilon_k \tag{7.6}$$

where

$\hat{y}_{k|k}$ = vector of updated hydraulic head estimates at time k,

$\hat{y}_{k|j}$ = vector of predicted hydraulic head estimates at time k, based on measurements obtained through time j ($j < k$),

ϵ_k = measurement residual or correction at time k.

The algorithm requires that we specify procedures for computing (1) the predicted head estimate, and (2) the measurement correction. These are briefly discussed below.

The deterministic model of (7.5) shows how head estimates should be predicted when model uncertainties are ignored. In the presence of uncertainty, this model needs to be modified to account for random effects. Unfortunately, a rigorous statistical analysis of all error sources is difficult, particularly given the limited amount of accurate experimental information usually available for this purpose. Instead, simple statistical error models must be postulated and the error magnitudes adjusted to roughly compensate for unmodeled sources of uncertainty. For the groundwater example, the following additive error model is convenient:

$$y_k = \phi(S,K)y_{k-1} + r_{k-1} + w_{k-1}, \tag{7.7}$$

where

w_{k-1} = random additive process noise.

We assume that this model is accurate and that the mean and covariance of w_{k-1} are correctly known to be q_{k-1} and Q_{k-1}. Note that, in general, w_{k-1} is a nonstationary, correlated random process.

Statistical estimation theory shows that, if (7.7) applies, the best way to predict $\hat{y}_{k|j}$ from an earlier estimate $\hat{y}_{j|j}$ is to use the following equation:

$$\hat{y}_{k|j} = [\phi(S,K)]^{(k-j)} \hat{y}_{j|j} + \sum_{\ell-k}^{k-1} \phi(S,K)^{(k-\ell-1)} (r_\ell + q_\ell) \qquad (7.8)$$

Here the effect of the process noise has been simply accounted for by the mean term q_ℓ. We shall see that the process noise also affects the measurement correction term through the optimal gain matrix.

Before considering the measurement correction term, we must postulate a model of the measurement process. Again, the effect of uncertainty is most conveniently represented by an additive error term:

$$z_k = H_k y_k + v_k \qquad (7.9)$$

where

z_k = vector of measured hydraulic heads at N_z observation points ($N_z \leq N_y$),

v_k = random additive measurement noise, and

H_k = N_z x N_y measurement matrix that specifies which heads are measured at time k. Normally, H_k contains only ones and zeros.

We assume that this model is accurate and that the mean and covariance of v_k are correctly known to be m_k and M_k. It is common, but not necessary, to assume that the noise processes w_k and v_k are each uncorrelated in time as well as uncorrelated with one another. Correlated noise may be accounted for if the state vector is appropriately augmented (Gelb 1974).

Statistical estimation theory shows that the measurement correction provided by the optimal linear filter takes the following form when (7.9) applies:

$$\hat{y}_{k|k} = \hat{y}_{k|j} + G_k [z_k - H_k \hat{y}_{k|j} - m_k] \qquad (7.10)$$

where

G_k = N_y x N_z "gain" matrix that weights the measurement residuals computed at time k.

The optimal Kalman gain matrix is computed from a statistical algorithm which is summarised in the Introduction to Part One. Since the algorithm is well documented, the details of its derivation are not considered here.

The basic principles of the optimal filter may be easily understood by considering the Kalman gain computation for certain extreme cases. If all the nodes in the groundwater network are measured ($N_z=N_y$), H_k will be equal to the identity matrix and the equation for G_k will reduce to:

$$G_k = P_k(-) [P_k(-) + M_k]^{-1} \qquad (7.11)$$

Here $P_k(-)$ is the covariance of the error associated with the model's prediction $\hat{y}_{k|j}$ and M_k is the covariance of the measurement error. When the model is much

better than the measurements, $P_k(-)$ will approach zero and G_k will vanish. In this case the measurement correction term in (7.10) is ignored and the optimal estimate is:

$$\hat{y}_{k|k} = \hat{y}_{k|j} \qquad \text{(good model, poor measurements)} \tag{7.12}$$

On the other hand, when the measurements are much better than the model, M_k will approach zero and G_k will approach the identity matrix. Inspection of (7.10) shows that, in this case, the optimal estimate is:

$$\hat{y}_{k|k} = z_k - m_k \qquad \text{(poor model, good measurements)} \tag{7.13}$$

For situations between the two extremes considered above, the Kalman gain matrix will take on intermediate values that depend on the relative magnitudes of model and measurement certainties.

Application of Optimal Linear Filtering Algorithm

The example presented here describes experiments with a 63-state Kalman filter designed to predict hydraulic head fluctuations in the San Jacinto groundwater basin of southern California (McLaughlin 1975). This 121-km^2 basin was subdivided, for simulation purposes, into 63 model elements to represent geologic or hydrologic nonuniformities (see Fig. 7.2). The scheme used to make the model spatially discrete

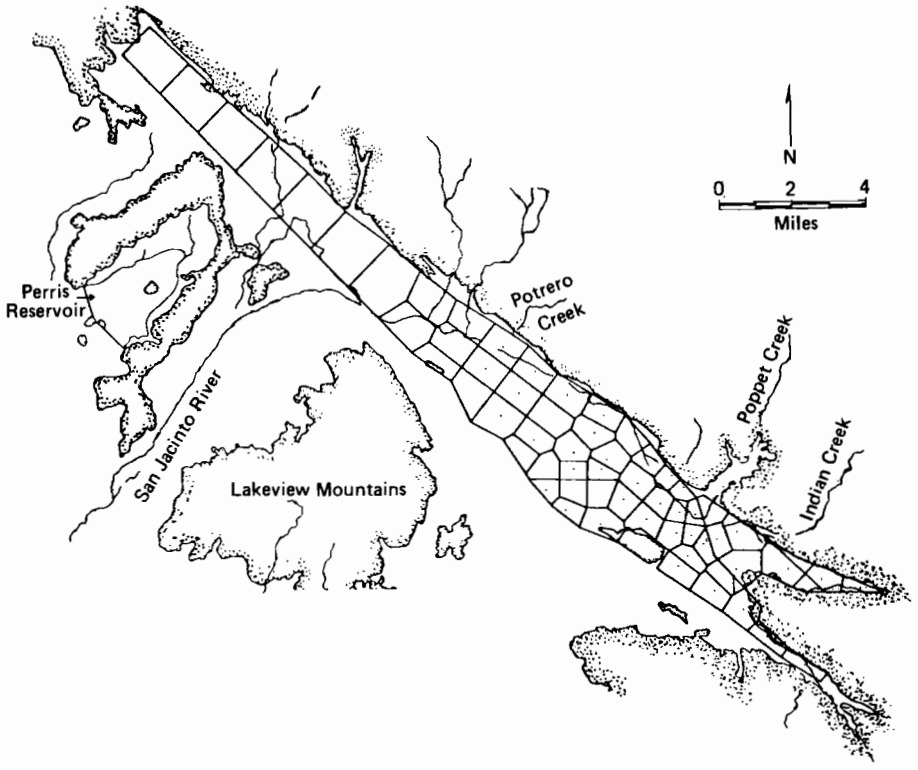


Fig. 7.2 San Jacinto groundwater basin nodal network.

was based on the link-node approach originally proposed by Tyson and Weber (1964). This approach solves a one-dimensional version of (7.5) along lines (links) connecting the centroids of the elements defined in Fig. 7.2. Although the link-node scheme was developed for two-dimensional applications, it is really based on one-dimensional flow assumptions. This simplification reduces the number of nodes (states) in the groundwater model, as compared with truly two-dimensional finite difference or finite element techniques.

Average water table elevation (hydraulic head) measurements were available throughout the San Jacinto basin for only 3 years out of the 10-year period being studied, although pumping and recharge figures were available on an annual basis. While this may appear to be a rather poor data base, it is not atypical of groundwater basins in the western United States. A review of available data indicated that the reliability of the water table elevation measurements (which were inferred from contour maps based on scattered well observations) varied somewhat, depending on time and location. Similar variations were associated with the pumping and recharge figures.

In the absence of detailed experimental data, rough reliability estimates had to be used to define the diagonal elements of the process and measurement noise covariance matrices. Off-diagonal terms were set equal to zero since no information on spatial correlation was available. The first year's measurements were used to initialize the filter algorithm and the initial estimation error covariance was set equal to the measurement noise covariance for that year.

Water table elevation estimates for two typical nodes in the San Jacinto basin are shown in Fig. 7.3. Note the tendency of the measurements to "pull" the filtered estimates toward the observed values. The unfiltered curve is the prediction obtained from the model alone, with all measurements ignored. Although the filter appears to be working as desired, an examination of estimation error standard deviations (obtained from the diagonal elements of the two estimation error covariance matrices) shows that the filtered predictions are really not significantly better than those of the model. The errors plotted in Fig. 7.4 decrease at the measurement times but soon return to the steady-state levels set by the model process noise. Clearly, measurements are not being taken often enough to provide any permanent reduction in estimation error. The implications of this result, together with a number of other implementation issues, are considered in the next section.

APPLICATION PROBLEMS AND POSSIBLE SOLUTIONS

A number of significant computational problems can arise when statistical filtering techniques are applied to environmental systems. The groundwater basin example discussed in the previous section provides a good opportunity to review some of these problems as well as some of the simplifying assumptions typically used in optimal filtering applications. Our ultimate goal is to identify the kinds of research needed to properly deal with environmental filtering problems and to suggest some possible suboptimal solutions. For convenience, the main discussion is divided into three parts corresponding to three major problem areas - high dimensionality, uncertain statistics, and limited measurements. A short review of suboptimal filtering and stochastic approximation is presented as the end of the section.

High Dimensionality

The fundamental difficulty with environmental filtering applications is the high dimensionality that arises whenever the underlying system model is spatially distributed, i.e., represented by a partial differential equation dependent on both

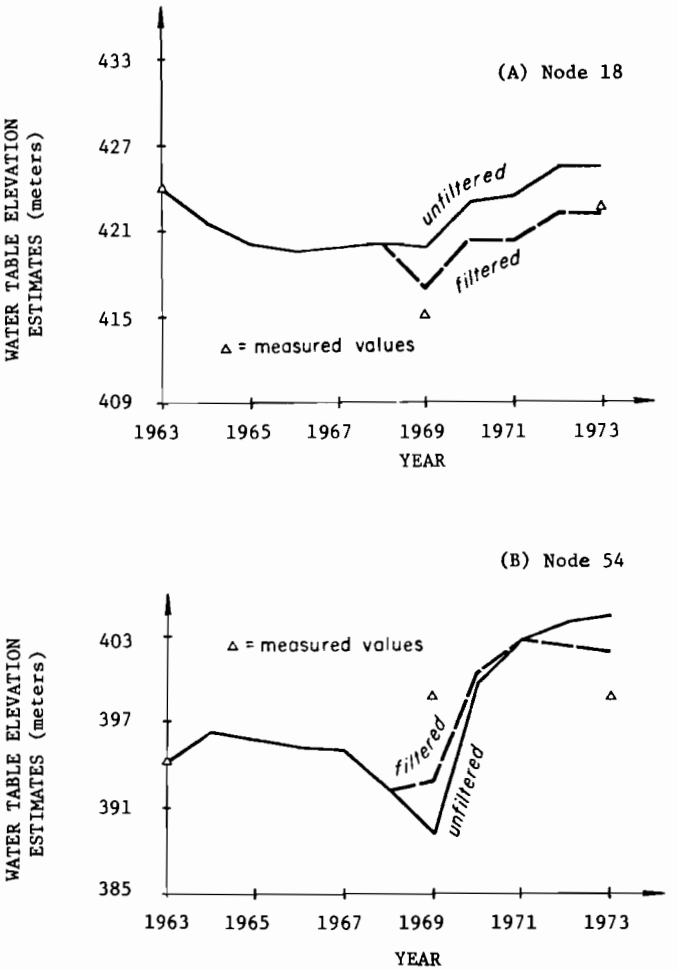


Fig. 7.3 Water table elevation estimates.

time and space. It is true that some important environmental models (in hydrology, for example) avoid this difficulty by using an aggregated input-output approach. Unit hydrograph analyses such as that of Hino (1973) are typical examples. (This approach is also discussed in Chapter 8.) But in many other applications, such as the groundwater problem considered in this paper, the system model must be based on fundamental equations of fluid flow that are spatially distributed.

The partial differential equations used to describe spatially distributed systems must be made discrete in order to be solved in all but the most specialized applications. Techniques such as those illustrated in Fig. 7.1 have been designed to provide accurate solutions under conditions of highly variable inputs, coefficients, and boundary conditions. Unfortunately, the most accurate finite difference and finite element techniques require large numbers of discrete nodes (i.e., state variables) in practical multidimensional applications. Isoparametric finite element models could easily require 200 states for a realistic two-dimensional saturated flow groundwater analysis. Incorporation of unsaturated flow or three-dimensional effects could increase this number significantly. Link-node discreteness does not require as many states (see the San Jacinto example of the previous

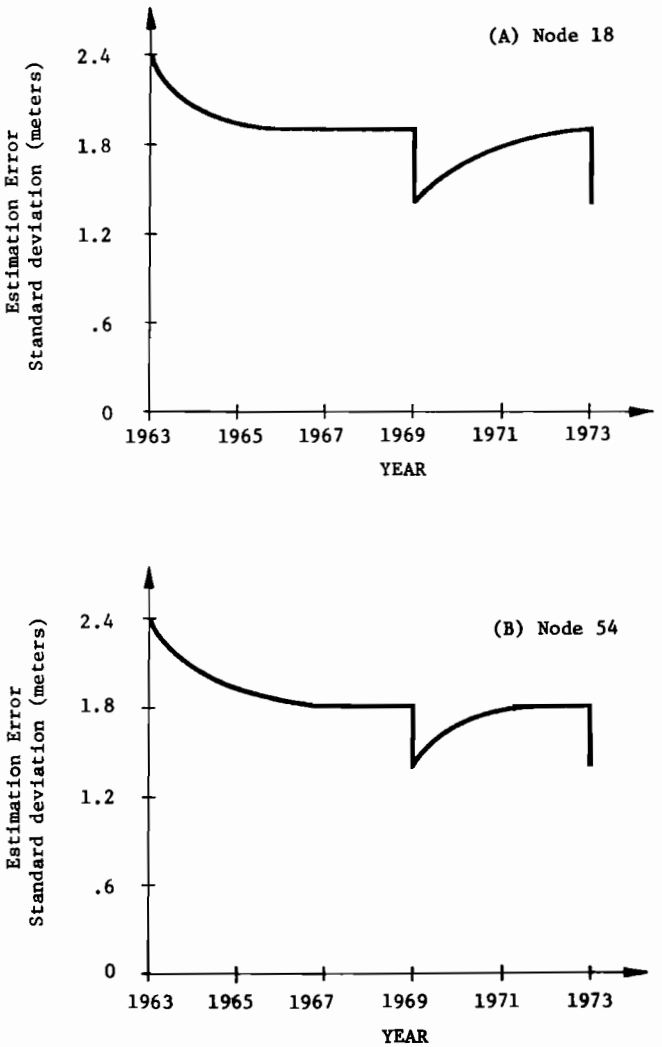


Fig. 7.4 Standard deviations of filtered water table elevation error.

section), but it is limited to applications where the predominant direction of flow is known.

Kalman filters having hundreds of state variables pose rather serious computational problems both in terms of storage and processing time. The covariance matrices $P_k(-)$, $P_k(+)$, and Q each require $N_y + |N_y(N_y-1)|/2$ storage locations (since they are symmetric), a number that becomes very large when N_y exceeds one hundred or two hundred. Processing time requirements are also significant since the matrix multiplications and inversions included in the filtering algorithm proceed slowly when the matrices are very large. This becomes particularly troublesome in non-linear problems that must be solved iteratively (Jazwinski 1970). Peripheral devices used to deal with the matrix storage problem also increase processing time when they are accessed frequently.

Obviously, the optimal filtering algorithm must be reformulated if it is to be successfully applied to multidimensional water resource problems. Several possible

applications have been suggested in the literature. Generally, these fall into the following categories:

- Spatial discreteness schemes that rely on special properties of the system
- Formulations of the recursive filter equations
- Multilevel decomposition of the filtering problem

A potentially useful spatial discreteness technique based on a separation-of-variables approach was applied by Pimentel (1975) to a Kalman filtering problem. Pimentel's application was to a diffusion model similar to the linear groundwater model described in (7.3). The solution to the equation may be written as:

$$y(x,t) = \sum_{n=1}^{\infty} a_n(t)\, b_n(x) \tag{7.14}$$

where $b_n(x)$ are orthogonal eigenfunctions defined over the region of interest and $a_n(t)$ are coefficients dependent only on time. It may be shown that once the eigenfunctions are specified, the $a_n(t)$ may be written in the following form:

$$\frac{da_n(t)}{dt} = \lambda_n a_n(t) + f_n(t) \quad n = 1,2,\ldots \tag{7.15}$$

where λ_n is a constant eigenvalue coefficient and $f_n(t)$ is a modal input related to the source terms of (7.3). Equation (7.15) may be thought of as an infinite-dimensional discrete groundwater model. If, however, the series of (7.14) is truncated at some point n=N, the model becomes finite with N states.

The usefulness of the orthogonal function discreteness scheme depends on the value of N needed to obtain a good representation of the underlying continuous groundwater flow model. Obviously, this depends in turn on the application as well as on the choice of eigenfunctions. In some cases, when a few dominant exponential or oscillatory modes can be identified, the orthogonal representation could provide a significant reduction in the number of states required in the filter. General assessments of its utility are, however, difficult to make.

Other investigators, including Desalu (1974), have concentrated on formulations of the recursive filtering algorithm. Desalu's approach is based on propagation of an incremental covariance matrix which has lower rank than the usual covariance matrix $P_k(+)$. The low-rank incremental covariance is written as the product of a series of factor matrices that are each smaller than $P_k(+)$. Since the Kalman gain is computed directly from the factor matrices, the filtering algorithm is not needed and storage requirements can be substantially reduced. Although the alternative algorithm described by Desalu avoids the need to compute inverses or to store $N_y \times N_y$ matrices, it is rather complex, and a lot of multiplications are required. Since the degree of improvement depends on the number of inputs (N_u) and number of measurements (N_z) included in the filtering problems, the usefulness of the algorithm depends on the application. Nevertheless, the basic concepts merit further investigation.

Another possible solution to the problem of filter dimensionality is decomposition of the system into a number of smaller subsystems that are treated more or less independently. Since computational time and storage requirements increase with the square or cube of the number of states, several small subsystems are much less demanding than one large system. Both intuition based on physical concepts and the form of the state transition matrix suggest that matrix equations such as (7.5) can be partitioned without a significant loss in accuracy, as long as all the

components are properly related when they are reassembled. System decomposition and multilevel optimization is an active area of research, and significant results appear regularly (Ho and Mitter 1976). Applications in the water resources field have been limited (see McLaughlin 1975 , for example) although it is likely that significant progress will be made.

Uncertain Statistics

It became apparent in the San Jacinto groundwater filtering example that the model coefficients and statistical parameters required by the optimal filter cannot be accurately specified in many environmental applications. Since the standard Kalman filter formulation assumes that all parameters are perfectly known, its covariance predictions are likely to be overly optimistic. In some cases, parameter uncertainties can even lead to divergence and instability (Gelb 1974). For these reasons, it is important to consider how uncertainties in the filter inputs can be accounted for.

In the groundwater application, two types of input uncertainty are of particular interest - uncertainty in the hydrologic parameters (S and K) and uncertainty in the noise covariances (Q_k and M_k). Each of these has been extensively examined in the literature and a variety of solutions are available.

The problem of estimating uncertain hydrologic parameters from field measurements is the well-known "inverse problem" of groundwater modeling. A number of deterministic solutions to this problem have been proposed and applied [see, for example, McLaughlin (1975), Yeh (1975), and Distefano and Rath (1975)]. Unfortunately, deterministic techniques make the unrealistic assumption that the measurements used for parameter estimation are noise-free. In addition, these techniques usually rely on iterative optimization algorithms that can be time consuming for large problems. Statistical parameter estimation techniques that account for noisy measurements are also available. They include the extended Kalman filter, which treats the parameters as uncertain states (Gelb 1974), and maximum likelihood algorithms (Schweppe 1973). Such algorithms invariably increase the computational burden of the filtering algorithm, either by increasing the number of states or by adding more computational operations. In general, on-line hydrologic parameter estimation appears to be feasible only when the filtering problem can be reduced to a manageable size (possibly through decomposition).

The problem of estimating uncertain noise covariances may be approached in a number of ways. Mehra (1970) discusses the problem in some detail and proposes an adaptive solution procedure based on an examination of the measurement residual. This procedure has proved to be useful in some cases, but unreliable in others, particularly when the system dimensionality is large (Gelb 1974). As might be expected, adaptive covariance estimation increases the computational requirements of the optimal filtering algorithm.

Limited Measurements

The problem of limited measurements is well illustrated by the San Jacinto example in which the updated measurement error gradually returns to its old level after an initial improvement. It is evident from Fig. 7.4 that more frequent measurements (annual or quarterly) would have provided a significant *permanent* improvement in accuracy. Unfortunately, expansion of field measurement programs is often expensive or logistically difficult, particularly when more sampling sites (i.e., wells) are required. This makes efficient monitoring design and data collection especially important in environmental applications.

A number of studies concerned with efficient design of monitoring networks and sampling schedules have been conducted and references are given in the literature

review in the Introduction to Part One. We now have a better understanding of the trade-offs involved in the monitoring problem, but not of the setting up of an "optimal" sampling network. Several significant problems must be resolved before optimal network design becomes a reality. These include proper specification of all the costs involved and definition of important logistic constraints such as site accessibility. The optimization process itself needs further investigation since, in many applications, it involves solution of a difficult integer programming problem.

Obviously, improvements can be made in monitoring strategies even when the optimal strategy is not known. Analysis of the predicted filter covariance can be used to compare the effectiveness of monitoring alternatives that are proposed on the basis of good engineering judgement. Such comparative analyses will probably prove to be very useful in applications of statistical filters.

Suboptimal Filtering and Stochastic Approximation

The basic purpose of statistical filtering is to improve the predictions of an uncertain model by using field measurements that may be equally uncertain. This is to be done with a filtering algorithm having the general form of (7.6). In the case of linear filtering, the measurement correction term ϵ_k is given by the following equation:

$$\epsilon_k = G_k[z_k - H_k \hat{y}_{k|j} - m_k] \tag{7.16}$$

For the *optimal* filter, G_k is computed from the rather troublesome recursive Kalman filter algorithm. For suboptimal filters, G_k is computed some other way - perhaps from a simpler, more efficient algorithm. As pointed out in the discussion of optimal filtering, it is desirable for G_k to vanish when the model is much better than the measurements and to approach the identity matrix when the measurements are much better than the model (this applies when $N_y = N_z$). Gains that exhibit this behavior appear to offer some advantages, even if they are not optimal.

We must be very careful in specifying suboptimal gains, however, since suboptimal filters can easily diverge - i.e., the true estimation error can grow without limit. Gelb (1974) presents a useful discussion of the problems of divergence and suboptimal design that emphasizes caution. Of course, it should be evident by now that even the so-called optimal filter of the previous section is actually suboptimal since its gains are computed from an erroneous model that ignores parameter uncertainties. Various techniques, such as finite memory filtering (Jazwinski 1970) or the "epsilon technique" (Schmidt 1970) have been developed to deal with the potential divergence of suboptimal filters, although none of these can be guaranteed to work in all cases.

The concept of computing G_k suboptimally is very attractive for environmental applications, even with the potential divergence problems mentioned above, because simple suboptimal computational algorithms could solve most of the filter implementation difficulties discussed in this paper. What is needed is a general theory that can be used to specify stability conditions for a particular suboptimal filter in a particular application. Although such a theory does not yet appear to be available, there are some close parallels in the literature on stochastic approximation (Robbins and Monroe 1951, Ho 1962, Albert and Gardner 1967).

Like Kalman filtering algorithms, stochastic approximation algorithms take the form of (7.6) above. The correction term is, however, not selected to minimize the estimation error at each time, but rather, to ensure *mean square convergence* (Robbins and Monroe 1951). Several investigators, including Albert and Gardner (1967) have derived convergence criteria for the gains of algorithms written in the

form of (7.16) for the special case of constant states (i.e., $\phi(K,S)$ equal to the identity matrix). Extension of these criteria to the time-varying case requires that we settle for "boundedness" rather than convergence. Even then, design criteria have only been derived for certain special cases (Gelb 1974, Deyst and Price 1968).

We believe that the problem of finding stable and efficient algorithms for computing suboptimal filter gains should be given a high priority in statistical estimation research. Once this problem is solved, widespread application of filtering to environmental prediction will be much more likely.

REFERENCES

Albert, A.E., and L.A. Gardner, Jr. (1967). Stochastic Approximation and Nonlinear Regression. MIT Press, Cambridge, Massachusetts.

Bras, R.L., and I. Rodriguez-Iturbe (1975). Rainfall-runoff as spatial stochastic processes; data collection and synthesis. Report No. 196, Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Deguida, R.N., J.J. Connor, and B.R. Pearce (1976). Application of estimation theory to design of sampling programs for verification of coastal dispersion predictions. Presented at the International Conference on Finite Elements in Water Resources, Princeton, New Jersey.

Desalu, A.A. (1974). Dynamic air quality estimation in a stochastic dispersive atmosphere. MIT Electronic Systems Laboratory publication ELS-R-542, Cambridge, Massachusetts.

Deyst, J.J., Jr., and C.F. Price (1968). Conditions for asymptotic stability of the discrete minimum variance linear estimator. IEEE Trans. Autom. Control (6).

Distefano, N., and A. Rath (1975). An identification approach to subsurface hydrological systems. Water Resour. Res. 11(6).

Eagleson, P.S. (1970). Dynamic Hydrology. McGraw-Hill, New York, pp. 261-282.

Eykhoff, P., Ed. (1973). Identification and System Parameter Estimation. American Elsevier, New York.

Freeze, A. (1975). A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media. Water Resour. Res. 11(5).

Gelb, A. (1974). Applied Optimal Estimation. MIT Press, Cambridge, Massachusetts.

Greenkorn, R.A., and D.P. Kessler (1969). Dispersion in heterogeneous nonuniform anisotropic media. Ind. Eng. Chem. 61(9).

Hino, M. (1970). Runoff forecasts by linear predictive filter. Proc. ASCE, J. Hyd. Div. Hy 3.

Hino, M. (1973). On-line prediction of a hydrologic system. Presented at the XV Congress of IAHR, Istanbul, Sept.

Ho, Y.C. (1962). On the stochastic approximation method and optimal filtering theory. J. Math. Anal. Appl. 6.

Ho, Y.C., and S. Mitter, Eds. (1976). Directions in Decentralized Control, Many-

Person Optimization, and Large Scale Systems. Plenum, New York.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Lettenmaier, D.P. (1975). Design of monitoring systems for detection of trends in stream quality. Tech. Rep. No. 39, Charles W. Harris Hydraulics Laboratory, University of Washington, Seattle, Washington.

McLaughlin, D. (1975). Investigation of alternative procedures for estimating groundwater basin parameters. Report prepared for the Office of Water Research and Technology, U.S. Department of Interior, Washington, D.C.

Mehra, R.K. (1970). On the identification of variances and adaptive Kalman filtering. IEEE Trans. Autom. Control 15(2).

Moore, S.F. (1973). Estimation theory applications to design of water quality monitoring systems. Proc. ASCE, J. Hyd. Div. Hy. 5.

Pimentel, K. (1975). Toward a mathematical theory of environmental monitoring: The infrequent sampling problem. Lawrence Livermore Laboratory publication UCRL-51837. University of California, Livermore, California.

Robbins, H., and S. Monroe (1951). A stochastic approximation method. Ann. Math. Stat. 22(1).

Sage, A.P., and G.W. Husa (1969). Adaptive filtering with unknown prior statistics. Proc. 1969 Joint Autom. Control Conf.

Schmidt, S.F. (1970). Computational techniques in Kalman filtering. Theory and applications of Kalman filtering. Advisory group for aerospace research and development (AD 704-306), AGARDograph No. 139.

Schweppe, F.C. (1973). Uncertain Dynamic Systems. Prentice-Hall, Englewood Cliffs.

Szöllösi-Nagy, A. (1975). An adaptive identification and prediction algorithm for the real-time forecasting of hydrologic time series. RM-75-22, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Takeuchi, K. (1976). Application of the Kalman filter to cyclone forecasting. RM-76-9. International Institute for Applied Systems Analysis, Laxenburg, Austria.

Tyson, H., and E. Weber (1964). Groundwater management for the nation's future - computer simulation of groundwater basins. Proc. ASCE, J. Hyd. Div. Hy 7.

Yeh, W.W.G. (1975). Aquifer parameter identification. Proc. ASCE, J. Hyd. Div. Hy 9.

8 Short-Term Forecasting of Rainfall and Runoff

Rafael L. Bras

On-line hydrologic forecasting is an estimation-identification-prediction problem. A model is used to forecast the state of the system. The forecast is compared with the noisy observed state to estimate the true state and to calibrate the model with new information. The adjusted model is then used in the next forecast.

In this chapter, on-line hydrologic forecasting is applied to short-term forecasting of discharges into a river basin and to rainfall over small urban areas. The first application is relevant to flood warning methodologies such as the National Weather Service River Forecasting System. The second application is valuable in real-time operation of urban runoff retention facilities for the control of local flooding and the prevention of water quality degradation in receiving water bodies.

Deterministic models of runoff and rainfall with an added stochastic component are used. The Kalman filter will be used as a linear, time varying sequential procedure for filtering information, estimating present state values, and identifying necessary modifications in model parameters.

SHORT-TERM FORECASTING OF RUNOFF

We consider the lumped and distributed alternatives.

The Lumped Alternative (Hino 1973)

Runoff or discharge can be generally represented by an equation of the following form. This is a stationary linear representation similar to the unit hydrograph concept.

$$
\begin{aligned}
Q_{1,t} = {} & \alpha^1_{11} Q_{1,t-1} + \alpha^1_{12} Q_{1,t-2} + \cdots + \alpha^1_{1j} Q_{1,t-j} \\
& + \alpha^1_{21} Q_{2,t-1} + \alpha^1_{22} Q_{2,t-2} + \cdots + \alpha^1_{2j} Q_{2,t-j} \\
& + \cdots \\
& + a^1_{11} i_{1,t-1} + a^1_{12} i_{2,t-2} + \cdots + a^1_{1\ell} i_{1,t-\ell}
\end{aligned}
$$

$$+ a^1_{21}\, i_{2,t-1} + a^1_{22}\, i_{2,t-2} + \cdots + a^1_{2\ell} i_{2,t-\ell}$$

$$+ \cdots \qquad (8.1)$$

$$+ W_{1,t-1} \cdot$$

$$Q_{m,t} = \alpha^m_{11}\, Q_{1,t-1} + \alpha^m_{12}\, Q_{1,t-2} + \cdots + \alpha^m_{1j}\, Q_{1,t-j}$$

$$\cdots$$

$$+ a^m_{11}\, i_{1,t-1} + a^m_{12}\, i_{1,t-2} + \cdots + a^m_{1\ell}\, i_{1,t-\ell}$$

$$\cdots$$

$$+ W_{m,t-1} \cdot$$

This is the general representation of discharge Q at m (i = 1 to m) locations at time t. It is clear that in this general formulation, discharge at a point i at time t can be a function of j previous discharges and ℓ previous rainfall inputs, i, at any of n different locations. This relation can be considerably simplified since discharge at any location is usually assumed to be a function of previous discharges at that location and rainfall at contributing higher locations.

Equation (8.1) can be expressed in matrix notation as

$$\underline{Q}(t) = \underline{M}(t-1)\, \underline{h}(t-1) + \underline{W}(t-1) \,. \qquad (8.2)$$

$\underline{Q}(t)$, the m-dimensional vector of discharges, is

$$\underline{Q}(t) = [Q_{1,t},\ Q_{2,t},\ \cdots\ Q_{m,t}] \qquad (8.3)$$

$\underline{h}(t-1)$ is a vector with $(j \times m + \ell \times n) \times m$ elements. The form of this vector is:

$$\underline{h}(t-1) = \Big[\alpha^1_{11}\alpha^1_{12}\cdots\alpha^1_{1j};\ \alpha^1_{21}\cdots\alpha^1_{2j}\cdots a^1_{11}a^1_{12}\cdots a^1_{1\ell};\ a^1_{21}a^1_{22}\cdots a^1_{2\ell}\cdots$$
$$\alpha^2_{11}\alpha^2_{12}\cdots\alpha^2_{1j};\ \alpha^2_{21}\cdots\alpha^2_{2j}\cdots a^2_{11}a^2_{12}\cdots a^2_{1\ell};\ a^2_{21}a^2_{22}\cdots a^2_{2\ell};\cdots\Big]^T \,. \qquad (8.4)$$

The elements of the vector are the coefficients of the discharge and input variables that appear in (8.1).

M is a matrix of m rows and $(j \times m + \ell \times n) \times m$ columns. The matrix is diagonal with terms of the form:

$$y = \Big[Q_{1,t-1}\ Q_{1,t-2}\cdots Q_{1,t-j};\ Q_{2,t-1}\cdots Q_{2,t-j}\cdots$$

$$Q_{m,t-1}\cdots Q_{m,t-j};\ i_{1,t-1}\ i_{1,t-2}\cdots i_{1,t-\ell};\qquad (8.5)$$

$$i_{2,t-1}\cdots i_{2,t-\ell}\cdots i_{n,t-1}\cdots i_{n,t-\ell}\Big]$$

$\underline{W}(t-1)$ is a vector of m elements equal to the white noise component $W_{1,t-1}\cdots W_{m,t-1}$ shown in (8.1)

Substituting a variable vector $\underline{Z}(t)$ for $\underline{Q}(t+1)$ in (8.2) results in an equation of the form:

$$\underline{Z}(t) = \underline{M}(t)\ \underline{h}(t) + \underline{W}(t)\ . \qquad (8.6)$$

The above can be interpreted as a measurement equation where $\underline{Z}(t)$ is the observation of the state vector $\underline{h}(t)$.

$\underline{Z}(t)$ is the vector of observed discharges at time t. The dynamics of the state vector $\underline{h}(t)$ are simply of the form

$$\underline{h}(t+1) = \underline{h}(t) + \underline{V}\ (t) \qquad (8.7)$$

where $\underline{V}(t)$ is a white noise component that adds uncertainty to the otherwise invariant state.

Equations (8.6) and (8.7) define a linear dynamic system that can be modeled within the Kalman filter framework. Using a Kalman filter on the system and measurement equations results in the following estimates:

$$\underline{\hat{h}}(t+1|t) = \underline{\hat{h}}(t|t)$$

$$\underline{\hat{h}}(t+1|t+1) = \underline{\hat{h}}(t|t) + \underline{K}(t+1)\ \{\underline{Z}(t+1) - \underline{M}(t+1)\ \underline{\hat{h}}(t|t)\}$$

$$\underline{K}(t+1) = \underline{P}(t+1|t+1)\ \underline{M}^T(t+1)\ \underline{R}^{-1}(t+1)$$

$$\underline{P}(t+1|t) = \underline{P}(t|t) + \underline{C}(t) \qquad (8.8)$$

$$P(t+1|t+1) = P(t+1|t) - P(t+1|t)M^T(t+1)\ \big[R(t+1) +$$

$$M(t+1)\ P(t+1|t)M^T(t+1)\big]^{-1}\ \underline{M}(t+1)\ \underline{P}(t+1|t)$$

where

$$R(t) = E\big[W(t)W(t)\big]^T \text{ and } C(t) = E\big[V(t)V(t)^T\big]$$

Equation (8.8) gives an estimate of the runoff model parameters sequentially utilizing information. These new parameters can then be used to forecast until a new observation becomes available.

The above lumped formulation is particularly attractive for use with the traditional unit hydrograph approach; the runoff model simplifies considerably because discharge is only a function of past inputs. Matrix $\underline{M}(t)$ represents observed rainfall, and vector $\underline{h}(t)$ elements are the unit hydrograph ordinates.

There are several weaknesses of the lumped formulation:

- It has no explicit relation between the identified unit hydrograph and the physical parameters of the basin.
- The interpretation of the measurement equation (8.6) eliminates the explicit consideration of true observation errors due to instruments or techniques.
- The model uses an essentially time-invariant (see (8.7)) linear system. Since we know that runoff is nonlinear in nature, we cannot expect to achieve a steady-state solution for the catchment response function, $\underline{h}(t)$.
- Error terms $\underline{W}(t)$ and $\underline{V}(t)$ are hard to describe or quantify in terms of their covariances. Particularly, the term $\underline{V}(t)$ has no obvious physical significance which could be used in its estimation.
- The dimensions of the matrices and vectors can be unmanageable, if fine resolution of the unit hydrograph is required, if the "memory" ℓ is large, and if we consider discharge at more than one location.

The possibility of using the above approach in conjunction with some of the existing nonlinear, distributed rainfall-runoff models is being studied. The refined models can be used in determining good initial solutions.

The Distributed Alternative

Spatially distributed rainfall-runoff models are widely used in hydrology. Advances in computer technology have permitted the numerical solution of the continuity and momentum equations over simple models of real basins. Most of these numerical solutions can be expressed as a nonlinear state-space difference equation of the following form (see Bras and Rodriguez-Iturbe 1975 and Muzik 1974).

$$\underline{Q}(t) = \underline{F}(Q(t-1)) + \underline{Bi}(t-1) + \underline{GW}(t-1) \tag{8.9a}$$

where

$\underline{Q}(t)$ = vector of discharges at n different locations

$\underline{i}(t)$ = vector of effective rainfall intensities at m locations

$B = (k_1 + \ldots + k_M + \ell) \times M$ input transformation matrix

$\underline{W}(t)$ = vector of white noise

$\underline{G}$ = noise transformation matrix

$\underline{F}$ = functional matrix operating on the previous time state vector

The system given by (8.11) can be observed by a monitoring network of the form

$$\underline{Z}(t) = \underline{H}\,\underline{Q}(t) + \underline{V}(t) \tag{8.9b}$$

where

$\underline{Z}(t)$ = k vector ($k \leq n$) of discharge observations

$\underline{V}(t)$ = k vector of white noise measurement errors

$\underline{H}$ = k x n matrix representing the discharge monitoring system

This matrix is sparse with only one nonzero element per row in the column corresponding to the desired discharge observation (see Bras and Rodriguez-Iturbe 1975).

There are $k_1 + k_n + \ell$ elements in this vector, where k_1 is the degree of discreteness of the element in the finite difference solution; ℓ is the number of nodes in the basin.

Except for the nonlinearity of (8.9a) arising from the matrix F, the system clearly falls within the Kalman filter framework. Fortunately, the nonlinearities of (8.9a) are small (exponents in the range 1.33 to 1.67). This permits the use of the so-called extended Kalman filter, which is just a linearization of the system around the previous estimate of the state variable (see Schweppe 1973). For the nonlinear system of (8.9a) and (8.9b) the extended Kalman filter formulation becomes:

$$\hat{\underline{Q}}(t+1|t+1) = \hat{\underline{Q}}(t+1|t) + \underline{K}(t+1)\left[\underline{Z}(t+1) - \underline{H}\,\hat{\underline{Q}}(t+1|t)\right]$$

$$\hat{\underline{Q}}(t+1|t) = \underline{F}(\hat{\underline{Q}}(t|t)) + \underline{B}\,\underline{i}(t)$$

$$\underline{K}(t+1) = \underline{P}(t+1|t+1)\,\underline{H}^T\,\underline{R}^{-1}(t+1)$$

$$\underline{P}(t+1|t+1) = \underline{H}^T\underline{R}^{-1}(t+1)\,\underline{H} + \underline{P}^{-1}(t+1|t)^{-1} \qquad (8.10)$$

$$\underline{P}(t+1|t) = \underline{F}^1(\hat{\underline{Q}}(t|t))\,\underline{P}(t|t)\,\underline{F}^{1^T}(\hat{\underline{Q}}(t|t)) + \underline{G}\,\underline{C}(t)\,\underline{G}^T$$

$$\underline{P}(0|0) = \underline{\Psi}$$

$$\hat{\underline{Q}}(0|0) = \underline{0}$$

$\hat{\underline{Q}}(t|t)$ is the estimated state vector at time t given information up to time t. $\underline{F}^1(\hat{\underline{Q}}(t|t))$ is the first derivative of the functional matrix $\underline{F}$ evaluated at $\hat{\underline{Q}}(t|t)$. $\underline{P}(t|t)$ is again the mean square error of the estimation matrix at time t given information up to time t. The matrices $\underline{R}(t)$ and $\underline{C}(t)$ are the covariance matrices of the measurement and model white noise components, respectively. All other terms have been previously defined.

It is clear that (8.10) is an on-line estimating-forecasting procedure. As data or observations of discharge and rainfall become available a new estimate of discharge, $\hat{\underline{Q}}(t|t)$, can be made. A forecast into the future, before observations at t+1 are available, is made by using the second equation in (8.10) or,

$$\hat{\underline{Q}}(t+1|t) = \underline{F}(\hat{\underline{Q}}(t|t)) + \underline{B}\,\underline{i}(t) \qquad (8.11)$$

This forecasting procedure uses all available information to obtain an on-line prediction one time step ahead of real-time. The time step of prediction depends on the definition of the matrices $\underline{F}$ and $\underline{B}$ (see Bras and Rodriguez-Iturbe 1975).

The distributed alternative has several advantages. First, the model allows a spatially varying basin representation. Second, the model is theoretically correct and based on the equations of motion and mass conservation. Model parameters $\underline{F}$ and $\underline{B}$ are functions of measurable basin characteristics and of the numerical techniques used. Also, the model is nonlinear and thus realistic. The linearities involved are trivial relative to those in the lumped alternative. Third, the model fits perfectly and neatly into the Kalman filter framework. Finally, this model framework explicitly takes into account discharge measurement error and could be expanded to consider rainfall measurement errors.

Further refinement of the forecasting system of (8.10) is possible by adding a parameter identification step that could allow adjustment of the model parameters. The model error, $\underline{W}(t)$, and its covariance function, $\underline{C}(t)$, are parameters that are particularly difficult to define and could be estimated using observed data. The procedure would consist of maximizing a logarithmic likelihood function in order to find the optimal parameters. The log likelihood function for systems like that of (8.9) and (8.10) has been derived assuming Gaussian or near Gaussian distributions (see Schweppe 1973). This likelihood function fits into the framework of the Kalman filter and can be obtained simultaneously with the solution to (8.10). The maximum likelihood estimators for the parameters would then be used in the forecasting procedure until new observations become available.

The major disadvantages of the distributed approach are:

- The identification, if desired, of model, $\underline{W}(t)$, and measurement, $\underline{V}(t)$, error may involve considerable difficulties.

- The number of discharge locations considered will be limited by numerical efficiency considerations.

- The forecast will be good for time intervals less than the time of concentration. This condition allows the best application of the system to large basins. This restriction is introduced because (8.9) expresses discharge at time t, $\underline{Q}(t)$, as a function of rainfall at time t-1, $\underline{i}(t-1)$. This is approximately true and the approximation will be best for time intervals as stated above.

SHORT-TERM FORECASTING OF RAINFALL

In the previous section, a methodology for on-line runoff forecasting was presented. It was then mentioned that the methodology is valid for forecasting in large, slow-responding river basins. The forecast lag time cannot be more than the time of concentration of the basin. The reason for the limitation was that no attempt was being made to foretell the state of the rainfall input. In small urban areas, where the hydrologist may worry about local flooding and runoff pollution loads, the above conditions are not satisfied. We are therefore compelled to attempt short-term forecasting of rainfall to deal with fast-responding urban areas and to be able to make forecasts beyond the concentration time of the basin.

Jamieson and Wilkinson (1972) suggested the use of a first-order autoregressive model as a tool for short-term forecasting. They found that such stationary models explained only about 45% of the rainfall time series. In fact, several other time series models could be suggested. For example, besides the autoregressive model, we can possibly use stationary moving average models, mixed models, or

Box and Jenkins (1970), nonstationary, autoregressive integrated moving average models (ARIMA). The main limitations of the above are that they have no parameters representing real conditions. Their parameters must be statistically determined from data. In an on-line, real-time forecasting situation data is limited and available piecewise. To "fit" one of the statistical models, the storm event would have to be observed in spatial detail for extremely long times before arrival at the location of interest. Even then, the statistical meaning of the obtained parameters could be poor.

The above discussion obviously implies that we need some knowledge of the structure and behavior of the rainfall process so as to minimize data requirements. At the extreme, we would need a completely deterministic model of rainfall. Meteorologists have only imperfect and very complex deterministic storm models based on atmospheric conditions. They are more an exercise to study atmospheric interactions of climatic variables than rainfall predictors. Their parameters and structure would not lend themselves to efficient on-line modification resulting from observed data.

It is clear that the short-term rainfall prediction problem is a very hard one. We are only starting to see its possible solutions. The folllowing is an approach that makes some reasonable assumptions to allow us to study the problem.

Bras and Rodriguez-Iturbe (1976) suggested a nonstationary, time-varying, multidimensional rainfall generator that preserves first- and second-order statistics in time and space, and assumes that storms have a basic structure. The details of the model will not be discussed here; only a summary of important points will be presented. The reader is referred to Bras and Rodriguez-Iturbe (1976) for the development and some examples, as well as to Wilson (1976) for applications.

The model assumes that the interior of a storm event with a given depth and duration can be modeled as:

$$i(\underline{x}_i,t) = i_\mu(\underline{x}_i,t) + \eta(\underline{x}_i,t) \tag{8.12}$$

where

$i(\underline{x}_i,t)$ = rainfall intensity at a point with coordinate vector $\underline{x}_i$ at time t.

$i_\mu(\underline{x}_i,t)$ = mean intensity at $\underline{x}_i$ and t.

$\eta(\underline{x}_i,t)$ = noisy residuals obeying a certain covariance function in time and space.

It is assumed that each point in space will have the same average mass distribution. This mass distribution is obtained for various types of storms (convective, frontal) in terms of a mass accumulation curve without dimension. Since each point in space will, on the average, have the same accumulation history, the term $i_\mu(\underline{x}_i t)$ will be the same everywhere in space, but will be translated by the velocity of the event. Assuming that the storm were moving in the x direction with velocity U then,

$$i_\mu(\underline{x}_i,t) = i_a(t - \frac{x_i}{U}) , \tag{8.13}$$

where

$i_a(t)$ = average precipitation at time t where t is now the time since the beginning of rainfall at a given point; this is obtained from the average dimensionless accumulation curve determined for various types of storms.

The space and time correlation of the process is embodied in the covariance function of the residuals. To define this covariance function, we assume that storms, like other meteorological processes, obey Taylor's hypothesis of turbulence. This is generally useful for translating processes with relatively weak time dependence within their moving coordinate system such that time dependence in a fixed coordinate system is dominated by the average translating motion (see Bras and Rodriguez-Iturbe 1976). This allows us to suggest the following form for the covariance function, assuming for simplicity that the storm is moving in the x direction:

$$E[\eta(\underline{x}_i,t')\ \eta(\underline{x}_j,t'')] = \sigma(\underline{x}_i,t')\ \sigma(\underline{x}_j,t'')\ r(\underline{x}_i,t';\ \underline{x}_j t'')$$

$$= \sigma_a(t' - \frac{x_i}{U})\ \sigma_a(t'' - \frac{x_j}{U}) \qquad (8.14)$$

$$r(y_i,x_i + Ut';y_j,x_j + Ut'')$$

where

$r(\cdot)$ = correlation function of process

$\sigma(\underline{x}_i,t)$ = standard deviation of rainfall intensity at point $\underline{x}_i$ and time t

U = velocity in x direction

x_i,y_i = x and y coordinates of point i

Based on some available data, it is further assumed that rainfall intensities have isotropic spatial correlation which then results in:

$$r(x_i,t';x_j,t'') = r(v) = r(\sqrt{(y_j-y_i)^2 + ((x_j + Ut'') - (x_i + Ut'))^2}) \qquad (8.15)$$

Equation (8.15) still assumes that storm movement is in the x direction. Notice that isotropy is on the variables $x_1 = y$ and $x_2 = x + Ut$.

The standard deviations, $\sigma(\underline{x}_i,t') = \sigma_a(t' - \frac{x_i}{U})$, are obtained from data on the variation around the mean behavior $i_a(t)$. Again $\sigma_a(t)$ should be typical for certain types of storms. Storm interiors are then generated from an equation

$$\underline{i}(\underline{x}_i,t) = i_a(t - \frac{x_i}{U}) + \sigma_a(t - \frac{x_i}{U})\ R(\underline{x}_i,t) \qquad (8.16)$$

where $R(\underline{x}_i,t)$ are mean 0, variance 1 residuals obeying the time and space correlation given in (8.15). These residuals are generated using spectral representations of the correlation function (see Bras and Rodriguez-Iturbe (1976) for details).

The characteristics of the previous model that are important for the forecasting

exercise are: its "banded" structure, where the mean and standard deviations in a band perpendicular to storm movement are homogeneous or constant, and that the covariance in time and space has a structure given by (8.15). The function r(v) can take different forms, for example:

Single Exponential

$$r(v) = e^{-\alpha|v|} \tag{8.17}$$

Quadratic Exponential

$$r(v) = e^{-\alpha^2 v^2}$$

Bessel Form

$$r(v) = |v| \quad bk_1(|v|\ b)$$

where k_1 is a first-order Bessel function of the second kind.

Therefore the covariance is explained in terms of a parameter, α, α^2, or b^2 and the velocity U.

Bras and Rodriguez-Iturbe (1975) suggest that the above continuous model can be approximated by a multivariate form:

$$i(t) = i_\mu(t) + A(t-1)(i(t-1) - i_\mu(t-1)) + B(t-1)W(t-1) \tag{8.18}$$

where

$i(t)$ = n x 1 vector of rainfall intensities at n locations at time t

$i_\mu(t)$ = n x 1 vector of mean intensities

$A(t)$, $B(t)$ = time-varying n x n matrices that can be estimated from the given covariance function (see Bras and Rodriguez-Iturbe 1975)

$W(t)$ = n x 1 vector of white noise with zero mean and unit variance

$E[W(t)W^T(t)] = I$

The above approximation is particularly true for a small time correlation of exponential decay.

We assume that there exists a fairly extensive telemetric rainfall network surrounding an area of interest and that one of the correlation function forms given by (8.17) is applicable. As the storm moves into the area rainfall intensity data are recorded. After some time, before the storm arrives at the area of interest, a "snapshot" of intensities at a given time, t_1, over the area is taken. At a time $t_2 > t_1$ (Δt not too big), take another sampling of the intensities, another "snapshot" of conditions in the area A at the two times.

$$C(\alpha,\beta) = \frac{1}{A}\int_A Z(x,y,t_1)\ Z(x+\alpha,\ y+\beta,\ t_2)dx\ dy \tag{8.19}$$

should be a maximum, if Taylor's Hypothesis holds, for a given α_0 and β_0, which are

some displacements in the x and y direction. $Z(x,y,t)$ is a functional description of the vector Z of observed rainfall values. Z is an n x 1 vector. $Z(x,y,t)$ could be obtained using surface fitting techniques. The average velocity of translation of the storm between time t_1 and t_2 can be approximated by

$$V_x = \frac{\alpha_0}{t_2-t_1};\ V_y = \frac{\beta_0}{t_2-t_1} \tag{8.20}$$

Now we assume that the obtained velocity and direction will prevail over the next time step. Similarly, taking averages over bands perpendicular to storm movement results in estimates $\hat{i}_\mu(\underline{x}_i,t_2)$. Standard deviations can also be obtained over those bands to define $\hat{\sigma}(\underline{x}_i,t_2)$.

Dividing by the obtained variances and subtracting the means, the normalized intensity values could be used to fit a given spatial correlation function form. For example, assuming a single exponential form (8.17) and using the isotropy assumption, the parameter α could be estimated as $\hat{\alpha}$. Using the estimated velocities $\hat{V}_x$ and $\hat{V}_y$, the assumed correlation form, and the estimated parameter $\hat{\alpha}$, we can obtain estimates of the matrices, $A(t_3 - \Delta t)$ and $B(t_3 - \Delta t)$ in (8.18) and express them as $\hat{A}(t_3 - \Delta t)$ and $\hat{B}(t_3 - \Delta t)$.

An estimate of the state vector at time t_3 with information up to time t_2 can be:

$$\hat{i}(t_3|t_2) = \hat{A}(t_3-\Delta t)\ \{Z(t_2) - \hat{i}_\mu(t_2)\} + \hat{i}_\mu(x_i-\hat{V}_x\Delta t,\ y_i-\hat{V}_y\Delta t, t_2) \tag{8.21}$$

where the last term in (8.21) is the mean vector at time t_2 translated (rearranged) by the estimated velocities. The vector $Z(t_2)$ is an n x 1 vector of the observed rainfall at the n points in space. It is given by

$$Z(t) = H\ i(t) + V(t) \tag{8.22}$$

where

i(t) = true vector of rainfall values

H = n x n matrix defining measurement network

V(t) = error of observation

At time t_3 (8.19) can again be optimized using "snapshots" of rainfall intensities over the area at time t_2 and t_3. New velocity estimates, V_x and V_y, can be obtained using (8.22). Band averaging should again yield estimates at $\hat{i}_\mu(\underline{x}_i,t_3)$ and $\hat{\sigma}(\underline{x}_i,t_3)$. The correlation function coefficient $\hat{\alpha}$ can be re-estimated, again using observed intensity values at time t_3. Since this parameter should in theory be constant we suggest that some sort of filtering or estimation algorithm should be used to avoid radically different estimates at different times. Now we can obtain new matrices, $\hat{A}(t_4 - \Delta t)$, $\hat{B}(t_4 - \Delta t)$ using the covariance function with its estimated parameters. A new forecast to time t_4 can be obtained with an equation similar to (8.21):

$$\hat{i}(t_4|t_3) = \hat{A}(t_4 - \Delta t)\ \{\hat{i}(t_3|t_3) - \hat{i}_\mu(t_3)\} + \hat{i}_\mu(x_i-\hat{V}_x\Delta t,\ y_i-\hat{V}_y\Delta t,\ t_3) \tag{8.23}$$

where $\hat{i}(t_3|t_3)$ is a filtered estimate resulting from a weighted average of the model prediction at time t_3 and the observations at time t_3, $Z(t_3)$. The filtering could be done using the linear Kalman filter which results in:

$$\hat{i}(t_3|t_3) = \hat{i}(t_3|t_2) + K(t_3)\{Z(t_3) - H\,\hat{i}(t_3|t_2)\}$$

$$K(t_3) = P(t_3|t_3)\,H^T R^{-1}$$

$$P(t_3|t_3) = P(t_3|t_2) - P(t_3|t_2)\,H^T\{R + H\,P(t_3|t_2)H^T\}^{-1} H\,P(t_3|t_2) \qquad (8.24)$$

$$P(t_3|t_2) = \hat{A}(t_3-\Delta t)\,P(t_2|t_2)\,\hat{A}^T(t_3-\Delta t) + \hat{B}(t_3-\Delta t)\hat{B}^T(t_3-\Delta t)$$

$$P(t_2|t_2) = E[i(t_2)\,i^T(t_2)]$$

$\equiv$ process covariance matrix given by the estimated covariance function at time t_2

$$E[V(t)\,V^T(t)] = R$$

Forecasting for all other times would follow the same pattern.

CONCLUSIONS

These ideas are obviously rough, but they are being refined and studied.

The difference between long- and short-term forecasting are important. Long-term forecasting is usually an off-line exercise intended to generate random events that preserve, to a certain extent, the statistical behavior of historical data. We are usually dealing with aggregate parameters like monthly discharges, daily rainfalls and so forth. Short-term forecasting is done in real-time, on-line, at the same time as the event. We are usually interested not in aggregate events but in time histories of occurrences. We want to continuously update the forecasting procedure by using observed occurrences. We are interested not only in statistical behavior but in the absolute value of forecasted variables. Furthermore, we want to forecast future behavior with some degree of accuracy. In long-term forecasting, we usually have ample historical data from which to statistically estimate model parameters. In short-term forecasting this is not the case. We therefore, in order to satisfy the accuracy and prediction criteria, must use all possible prior information about the process to be able to define the structure and parameters of underlying behavior as much as possible. The more we define an underlying structure, the more our statistical requirements are reduced and accuracy increases. Nevertheless, our model structure must be flexible and simple enough to allow continuous parameter updating as data become available.

This paper has presented possible approaches to short term forecasting of runoff and rainfall. The mechanics and the models used, particularly the rainfall model, may be questioned. Nevertheless, the approach and general concept we believe to be correct.

Kalman filtering has been used throughout owing to the linear or near-linear behavior of the models used. It is certainly a very powerful technique that handles the deterministic (physical) and stochastic components of the problem in a unified way. It accounts for both model and measurement uncertainty in a stepwise, sequential (on-line) manner. It also can handle parameter estimation problems when they arise.

ACKNOWLEDGEMENTS

This paper was presented and prepared with partial support from M.I.T.'s Department of Civil Engineering and its Water Resources Division. The U.S. Department of Commerce, National Weather Service, also sponsored this presentation and the development of the ideas presented here.

REFERENCES

Amorocho, I. (1967). The non-linear prediction problem in the study of the runoff cycle. Water Resour. Res. 3 (3).

Barrera, A., and F.E. Perkins (1967). An extension of the role of linear systems analysis in hydrograph theory. Hydrodynamics Lab. Rep. No. 106, M.I.T. Dept. of Civil Engineering, Cambridge, Massachusetts, Sept.

Blank, D., and J.W. Delleur (1968). A program for estimating runoff from Indiana watersheds, Part I: Linear system analysis in surface water hydrology and its applications to Indiana watersheds. Technical Report No. 4, Purdue University, Water Resources Research Center, Lafayette, Indiana, Aug.

Box, G.E.P., and G.M. Jenkins (1970). Time Series Analysis - Forecasting and Control. Holden-Day, San Francisco.

Bras, R.L., and I. Rodriguez-Iturbe (1975). Rainfall-runoff as spatial stochastic processes: Data collection and synthesis. Hydrodynamics Lab, Report No. 196, M.I.T. Dept. of Civil Engineering, Cambridge, Massachusetts.

Bras, R.L., and I. Rodriguez-Iturbe (1976). Rainfall generation: A non-stationary time varying multidimensional model. Water Resour. Res. 12 (3).

Eagleson, P.S. (1967). Optimum density of rainfall networks. Water Resour. Res. 3 (4): 1921-1033.

Eagleson, P.S., R. Majia, and F. March (1966). The computation of optimum realizable unit hydrographs. Water Resour. Res. 2 (4): 755-764; and discussion in 4 (1): 213-217, 1968.

Gustavsson, I. (1972). Comparison of different methods for identification of industrial processes. Automatica 8: 127-142.

Hino, M. (1970). Runoff forecast by linear predictive filter. Proc. ASCE, J. Hyd. Div. Hy3: 681-701.

Hino, M. (1973). On-line prediction of a hydrologic system. Presented at XV Congress of IAHR, Istanbul, Sept.

Jamieson, D.G. (1974). Seminar presented at M.I.T. Dept. of Civil Engineering.

Jamieson, D.G., and J.C. Wilkinson (1972). River Dee research program 3, A short-

term control strategy for multi-purpose reservoir systems. Water Resour. Res. 8 (4): 911-920.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Kalman, R.E. (1960). A new approach to linear filtering and prediction problems. Trans. ASME, J. Basic Eng. 82: 35-44.

Kalman, R.E., and R.S. Bucy (1961). New results in linear filtering and prediction theory. Trans. ASME, J. Basic Eng. 83: 95-108.

Koivo, A.J., and G.R. Phillips (1971). Identification of mathematical models for DO and BOD concentrations in polluted streams from noise corrupted measurements. Water Resour. Res. 7 (4): 853-862.

Koivo, A.J., and G.R. Phillips (1972). On determination of BOD and parameters in polluted stream models from DO measurements only. Water Resour. Res. 8 (2): 478-486.

Lee, E.S., and I. Hwang (1971). Stream quality modeling by quasilinearization. J. Water Pollut. Control Fed. 43 (2): 306-317.

Leondes, C.T., Ed. (1970). Theory and Applications of Kalman Filters. Advisory Group for Aerospace Research and Development, NATO, Paris, France.

McGarty, T.P. (1971). The estimation of the constituent densities of the upper atmosphere by means of a recursive filtering algorithm. IEEE Trans. Autom. Control 16 (6): 817-823.

Moore, R.L., and F. Schweppe (1972). Adaptive control for nuclear power plant load changes. Fifth World IFAC Congress, Paris.

Moore, S.F. (1971). The Application of Linear Filter Theory to the Design and Improvement of Measurement Systems for Aquatic Environments. Thesis presented to the University of California, Davis, California, in partial fulfillment of Ph.D.

Muzik, I. (1974). State variable model of overland flow. J. Hydrol. 22: 347-364.

Nash, J.E. (1960). A unit hydrograph study with particular reference to British Catchments. Proc. Inst. Civ. Eng. 17: 249-282.

Rhodes (1971). A tutorial on estimation and filtering. IEEE Trans. Autom. Control, Dec.

Schweppe, F.C. (1973). Uncertain Dynamic Systems. Prentice-Hall, Englewood Cliffs.

Whitehead, P., and P. Young (1974). A dynamic stochastic model for water quality in part of the Bedford-Ouse River system. IFIP Working Conference on Modelling Water Resource Systems, Ghent, July.

Wiener, N. (1949). The Extrapolation, Interpolation and Smoothing of Stationary Time Series. Wiley, New York.

Wilson, C.B. (1976). Effects of Spatially-Distributed Rainfall on Storm Runoff from a Small Catchment. M.S. Thesis, M.I.T. Dept. of Civil Engineering.

Zoch, R.T. (1934, 1936, 1937). On the relation between rainfall and streamflow. Mon. Weather Rev. Part I, 62: 315-322; Part II, 64: 105-121; Part III, 65: 135-147.

9 A Self-Tuning Predictor Applied to River Flow Forecasting

Torkil Ganendra

INTRODUCTION

Stochastic difference equation models are ideally suited for real-time streamflow forecasting, particularly because the autoregressive terms with updating of information prevent cumulative forecast divergence. Furthermore, the identification of the rainfall-runoff process can be bypassed by formulating the model as a predictor. The predictor algorithm should have a small computational and storage requirement to allow for minicomputer or microprocessor implementation.

The predictor that adjusts its parameters in real-time using least squares, such that in the limit the minimum square error predictor is obtained, is referred to as the self-tuning predictor (Wittenmark 1974). The predictor has been tested on five sets of data; in this paper only results from the Brosna Catchment in Ireland will be presented.

OPTIMAL PREDICTION

The assumption that must be made is that the system to be modeled is linear and time-invariant. In the predictor derivation, the deterministic input is omitted; it can be introduced later as an auxiliary variable. Consider the simple n-th order stochastic process,

$$A(q^{-1})\, y(t) = C(q^{-1})\, e(t) \tag{9.1}$$

where $e(t)$ is an independent $N(0,\sigma)$ random variable, $y(t)$ the measured output, and the polynomials $A(q^{-1})$ and $C(q^{-1})$ are described by

$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_n q^{-n}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \ldots + c_n q^{-n}$$

q^{-i} being the backward-shift operator

$$q^{-i}\, y(t) = y(t-i)$$

The following notation is introduced. The k-step ahead prediction of the output y(t+k), based on the past output measurements y(t), y(t-1), ... is denoted by $\hat{y}(t + k|t)$ and the prediction error is given by

$$\varepsilon(t+k) = y(t+k) - \hat{y}(t + k|t) \tag{9.2}$$

The minimum square error predictor, i.e., the predictor that minimizes the loss function

$$V = E\{\varepsilon(t)^2\} \tag{9.3}$$

can be obtained by introducing the identity (Aström 1970, p. 167)

$$C(q^{-1}) = A(q^{-1})\, F(q^{-1}) + q^{-k} G(q^{-1}) \tag{9.4}$$

where

$$F(q^{-1}) = 1 + f_1 q^{-1} + \cdots + f_{k-1} q^{-k+1}$$

$$G(q^{-1}) = g_0 + g_1 q^{-1} + \cdots + g_{n-1} q^{-n+1}$$

Note that q^{-k} operates on $G(q^{-1})$ in $C(q^{-1})$ and that $F(q^{-1})$ has k terms. Writing (9.1) as

$$y(t+k) = \frac{C(q^{-1})}{A(q^{-1})}\, e(t+k) \tag{9.5}$$

and replacing $C(q^{-1})$ by its identity (9.4) gives

$$y(t+k) = F(q^{-1})\, e(t+k) + \frac{G(q^{-1})}{A(q^{-1})}\, e(t) \tag{9.6}$$

Eliminating e(t) in (9.6) by using (9.1) gives

$$y(t+k) = F(q^{-1})\, e(t+k) + \frac{G(q^{-1})}{C(q^{-1})}\, y(t) \tag{9.7}$$

If $\hat{y}$ is an arbitrary function of y(t), y(t-1), ... y(0), and e(t+1), e(t+2), ... e(t+k) are independent of y(t), then

$$E\{[y(t+k)-\hat{y}]^2\} = E\{[F(q^{-1})e(t+k)]^2\} + E\{[\frac{G(q^{-1})}{C(q^{-1})}\, y(t)-\hat{y}]^2\}. \tag{9.8}$$

The variance of e(t) is

$$E\{[y(t+k)-\hat{y}]^2\} \geq \sigma^2(1 + f_1^2 + \ldots + f_{k-1}^2) \tag{9.9}$$

The predictor objective is a minimum square error and this is satisfied when $\hat{y}$ is $\hat{y}(t + k|t)$, that is,

$$\hat{y}(t + k|t) = \frac{G(q^{-1})}{C(q^{-1})} y(t) \tag{9.10}$$

Eliminating y(t) using (9.2) and $C(q^{-1})$ using (9.4), the general optimal predictor is obtained

$$\hat{y}(t + k|t) = \frac{G(q^{-1})}{A(q^{-1})F(q^{-1})} \varepsilon(t) \tag{9.11}$$

Finally, the drawback of the predictor is that it contains k-1 more parameters than the original system it represents.

THE SELF-TUNING PREDICTOR

If the parameters of the process were known, then the optimal predictor could be readily calculated. For unknown processes, the parameters of the predictor can be directly estimated without reference to the process parameters. The algorithm to be described consists of two components, a parameter estimator, and the predictor itself. The parameter estimator has the self-tuning property, that is, if the parameters of the predictor converge, then the predictor obtained in the limit will be the minimum square error predictor. Wittenmark (1974) has shown that the self-tuning predictor is a special case of the self-tuning regulator (Aström and Wittenmark 1973), an algorithm developed for controlling a system with unknown parameters. Consequently, the results of theoretical and experimental investigations of the self-tuning regulator are applicable to the self-tuning predictor.

Rewriting the predictor (9.11) with $A(q^{-1})F(q^{-1}) = 1+q^{-1}\underline{A}(q^{-1})$ and $G(q^{-1}) = \underline{C}(q^{-1})$,

$$\hat{y}(t + k|t) = - \underline{A}(q^{-1})\hat{y}(t+k-1|t-1) + \underline{C}(q^{-1})\varepsilon(t) \tag{9.12}$$

By defining $\hat{y}(t + k|t)$ as the prediction of the measured streamflow y(t+k) and by introducing the precipitation input u(t) as an auxiliary variable, the following predictor is obtained

$$\hat{y}(t + k|t) = - \underline{A}(q^{-1})\hat{y}(t+k-1|t-1) + \underline{B}(q^{-1})u(t) + \underline{C}(q^{-1})\varepsilon(t) \tag{9.13}$$

where

$$\underline{A}(q^{-1}) = \alpha_1 + \alpha_2 q^{-1} + \ldots + \alpha_n q^{n-1}$$

$$\underline{B}(q^{-1}) = \beta_1 + \beta_2 q^{-1} + \ldots + \beta_m q^{m-1}$$

$$\underline{C}(q^{-1}) = \gamma_1 + \gamma_2 q^{-1} + \ldots + \gamma_\ell q^{\ell-1}$$

The parameters n, m, and ℓ are simply the number of terms in the respective polynomials. These can be determined quite easily as shown in the next section.

In order to simplify the notation, let $\hat{y}(t + k|t)$ be $\hat{y}(t+k)$ and define

$$\phi(t) = [-\hat{y}(t+k-1),\ldots -\hat{y}(t+k-n),u(t),\ldots u(t-m+1),\varepsilon(t),\ldots \varepsilon(t-\ell+1)]$$

$$\theta^T = [\alpha_1, \cdots \alpha_n, \beta_1, \cdots \beta_m, \gamma_1, \cdots \gamma_\ell]$$

so the predictor can be written

$$\hat{y}(t+k) = \phi(t)\theta \tag{9.14}$$

The problem now is to estimate the parameter vector θ at each sampling point, using the latest data. For this purpose a recursive least-squares parameter estimator is used (Aström 1974) satisfying the self-tuning requirement. Thus, the estimator is given by the following equations which are evaluated at each time step.

$$\varepsilon(t) = y(t) - \hat{y}(t) \tag{9.15}$$

$$P(t) = P(t-1)-P(t-1)\phi^T(t-1)[1+\phi(t-1)P(t-1)\phi^T(t-1)]^{-1}\phi(t-1)P(t-1) \tag{9.16}$$

$$\theta(t) = \theta(t-1) + P(t)\phi^T(t-1)\varepsilon(t) \tag{9.17}$$

P is the error covariance matrix of the parameter estimates normalized with respect to the noise variance. This algorithm avoids direct matrix inversion (the inverse term in (9.16) is a scalar) and the past data does not have to be stored for subsequent calculations. The minimum square error prediction is then computed as if the values of the estimated parameters were true ones. Hence, the predictor for y(t+k) is given by (9.14).

If the handling of multiple inputs is required, the only modification that is necessary is the extension of the θ and ϕ vectors to include the additional variables.

PREDICTOR CONVERGENCE

For the case where $C(q^{-1}) = 1$ the least squares technique results in an unbiased estimator; however with $C(q^{-1}) \neq 1$ the predictor also gives good results. It can be shown (Aström and Wittenmark 1973) that the self-tuning predictor that has enough parameters converges to the optimal predictor when the covariances satisfy

$$\gamma_{\varepsilon\varepsilon}(\tau) = E\{\varepsilon(t)\varepsilon(t+\tau)\} = 0 \qquad \tau = k, \ldots k+\ell-1 \tag{9.18}$$

$$\gamma_{\varepsilon\hat{y}}(\tau) = E\{\varepsilon(t+\tau)\hat{y}(t+k)\} = 0 \qquad \tau = k, \ldots k+n \tag{9.19}$$

Furthermore, if the predictor contains enough parameters, $\varepsilon(t)$ should be uncorrelated for $\tau \geqslant \ell+k$. To illustrate, the relevant autocorrelation and cross-correlation functions, $\hat{f}_{\varepsilon\varepsilon}(\tau)$ and $\hat{f}_{\varepsilon\hat{y}}(\tau)$, using a two-step-ahead predictor are shown in Fig. 9.1. The predictor used does not have enough parameters so $\hat{f}_{\varepsilon\varepsilon}(\tau) \neq 0$ for

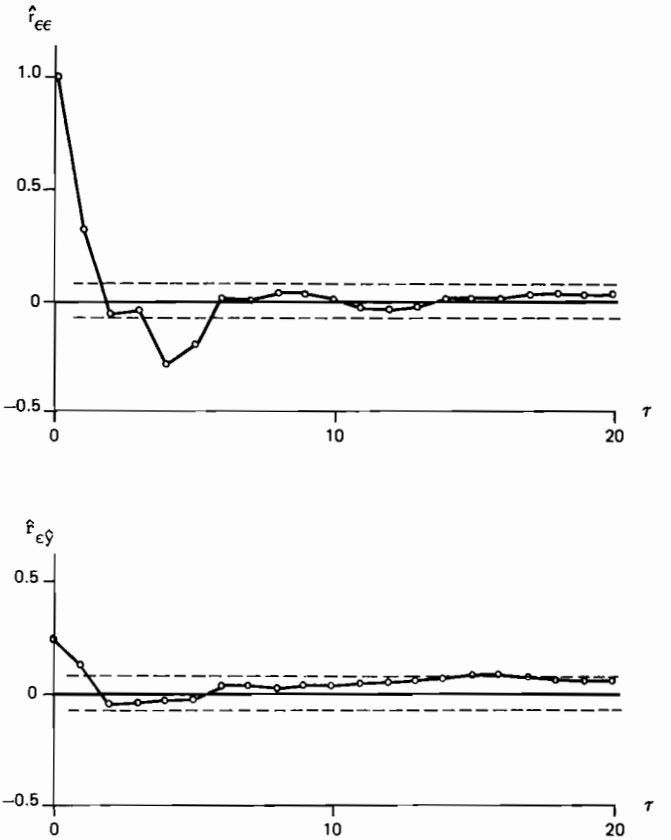


Fig. 9.1 Two-step ahead self-tuning predictor with n = 3, m = 2, and ℓ = 2. The dashed lines indicate the 95% confidence interval in which the covariances can be regarded as zero. $\hat{f}_{\varepsilon\varepsilon}$ is the autocorrelation function and $\hat{f}_{\varepsilon\hat{y}}$ is the cross-correlation function.

$\tau = 4$ and $\tau = 5$, although conditions (9.18) and (9.19) are satisfied. The number of parameters was n = 3, m = 2, ℓ = 2; the optimal predictor requires that m = 3, or that m = 2 when a pure time delay of one time step is applied to the input. The above procedure provides a convenient basis for determining the number of parameters, n,m, and ℓ.

The mean streamflow level

$$\bar{q} = \frac{1}{N} \sum_{t=1}^{N} q(t)$$

is nonzero, where q(t) is the measured streamflow. This mean level sometimes masks the information from the estimator such that (9.18) and (9.19) are not satisfied; so the output signal y(t) should be the deviation of the streamflow q(t) from the mean level

$$y(t) = q(t) - \bar{q}$$

In real-time, an estimated mean level is assumed which is then added to the streamflow prediction to obtain the true predicted flow.

The initial conditions for the algorithm can be taken as

$$P(0) = I \times 10^4$$

$$\theta(0) = \phi^T(0) = 0$$

however for $k > 1$ an approximation of $\phi^T(0)$ should be made. This averts an unstable condition during the first few iterations. In the case of the one-step-ahead predictor, only a few time steps are required before good predictions are obtained. It is a characteristic of the self-tuning predictor that near optimal predictions are obtained with apparently poor parameter estimates. This situation arises frequently during the initial stages of tuning.

Parameter convergence can be improved by preventing the elements of the P-matrix from diminishing too quickly. This is achieved by exponentially weighting the data so that more weight is given to recent data and less to older data with respect to the recursive parameter estimator. By introducing an exponential forgetting factor in (9.16), the following equation is obtained

$$\alpha(t)P(t) = P(t-1)-P(t-1)\phi^T(t-1)\left[\alpha(t)+\phi(t-1)P(t-1)\phi^T(t-1)\right]^{-1}\phi(t-1)P(t-1) \quad (9.20)$$

and $0 < \alpha(t) \leq 1$

where $\alpha(t)$ is the forgetting factor. The smaller the value of $\alpha(t)$, the less weight is given to past data. For no forgetting, $\alpha(t)$ is equal to 1. As the parameters converge, the degree of forgetting should decrease as $\alpha(t)$ tends to one; thus, greater confidence is given to the parameter estimates. $\alpha(t)$ is given by

$$\alpha(t) = \alpha(t-1)\alpha_0 + (1-\alpha_0) \quad (9.21)$$

The parameter α_0 is chosen so that $0 << \alpha_0 \leq 1$. Typically

$$\alpha(0) = 0.95 - 0.98$$

$$\alpha_0 = 0.99 - 0.999$$

The effect of forgetting past data (by using (9.20) and (9.21)) on the convergence

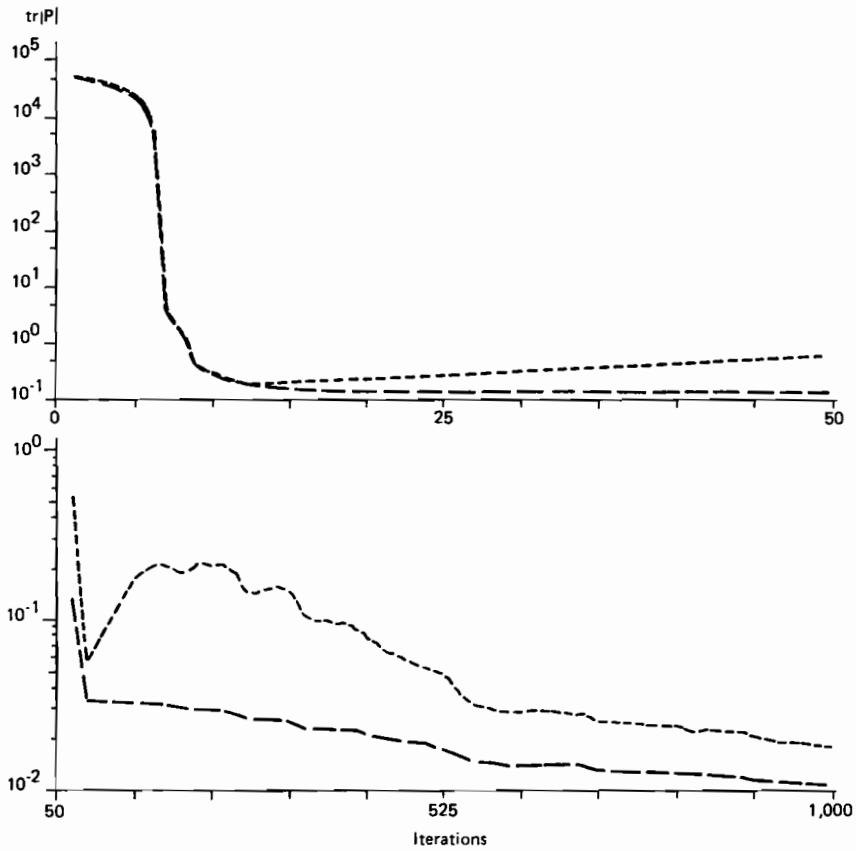


Fig. 9.2 Variation of tr|P| over 1,000 iterations with $P(0) = I \times 10^4$. ----- $\alpha_0 = 0.99$, $\alpha(0) = 0.95$; – – – – $\alpha_0 = 1.0$, $\alpha(0) = 1.0$.

of tr|P| is shown in Fig. 9.2. The case when $\alpha(0) = \alpha_0 = 1$ corresponds to no forgetting of past data; see (9.16). As expected, the forgetting of past data prevents tr|P| from becoming too small while the parameter estimates are still poor.

If the parameters θ are slowly time-varying, α_0 in (9.21) should be taken as 1 and the forgetting factor $\alpha(0)$ set accordingly to a value less than 1. However, if the data contain a high level of noise, as is often the case with hydrological data, then the estimator may cause the parameters to track the noise, as shown in Fig. 9.4. This can result in a poor predictor performance. Figure 9.3 shows the parameter variation when there is no forgetting of past data.

REAL-TIME FORECASTING - THE BROSNA CATCHMENT

The Brosna Catchment covers some 1,180 km^2 of mostly flat grazing land in the center of Ireland, and drains into the Shannon. Glacial movements have produced some hills, and there are also some peatbogs and woodland areas. The geology consists of carboniferous limestone covered by boulder clay. Rainfall is uniformly distributed throughout the year and is of low average intensity.

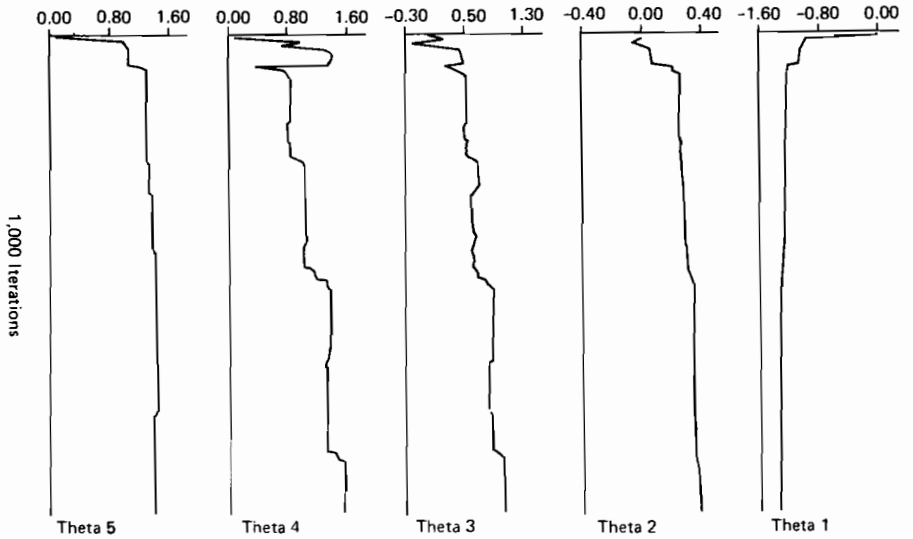


Fig. 9.3 Variation of θ with α_0 = 1.0 and $\alpha(0)$ = 1.0.
$\theta = [\alpha_1\ \alpha_2\ \beta_1\ \beta_2\ \varepsilon]$.

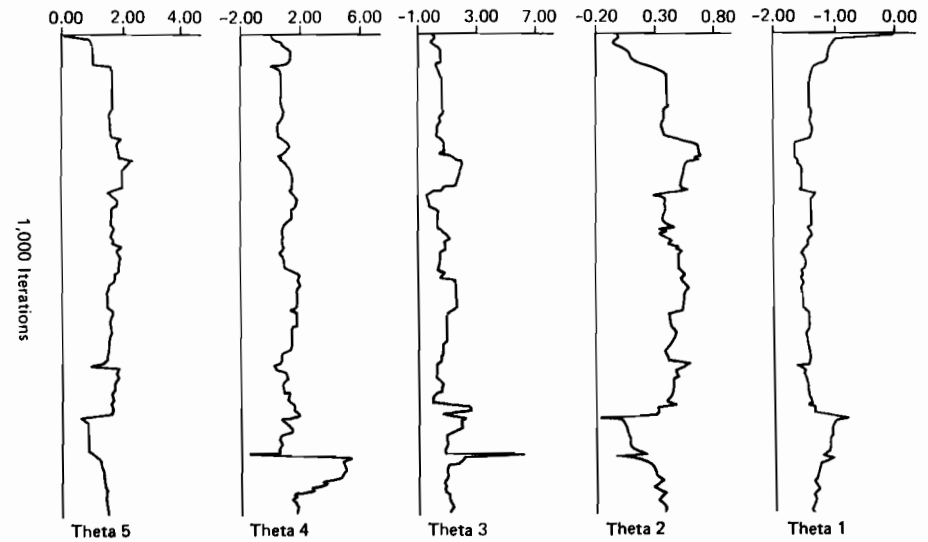


Fig. 9.4 Variation of θ with α_0 = 1.0 and $\alpha(0)$ = 0.98.
$\theta = [\alpha_1\ \alpha_2\ \beta_1\ \beta_2\ \varepsilon]$.

The data have been compiled from 5 years of discharge measurements at Ferbane and a centrally sited, continuously recording raingauge. Areal mean rainfall was computed using 18 raingauges distributed throughout the catchment area; data were evaluated every 3 hours with reference to the central recording gauge.

The results obtained for simulated real-time forecasts are shown in Figs. 9.5 and 9.6. The rainfall measurements shown are the average intensities over the 3-hr

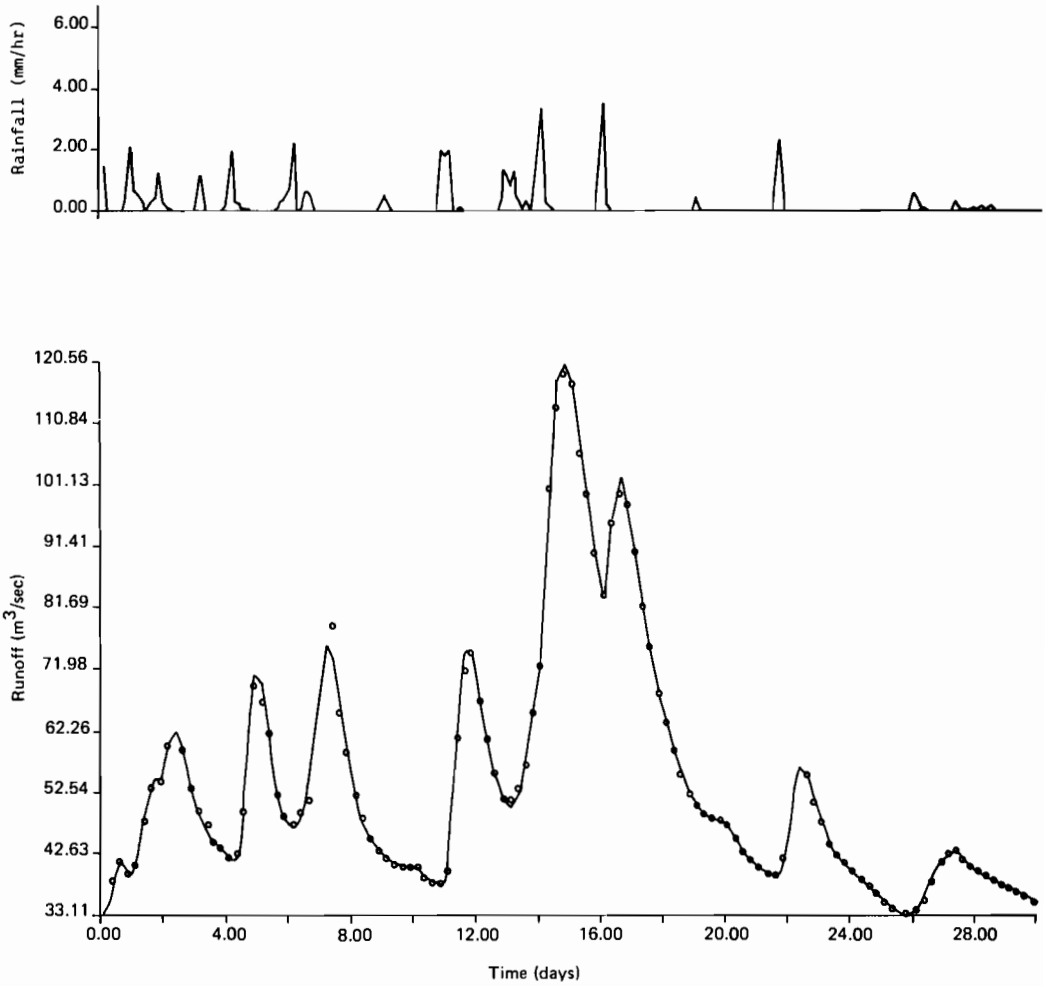


Fig. 9.5 One-step-ahead self-tuning predictor for the Brosna Catchment with m = 2, n = 2, and ℓ= 2. ——— observed; o 3-hr forecast.

sampling intervals. Fig. 9.5 shows a one-step-ahead forecast, and Fig. 9.6 shows a two-step-ahead forecast for the same data. The solid line is the measured streamflow and the points are the forecast streamflows.

It has been found that if the sampling rate is high compared to the process dynamics, as in Fig. 9.5, then it is generally better to use the one-step ahead predictor and sample the data at a larger time step than to use a k>1-step ahead predictor with shorter time steps.

CONCLUSIONS

A minimum square error predictor is postulated as a streamflow forecasting model because the rainfall-runoff process can be represented by a linear time-invariant stochastic model. The parameters of this predictor can be calculated from the

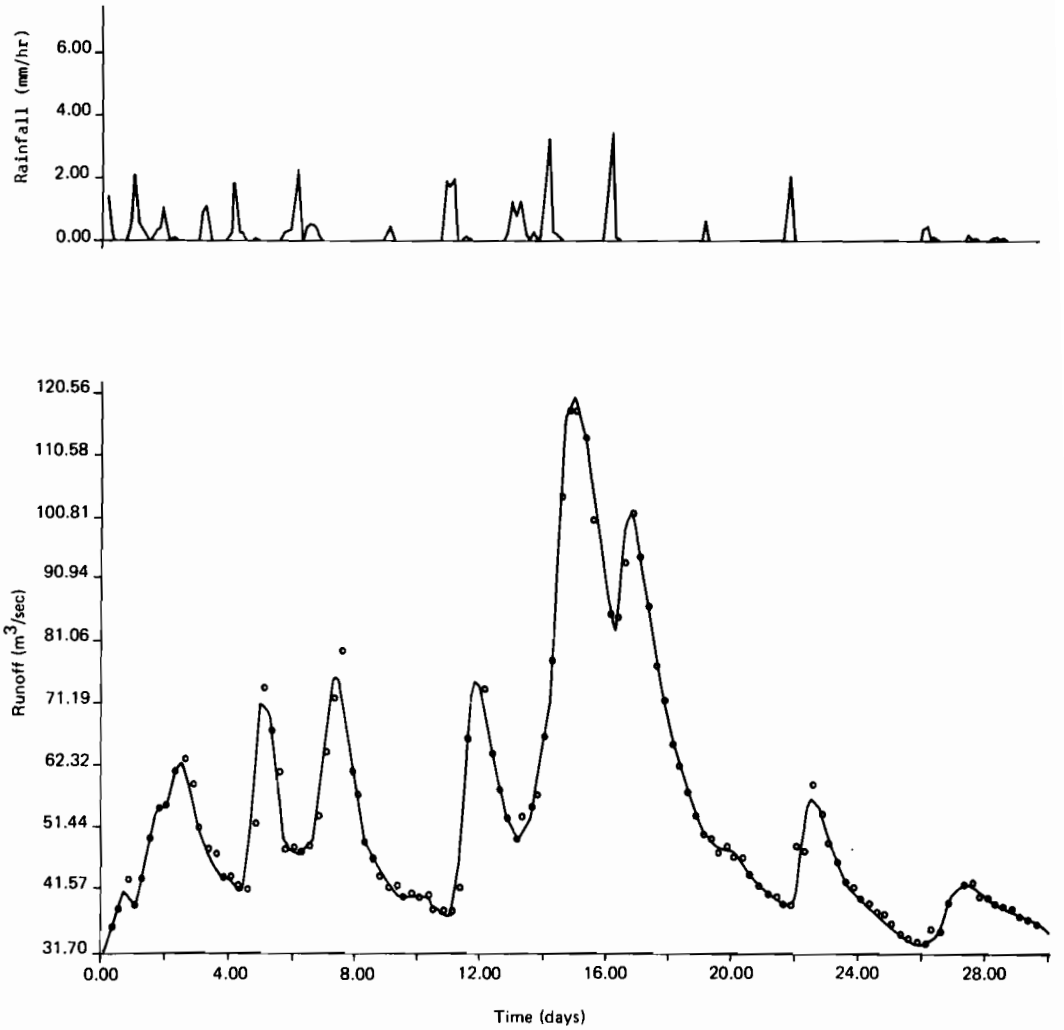


Fig. 9.6 Two-step-ahead self-tuning predictor for Brosna Catchment with m = 3, n = 2, ℓ = 2 and input delay = 1 (3 hr). ——— observed; o 6-hr forecast.

parameters of the stochastic model. The self-tuning predictor however bypasses the system identification problem because the predictor parameters are directly estimated. Also, the recursive least squares technique used results in an algorithm suitable for minicomputer or microprocessor implementation.

The predictor has been tested on several sets of data with good results, despite the incorrect linearity assumption. Basically, this problem is lessened because the on-line streamflow measurements prevent cumulative errors from building up. Clearly, with such a simple model there is possibility for improvements; the most obvious is to replace the total rainfall input by an adjusted rainfall input which takes into account the soil moisture deficit and evaporation effects.

REFERENCES

Aström, K.J. (1970). Introduction to Stochastic Control Theory. Academic Press, New York.

Aström , K.J. (1974). A self-tuning parameter estimator. Publ. No. 74/55, Dept. of Computing and Control, Imperial College of Science and Technology, London.

Aström, K.J. and B. Wittenmark (1973). On self-tuning regulators. Automatica 9: 185-199.

Wittenmark, B. (1974). A self-tuning predictor. IEEE Trans. Autom. Control 19(6): 848-851.

10 Nonlinear Forecasting Models of Water Flow and Quality by Heuristic Self-Organization Methods

Saburo Ikeda and Yoshikazu Sawaragi

Because pollution and eutrophication phenomena of surface water and groundwater have increased in urbanized areas, much attention has been given to research related to modeling water quality in rivers, lakes, and coastal areas, for the purpose of prediction and simulation. There are many factors that determine water quality; they may be socioeconomic, topographical, or hydrologic factors. For example, the combination of low flow, high water temperature, and an increase of pollutants results in poor water quality conditions. Needless to say, it is a basic problem to forecast water discharges from the available data on discharge and rainfall and to use the data in models of water quality.

Most of the dynamic models for river discharge prediction are based on hydrological and geographical characteristics of river basins (Linsley et al. 1949, Sugawara 1961, Eagleson 1970). However, only with many assumptions and simplifications of the real phenomena is it possible to formulate a physical or mathematical model. Therefore, these hydrologic models based on physical considerations may be suitable for a quantitive analysis of runoff mechanisms, but their applicability to exact quantitative forecasting may be rather limited because of the uncertainty of hydrologic variables such as areal precipitation, evapotranspiration, and underground storage. The modeling of water quality is far more difficult than the forecast modeling of runoff, since it is harder to find practical applications of the elaborate models. In fact, it is difficult to find the exact mass balance of pollutants, inflows and outflows of water in agricultural irrigation, municipal use, branchstreams, and so on. Even if these difficulties were overcome, there is still the more essential problem of a lack of sufficient water quality data to test the models because it is very laborious and expensive to measure and analyze water quality.

Another approach is based on time series analysis of runoff data; that is, the runoff model is identified from the measured flow and rainfall data by various statistical methods such as correlation analysis, maximum likelihood techniques, and least squares (Hino 1970, Kashyap and Rao 1973). However, the linear statistical models have some difficulties in treating the nonlinearity of runoff mechanisms, and lack sufficient length of input-output data on water quality. One of the general methods of identifying nonlinear systems uses the Volterra series (Eykhoff 1974). The use of the Volterra series, however, requires a large amount of data and computation of a high-dimensional matrix to determine the structure of the multivariable, nonlinear system.

The urgent need for water quality prediction and simulation for environmental

management will require development of these and related techniques. Prediction models have to have the following characteristics to be used in water resource systems:

- Simplicity in structure (without special or sophisticated knowledge of hydrology or geography of a basin),
- A small amount of computation time (well suited for real time forecasting),
- Applicability to the small amount of available data.

In this respect, a self-organization method for modeling or prediction of complex nonlinear systems, the Group Method of Data Handling (GMDH), has recently been developed by Ivakhnenko (1968, 1970, 1971). This algorithm is very useful for modeling complex nonlinear systems that have a great number of variables and parameters, structural uncertainties and a limited amount of collected data. We have already proposed two heuristic forecasting algorithms for the modeling of nonlinear complex input-output relations, which are improved versions of the original GMDH (Ikeda et al. 1976a, 1976b).

In this paper, a summary of our work and an application of the modified method to water quality modeling are presented in order to promote broad use of the heuristic self-organization methods. To show the effectiveness of the proposed method, numerical comparisons are made between the performance of the prediction models presented here and hydrologic models often found in the literature.

HEURISTIC SELF-ORGANIZATION MODELING (GMDH)

To simplify the illustration of GMDH, we shall first consider the following forecasting model of water flow;

$$q(t) = f(r(t - h), r(t - 2h), \ldots, r(t - mh), q(t - h), \ldots, q(t - nh)) \quad (10.1)$$

where $q(t)$ is an averaged daily flow (m^3/sec), $r(t)$ is daily rainfall (mm/day) in the river basin, and h is the sampling interval, assumed to be one day. The integers m and n represent the number of days included in the prediction (10.1), that is, the interval prior to the sample.

The problem is to determine the unknown nonlinear structure $f(r(t - h), \ldots, q(t - h), \ldots)$ from the available past data of rainfall and discharge. For operational use, a kind of cybernetic method, that is, a heuristic self-organization technique, is applied here instead of the conventional approaches.

The Basic GMDH Algorithm

Equation (10.1) can be rewritten more briefly as:

$$y(t) = f(x_1, x_2, \ldots, x_{N_o}) , \quad (10.2)$$

where $y(t) = q(t)$ is an output of interest and x_i ($i = 1, \ldots, N_o$) are inputs such as $x_1 = r(t - h)$, $x_{m+1} = q(t - h)$, and so on. Here, let us assume that the nonlinear input-output relation $f(x_1, \ldots, x_{N_o})$ is represented by a polynomial of a certain order with respect to x_i to formulate a universal nonlinear model.

The Kolmogorov-Gabor polynomial or the Volterra series for the stationary stochastic process (Gabor et al. 1961),

$$y = a_0 + \sum_i a_i x_i + \sum_i \sum_j a_{ij} x_i x_j + \sum_i \sum_j \sum_k a_{ijk} x_i x_j x_k + \cdots \qquad (10.3)$$

provides a conceptual basis for (10.2). However, a dimensionality problem is caused by the tremendous number of coefficients to be determined in (10.3). To overcome this difficulty, Ivakhnenko used the multilayered structure shown in Fig. 10.1. This structure is similar to that of the natural selection rule of plants

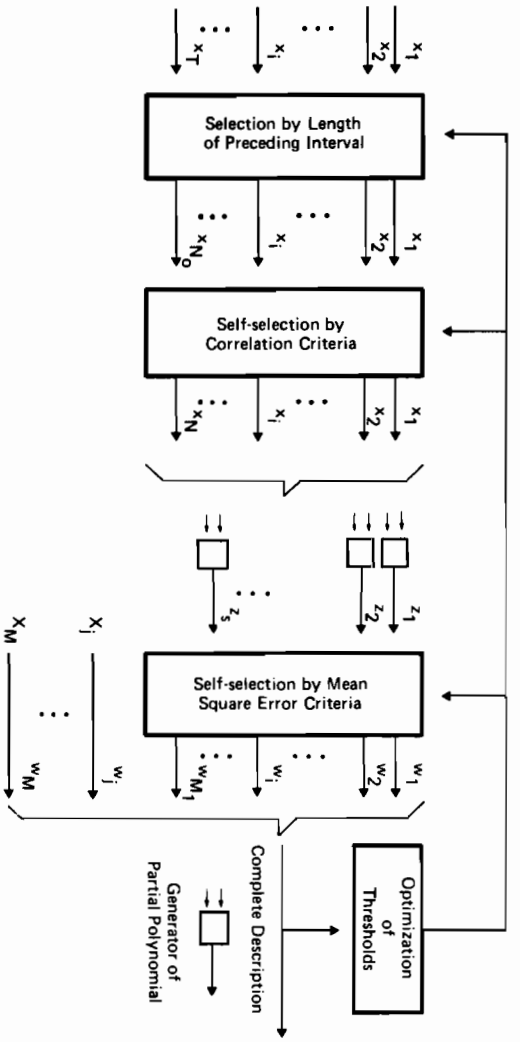


Fig. 10.1 Ivakhnenko's schematic structure of the GMDH algorithm.

or animals. The inputs from the previous layer are used for constructing all possible combinations of two inputs like the "crossing of seed". The better results are chosen on a heuristic basis or by threshold criteria, and then one proceeds to the next layer.

To avoid overfitting of the prediction, the concept of "regularization" in numerical analysis (Tikhonov 1963) is introduced. The available input-output data are divided into the "training" sequence for learning the system structure and the "checking" sequence for selecting the best results. This regularization technique is one of the most important features in the GMDH algorithm.

The procedure of the basic GMDH algorithm originally developed by Ivakhnenko can be summarized as follows:

Step 1: Determine the prior interval (m,n) and the corresponding input variables

Step 2: Choose N "useful" input variables x_i (i = 1,2,...,N), which are highly correlated with the output y, from among all input variables that belong to the prior intervals. For example, if m and n are picked to be two days, that is, m = n = 2 and h =1, then $x_1 = r(t - 1)$, $x_2 = r(t - 2)$, $x_3 = q(t - 1)$ and $x_4 = q(t - 2)$.

Step 3: Divide the original data into the training and checking sequences. The separation rule is a heuristic one. Usually, the training and checking sequences are taken alternately or on the basis of the magnitude of the variance from the mean value.

Step 4: Generate the following partial polynomials with all possible combinations of two inputs x_i and x_j:

$$z_k = a_{k0} + a_{k1}x_i + a_{k2}x_j + a_{k3}(x_i)^2 + a_{k4}(x_j)^2 + a_{k5}x_ix_j \ , \qquad (10.4)$$

$$i,j = 1,2,...,N,\ i \neq j;\quad k = 1,2...,N\cdot(N - 1)/2 \ .$$

The coefficients $a_{k\ell}(\ell = 0,1,...,5)$ are determined so as to minimize the mean square error:

$$\varepsilon_k^2 = \overline{(y - z_k)^2} \qquad (10.5)$$

where the bar denotes the sample mean.

Step 5: Compute the values of intermediate variables z_k (for the checking data) by (10.4) with the coefficients as determined in Step 4. Then select M intermediate variables $z_i(i = 1,...,M)$ which gives M smallest mean square errors (10.5) for the checking data.

Step 6: Repeat Step 4 and Step 5 but replace x_i with z_i and x_j with z_j until the mean square error, ε_m^2, of the best predictor, z_m, in the present layer exceeds that of the best one in the previous layer.

In addition to the above procedure, the selection rule of the intermediate variables and the separation rule for the training and checking sequences are changed so as to find the optimal prediction model, as shown in Fig. 10.1.

Application of the Basic GMDH Algorithm to River Flow Prediction

In this section, the basic GMDH algorithm is applied to modeling one-day-ahead river-flow prediction. Because it has been studied before, and for the convenience of comparisons with other models, the Karasu River basin in Japan was used

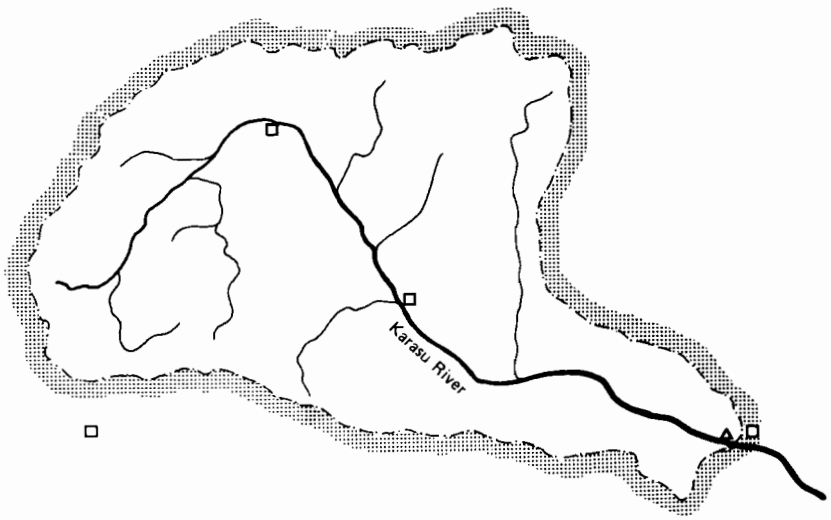


Fig. 10.2 The Karasu River basin. (Area = 156km^2)
□ denotes a precipitation station;
△ denotes a riverflow station. Reprinted, by permission, from Ikeda et al. (1976a).

(Tone-River-Dams Control Office 1967). The Karasu River basin is shown in Fig. 10.2. A station for the river flow data is denoted by a triangle and the precipitation stations by squares. The daily river flow q(t) (m^3/sec) and the daily mean areal precipitation r(t) (mm/day) are from the Tone-River-Dams Control Office report (1967). Figure 10.3 shows the autocorrelation function $\rho_q(t)$ of the river flow q(t) and the cross-correlation function $\rho_{rq}(t)$ between the river flow q(t) and the

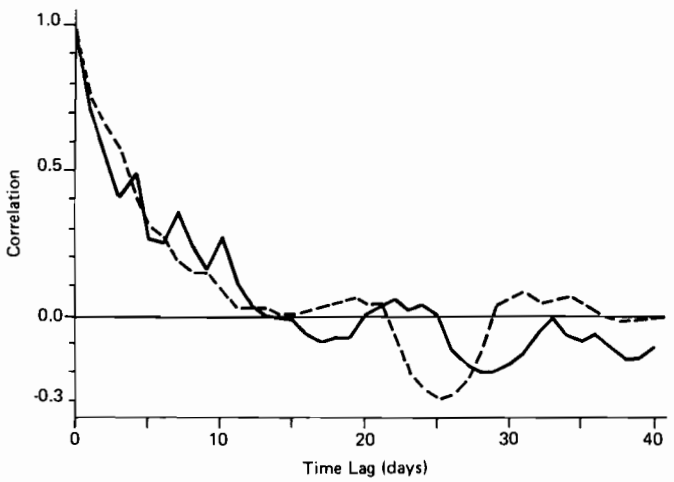


Fig. 10.3 The autocorrelation function $\rho_q(t)$ and the cross-correlation function $\rho_{rq}(t)$. ---- = $\rho_q(t)$; —— = $\rho_{rq}(t)$. Reprinted, by permission, from Ikeda et al. (1976a).

areal mean precipitation r(t) calculated for the data for the period 1 May to 31 August 1964.

$$\rho_{q}(t) = \frac{1}{123 - t} \sum_{s=1}^{123-t} q(s)q(s + t)$$

$$\rho_{rq}(t) = \frac{1}{123 - t} \sum_{s=1}^{123-t} r(s)q(s + t)$$

From Fig. 10.3, the length of the prior interval to be included in the model was chosen to be 10 days. First, the original input variables were selected from data from up to 10 days before the study:

$$\{r(t),r(t - 1),\ldots,r(t - 9);q(t),q(t - 1),\ldots,q(t - 9)\}$$

and the output variable is the one-day-ahead river flow q(t + 1). Next, with the use of the data from 1 May to 31 August 1964, the best forecasting model was found following the procedure from Step 3 to Step 6 in the basic GMDH algorithm.

From this, we obtained an input-output relation of a simple polynomial of second order given by

$$\begin{aligned} q(t + 1) = &- 0.2001 + 0.9846q(t) + 0.2245r(t) - 0.0180q^{2}(t) \\ &+ 0.004r^{2}(t) - 0.0048q(t)r(t) . \end{aligned} \tag{10.6}$$

Using the prediction model (10.6), the one-day-ahead prediction is carried out for the period 1 May to 30 June of the next year (1965). The tracking behavior of the predicted values $q_1(t)$ is shown in Fig. 10.4. The root-mean-square error for

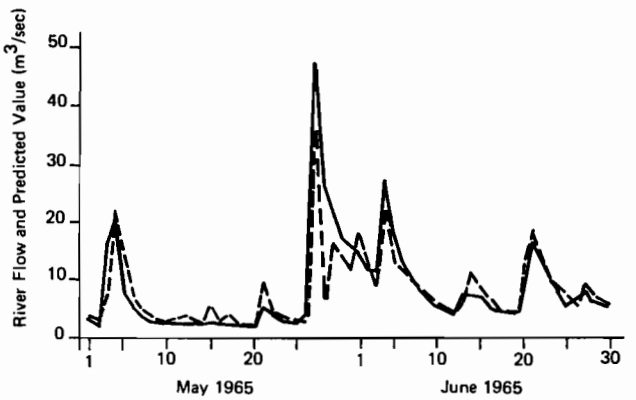


Fig. 10.4 The tracking behavior of predicted value $q_1(t)$ together with the actual flow q(t). —— = q(t); --- = $q_1(t)$. Reprinted, by permission, from Ikeda et al. (1976a).

that period becomes:

$$\sigma_1 = \frac{1}{60} \sum_{t=\text{May } 1}^{\text{June } 30} [q(t) - q_1(t)]^2 \Big]^{1/2} = 3.87 \ . \qquad (10.7)$$

It is surprising that the simple nonlinear prediction model (10.6) gives good estimates. However, it underestimated some peak points from 27 May to 4 June 1965.

MODIFIED GMDH ALGORITHMS

In order to improve the prediction of high river flows, two self-organized identification methods were used, which are modified versions of the basic GMDH algorithm as outlined in the previous section. The main improvements are in the selection rule for input variables and the generation rule for intermediate variables.

Procedure of the Modified GMDH Algorithms

Step 1: Select input variables.

Step 2: Divide the original data into training and checking data.

(These two steps are the same as those of the basic GMDH algorithm.)

Step 3: Suppose that we have the intermediate variables $z_i^{(k)}$, $(i = 1,2,\ldots M)$ for layer k. For $k = 1$ we set $z_i^{(1)} = x_i, (i = 1,2,\ldots,M)$. We determine the intermediate variables for layer $(k + 1)$ according to procedure A or B:

(A) (i) Transform the intermediate variables $z_i^{(k)}$ by the second-order polynomial:

$$\hat{z}_i^{(k)} = a_{i0} + a_{i1} z_i^{(k)} + a_{i2} (z_i^{(k)})^2 \ , \ i = 1,2\ldots M, \qquad (10.8)$$

where the coefficients a_{ij}, $(j = 0,1,2)$, are determined by using the training data so as to minimize the mean square error

$$\overline{(y - \hat{z}_i^{(k)})^2}$$

between the output y and the transformed variables $\hat{z}_i^{(k)}$. The upper bar denotes sample mean. The transformed variables $\hat{z}_i^{(k)}$, for checking data, should also be generated by (10.8).

(ii) Combine the transformed intermediate variables $\hat{z}_i^{(k)}$ and $\hat{z}_j^{(k)}$ generated in (10.8) by the partial polynomial:

$$\begin{aligned} z_\ell^{(k+1)} &= b_{\ell 0} + b_{\ell 1} \hat{z}_i^{(k)} + b_{\ell 2} \hat{z}_j^{(k)} + b_{\ell 3} (\hat{z}_i^{(k)})^2 \\ &\quad + b_{\ell 4} (\hat{z}_j^{(k)})^2 + b_{\ell 5} \hat{z}_i^{(k)} \hat{z}_j^{(k)} \ , \\ &\quad i,j = 1,2,\ldots,M \ , \ i \neq j \ , \\ &\quad \ell = 1,2\ldots,M(M-1)/2 \ , \end{aligned} \qquad (10.9)$$

where the coefficients $b_{\ell j}$ ($j = 0,1,\ldots 5$) are determined by using training data so as to minimize the mean square error,

$$\varepsilon_\ell^2 = \overline{(y - z_\ell^{(k+1)})^2} \quad , \qquad \ell = 1,2,\ldots,M(M-1)/2,$$

between the output y and the intermediate variables $z_\ell^{(k+1)}$. The same transformation (10.9) should be applied to checking data.

(B) Generate the following partial polynomials with all possible combinations of $z_i^{(k)}$ and $z_j^{(k)}$:

$$z_\ell^{(k+1)} = c_{\ell 0} + c_{\ell 1} z_i^{(k)} + c_{\ell 2} z_j^{(k)} + c_{\ell 3} z_i^{(k)} z_j^{(k)} \quad , \tag{10.10}$$

$$i,j = 1,2,\ldots,M \quad ,$$

$$\ell = 1,2,\ldots,M(M+1)/2 \quad .$$

When i = j, we use the following polynomial:

$$z_\ell^{(k+1)} = d_{\ell 0} + d_{\ell 1} z_i^{(k)} + d_{\ell 2} (z_i^{(k)})^2 \quad , \tag{10.11}$$

where the coefficients $c_{\ell j}$, ($j = 0,1,2,3$) and $d_{\ell j}$ ($j = 0,1,2$) are determined by using training data so as to minimize the mean square error ε_ℓ^2 between the output y and the intermediate variables $z_\ell^{(k+1)}$. We thus have the intermediate variables $z_\ell^{(k+1)}$ for layer (k+1).

Step 4: Select intermediate variables (say, $z_\ell^{(k+1)}$, $\ell = 1,2,\ldots,M_1$) which give M_1 smallest mean square errors ε_ℓ^2 for the checking data. In the case of (A), we set $M = M_1$. In the case of B, we add M_2 original input variables and set $M = M_1 + M_2$ (See Fig. 10.1).

Step 5: Replace k by k + 1 and go to Step 3. Repeat from Step 3 to Step 5 until the mean square error ε_m^2 of the best predictor z_m in the present layer exceeds that of the best one in the previous layer.

In addition to the above procedure, the selection rule of the intermediate variables and the separation rule of the training and checking sequences are changed so as to find the optimal prediction model in the same way as the basic GMDH.

It should be noted that the resultant complete description of the input-output relation in (A) is more complex than that of the basic GMDH for the same number of layers; that is, we can obtain a higher-order polynomial with a smaller number of layers. Meanwhile, in the case of (B), although the partial description (10.10) or (10.11) has only the simplest nonlinear term (cross products of each variable) we can expect to have a different type of nonlinear input-output relation, one which contains more features of the system in the form of variables that once might have been discarded at an earlier stage.

Adaptive Forecasting Models of River Flow

In addition to the modifications of the GMDH algorithm, an adaptive scheme of the forecasting model will be introduced to promote the applicability to river systems with limited input-output data. Let the original data from only the past 30 days be used on each day. The training data and checking data are taken alternately from the original data sequence. Since the content of the data sequence is renewed from day to day, the structure of model equation (10.1) is not always the same for each day. As a typical forecasting example, the above algorithm was applied to the modeling of the rainy season in Japan in June. Figure 10.5 shows the tracking behavior of the predicted values $q_2(t)$ and $q_3(t)$ of the river flow $q(t)$ for the period 1-30 June 1965, where

$q_2(t)$: the predicted value of the basic GMDH;

$q_3(t)$: the predicted value of the modified algorithm (A);

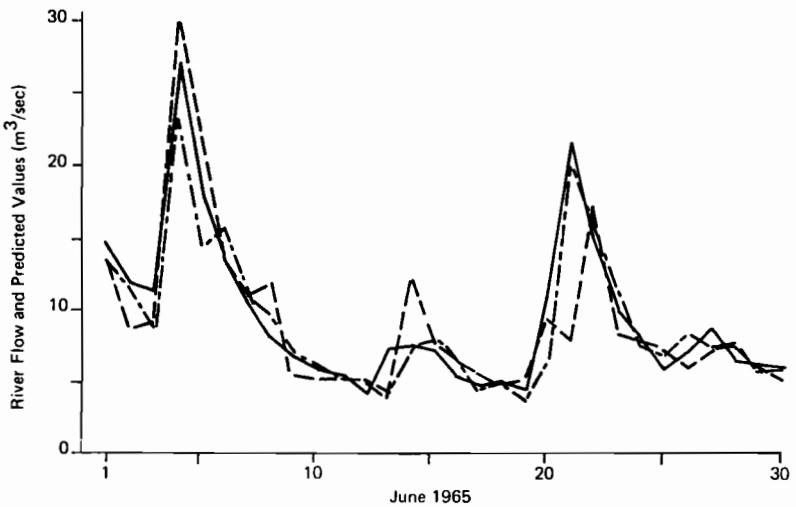


Fig. 10.5 The tracking behavior of predicted values $q_i(t)$ ($i = 2,3$) together with the actual flow $q(t)$. —— = $q(t)$; --- = $q_2(t)$; —·—·— = $q_3(t)$. Reprinted, by permission, from Ikeda et al. (1976a).

In these calculations, the number of "useful" input variables and the number of intermediate variables were optimally chosen to be ten and five for $q_3(t)$ and for $q_2(t)$ the optimal numbers were eight and five. In the modified algorithm (A), ten "useful" input variables were taken as

$$\{r(t), r(t-1), \ldots, r(t-5);\ q(t), q(t-1), \ldots, q(t-3)\},$$

and fixed throughout the experiment. As mentioned previously, the structure of the forecasting model for each day can vary. An example of the structure of the forecasting model, the final polynomial for $q_3(t)$ on 5 June 1965, is given in Table 10.1.

Table 10.1. Partial polynomials for $q_3(t)$ on 5 June 1965

Input variables	$x_1 = q(t-1),\quad x_2 = r(t-1),\quad x_3 = r(t-2)$
First Layer	$w_1^{(1)} = -0.7113 + 1.1809x_1 + 0.0148x_2 - 0.0220(x_1)^2 + 0.0024(x_2)^2 + 0.0101x_1x_2$
	$w_2^{(1)} = 2.0088 + 0.4540x_2 + 0.0706x_3 - 0.00065(x_2)^2 + 0.00055(x_3)^2 - 0.0103x_2x_3$
	$\hat{w}_1^{(1)} = -0.00008 + 1.0000w_1^{(1)} - 4.0 \times 10^{-7}(w_1^{(1)})^2$
	$\hat{w}_2^{(1)} = 0.0380 + 0.9922w_2^{(1)} + 0.00017(w_2^{(1)})^2$
Second Layer	$w_1^{(2)} = -0.3488 + 0.8171\hat{w}_1^{(1)} + 0.2833\hat{w}_2^{(1)} - 0.0084(\hat{w}_1^{(1)})^2$ $- 0.0240(\hat{w}_2^{(1)})^2 + 0.0304\hat{w}_1^{(1)}\hat{w}_2^{(1)}$
Predicted Value	$q_3^{(t)} = w_1^{(2)}$

In this case, the final description (the forecasting model for 5 June 1965) has passed only two layers, but it contains various orders of polynomials up to the sixth order. We can easily see from Fig. 10.5 that a considerable improvement is attained for predicting peak values by the modified algorithm A.

In fact, the root-mean-square errors

$$\sigma_i = \frac{1}{30} \sum_{t=\text{June } 1}^{\text{June } 30} [q(t) - q_i(t)]^2 \quad , \quad i = 2,3$$

are given as follows:

$$\sigma_2 = 3.27 \quad , \quad \sigma_3 = 1.86 \tag{10.12}$$

For comparing the performance of our model with hydrologic models, the tracking behavior of the predicted values $q_3(t)$, $q_4(t)$ and $q_5(t)$ of the river flow $q(t)$ is shown in Fig. 10.6, where:

$q_4(t)$: the predicted value of the "tank model" of Sugawara (1961);

$q_5(t)$: the predicted value of the "storage function method" (Linsley et al. 1949).

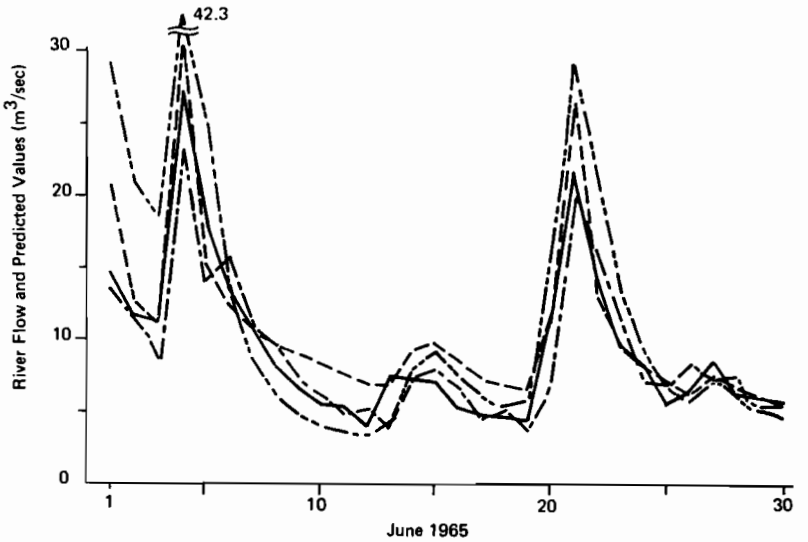


Fig. 10.6 The tracking behavior of predicted values $q_4(t)$ and $q_5(t)$ from hydrologic models together with $q_3(t)$ and the actual flow $q(t)$. —— = $q(t)$; - — - = $q_3(t)$; --- = $q_4(t)$; — -- — = $q_5(t)$. Reprinted, by permission, from Ikeda et al. (1976a).

Here, the predicted values $q_4(t)$ and $q_5(t)$ are taken from the Tone-River-Dams Control Office report (1967). The root-mean-square errors of hydrologic models are given as follows:

$$\sigma_4 = 2.23 \quad , \qquad \sigma_5 = 5.19 \tag{10.13}$$

We can see from (10.12) and (10.13) that the predicted values by the self-organization methods such as the basic GMDH and the modified algorithms are as accurate, or more accurate than those of the elaborate hydrologic models, such as the "tank model" or "storage function method" models. The modified algorithm (A) shows the best tracking behavior for sudden changes of the river flow.

Application to the Water Quality Forecasting Model

The modified GMDH algorithm has also been applied to forecasting models of water quality of the Tama River basin located in the Tokyo Metropolitan area (Ichikawa and Ikeda 1976). The river basin, shown schematically in Fig. 10.7, is about

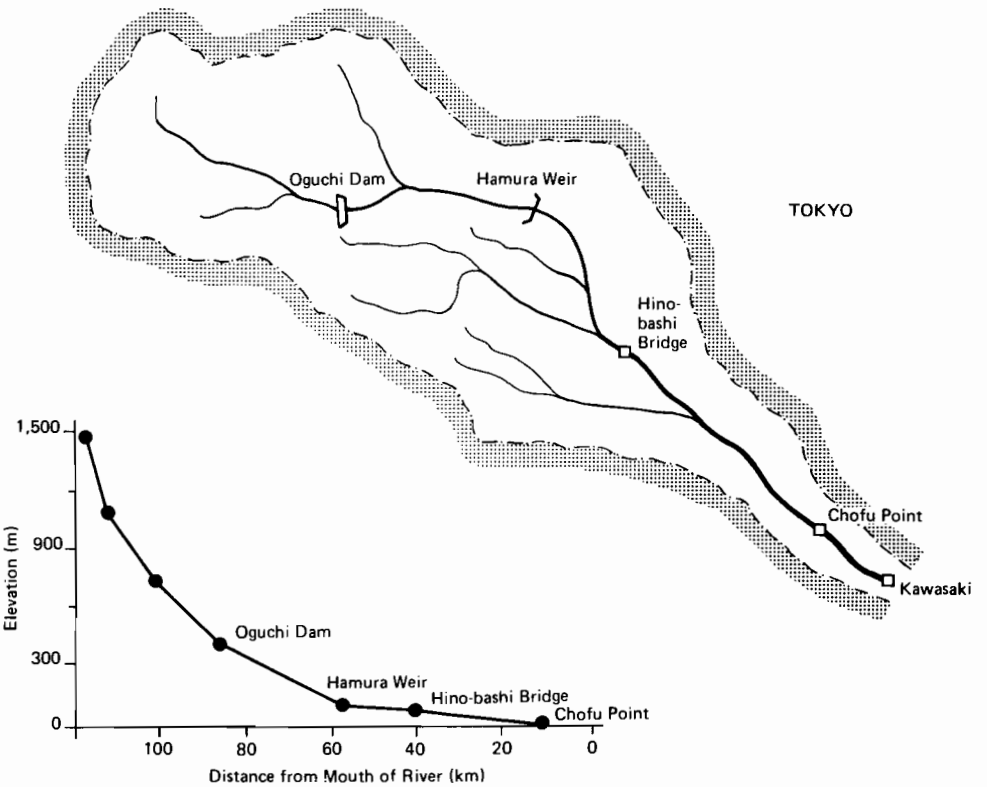


Fig. 10.7 The Tama River basin (Area: 1,200 km^2).

1,200km^2 in area and has several observation points. Water quality data used here was gathered from the Tokyo Metropolitan Government, the Ministry of Construction, and other organizations responsible for water quality. Therefore, it is inevitable that the data has a lot of errors.

Since each item of water quality, such as turbidity, chemical oxygen demand (COD), biological oxygen demand (BOD), NH_4-N, pH, and conductivity, depends on various elements (flow rate, temperature, population, and so on) and has interrelations with others, forecasting models of water quality must be identified with consideration of the nonlinear interactions between the variables.

First, time series data of each item were treated by standard statistical analysis;

that is, several statistical measures were computed (average, variance, correlation, power spectrum, correlogram, and so on). Next, from these statistical data and from physical considerations, we selected some useful input variables related to the water quality item of interest. Then, one-day-ahead optimal forecasting models were selected as multivariable nonlinear regression equations described by some items of water quality such as NH_4-N, COD, conductivity and flow rate, together with their delayed arguments.

The design parameters (heuristics) used in this modeling are as follows:

(1) Number of input factors: 4 elements of water quality and a prior interval of 5 days.

(2) Length of data sequence to be taken in the modeling: 30 to 50 days.

(3) Five to seven days taken in each layer.

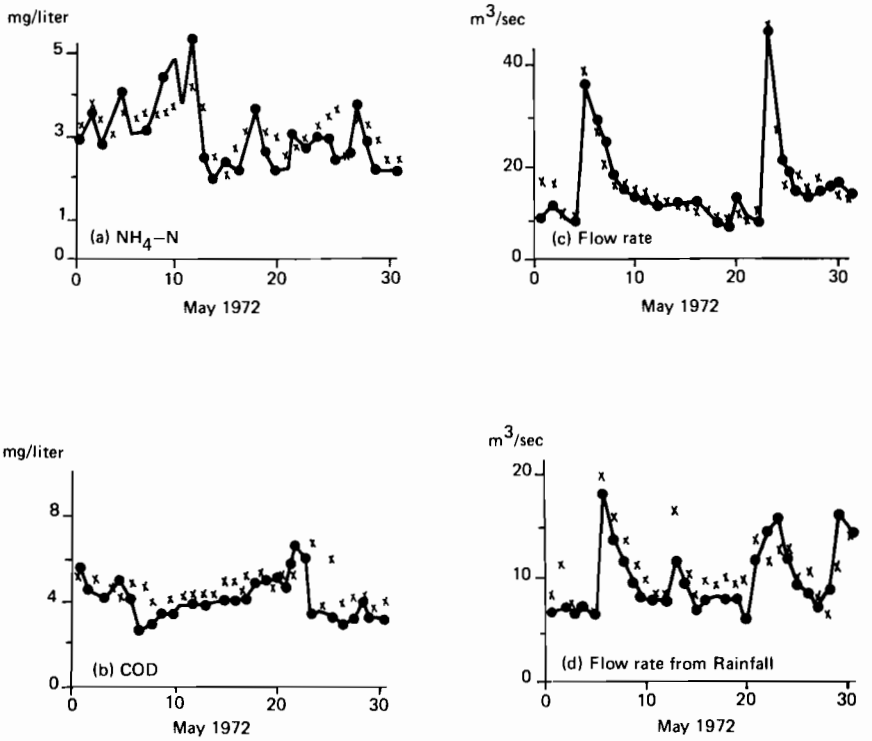


Fig. 10.8 One-day-ahead forecasts of water quality by the modified GMDH. (a) NH_4-N concentration; (b) COD concentration; (c) and (d) flow rates. ● - observed value; x - predicted value.

Figure 10.8 shows some examples of the prediction for NH_4-N, COD, and flow rate. The variables contained in these models are as follows:

(a) NH_4-N: COD, conductivity, NH_4-N, flow rate at Chofu;

(b) COD: COD, conductivity, NH_4-N, flow rate at Chofu;

(c) Flow rate: Flow rates of several upper points;

(d) Flow rate: Flow rate at Chofu and rainfall (averaged).

As far as water quality data are concerned, the number of facts were limited because of a lack of data, especially of consecutive daily data. Therefore, in selecting input factors, it is natural to use some heuristic considerations. For example, in Fig. 10.8c, which shows the good fit of the forecast, it is evident that flow rates of upper streams play essential and deterministic roles. Meanwhile, for a water quality index such as NH_4-N, COD, shown in Fig. 10.8a and b, it is very difficult to have an exact forecasting model owing to the use of daily instantaneous sampling data, although rough daily estimations were obtained. Figure 10.8d is the forecast of the runoff at Chofu from the past rainfalls and water levels. Better forecasts could be obtained if input factors were based on averaged rainfall of the upper, middle, and lower stream basins and not only on the averaged rainfall in the whole river basin.

One is able to say that the adaptive identification scheme of GMDH, described in this section, has a soft structure which supposes forecast models of water quality that can be adjusted to latest and limited numbers of available data. In particular, since the water quality varies with the seasons, and because of changes in artificial pollutants, this adaptive identification scheme can easily adjust the structure to use recent observation data. However, this algorithm sometimes has a certain kind of instability, that is, a tendency to overestimate peak points owing to the small amount of data. This sometimes yields a high-order polynomial which contains only a small number of variables. To handle this problem, a sequential GMDH will be given in the next section.

SEQUENTIAL GMDH ALGORITHM

In order to stabilize the structure of the prediction model and to shorten the computation time, a sequential formula of the GMDH was developed (Ikeda et al. 1976b). First, the optimal nonlinear structure of system (10.1) is determined by the use of the modified GMDH algorithms for the period of interest. Next, whenever new measurement data are obtained, the coefficients of partial descriptions (10.8)-(10.11) are changed sequentially according to the following procedure.

Sequential Procedure

Consider the following equation for each partial description:

$$A_p X = Y_p , \tag{10.14}$$

where A_p is a p x 1 matrix, X is a q x 1 coefficient vector, and Y_p is a p x 1 output vector. The integer q is equal to 3 for (10.8), 6 for (10.9), 4 for (10.10), and 3 for (10.11).

When A_p has a maximum rank, the least-squares solution is given by:

$$X^* = (A^T A_p)^{-1} A_p^T Y_p, \tag{10.15}$$

where $(A)^{-1}$ denotes the inverse of matrix A and A^T denotes the transpose of matrix A.

Now, suppose we have an additional equation:

$$B^T X = z_{p+1} \tag{10.16}$$

where B is a q x 1 vector obtained by a new measurement and z_{p+1} is the latest output. Then, combining (10.14) and (10.16), we have

$$A_{p+1}\, X = Y_{p+1}\ , \tag{10.17}$$

where

$$A_{p+1} \triangleq \begin{bmatrix} A_p \\ B^T \end{bmatrix}\ , \qquad Y_{p+1} \triangleq \begin{bmatrix} Y_p \\ z_{p+1} \end{bmatrix}\ .$$

For (10.17), the least-squares solution for (10.15) becomes:

$$x_{p+1}^* = (A^T_{p+1}\, A_{p+1})^{-1}\, A^T_{p+1}\, y_{p+1}\ . \tag{10.18}$$

Now, let us establish a recursive formula to simplify the computation of the optimal solution X_{p+1}^* with new measurement data. If we define

$$P_p = (A^T_p A_p)^{-1}\ , \tag{10.19}$$

then

$$P_{p+1}^{-1} = A^T_{p+1}\, A_{p+1} = A^T_p\, A_p + BB^T = P_p^{-1} + BB^T. \tag{10.20}$$

A matrix inversion lemma for (10.19) and (10.20) follows Sage (1968). If matrices P_{n+1}, P_n, H_{n+1}, and R_{n+1} satisfy the equation:

$$P_{n+1}^{-1} = P_n^{-1} + H^T_{n+1}\, R_{n+1}^{-1}\, H_{n+1}\ ,$$

and P_{n+1}, P_n^{-1}, and R_{n+1}^{-1} are nonsingular, and H_{n+1} is of maximum rank, then P_{n+1} is given by

$$P_{n+1} = P_n - P_n\, H^T_{n+1}\, (H_{n+1}\, P_n\, H^T_{n+1} + R_{n+1})^{-1}\, H_{n+1}\, P_n\ . \tag{10.21}$$

Here we set $H^T_{n+1} = B$ and $R_{n+1} = I_q$ (a q x q identity matrix), then from (10.21) we can obtain the following equation:

$$P_{p+1} = P_p - P_p\, B(B^T P_p\, B + 1)^{-1}\, B^T P_p\ . \tag{10.22}$$

Substituting (10.22) into the optimal solution (10.18), we obtain the recursive

equation:

$$X_{p+1}^* = P_{p+1}\,(A^T{}_p\, Y_p + B\, z_{p+1})$$

$$= X_p^* + P_p\, B(B^T P_p\, B + 1)^{-1}(z_{p+1} - B^T X_p^*)\ . \qquad (10.23)$$

The initial values X_0 and P_0 are given by the modified GMDH. Thus, we compute only the inverse of a scalar $(B^T P_p\, B + 1)$, instead of computing the inverse of a (p+1) x (p+1) matrix in (10.18). This procedure is repeated for the partial descriptions in each layer.

This algorithm is effective when the input-output relation of the system changes gradually, owing to the nonstationary elements or some other unknown factors. We change only the parameters of the obtained structure gradually, keeping a stable skeleton of the system's nonlinear structure. Computation time is smaller in comparison with that spent by the basic GMDH algorithm, owing to the sequential nature of the algorithm.

A Sequential Forecasting Model of River Flow

First, as a stable structure of the initial forecasting model we employ the following input-output relation (10.6) which was determined by using the modified GMDH algorithm and data from 1 May to 31 August 1964:

$$q(t+1) = -0.2001 + 0.9846\, q(t) + 0.2245\, r(t) - 0.0180\, q^2(t)$$

$$+ 0.0041\, r^2(t) - 0.0048\, q(t)\, r(t)\ . \qquad (10.6)$$

Next, sequential formula (10.23) is applied to the one-day-ahead prediction for the period 1 June to 30 June 1965.

The tracking behaviors of the actual flow q(t) and the predicted values $q_1^*(t)$ and $q_2^*(t)$ are shown in Fig. 10.9, where

$q_1^*(t)$: the predicted value by the sequential algorithm (10.23)

$q_2^*(t)$: the predicted value by model (10.6).

The root-mean-square errors, as defined in (10.7)

$$\sigma_i^* = \frac{1}{30} \sum_{t=\text{June } 1}^{\text{June } 30} (q(t) - q_i^*(t))^2 \Bigg)^{1/2}, \quad i = 1, 2\ ,$$

are

$$\sigma_1^* = 1.84, \qquad \sigma_2^* = 2.07$$

CONCLUDING REMARKS

There is no methodology that covers all aspects of forecasting models in various situations for water management systems. The GMDH algorithms have both strengths

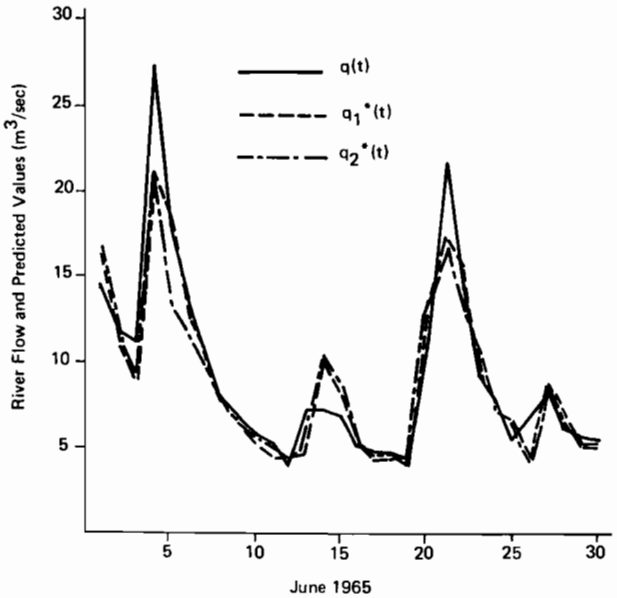


Fig. 10.9 Tracking behavior of predicted values $q_1^*(t)$ and $q_2^*(t)$ together with actual flow $q(t)$.

and weaknesses compared with other forecasting methods. Generally speaking, to apply hydrologic forecasting models to actual situations, adequate data, sophisticated computation, and a special knowledge of hydrology are required. On the other hand, our heuristic self-organized model is considered to have the following features: algorithmic simplicity (without any special knowledge of hydrology), short computation time, and applicability to a small amount of data. It is acknowledged that these models may not fully explain the physical characteristics of the runoff mechanism or transformation of water quality in streams.

The GMDH algorithm has great potential for absorbing heuristics from the specialists versed in hydrologic phenomena. It is therefore one of the strongest tools for modeling nonlinear complex systems for both specialists in hydrology, and non-specialists working in related areas.

ACKNOWLEDGEMENTS

The author would like to express his sincere gratitude to Professor A. Ichikawa of Tokyo University for his valuable suggestions, and special thanks should be extended to Dr. S. Fujishinge of Tokyo University and Ms. M. Ochiai of Kyoto University for their assistance in performing the research.

REFERENCES

Eagleson, P.S. (1970). Dynamic Hydrology. McGraw-Hill, New York.

Eykhoff, P. (1974). System Identification. Wiley-Interscience, London.

Gabor, D., W.P.L. Wilby, and R. Woodcock (1961). A universal nonlinear filter

predictor and simulator which optimizes itself by a learning process. IEE Proc. 108, Part B: 422-433.

Hino, M. (1970). Runoff forecasts by linear predictive filter. Proc. ASCE, J. Hyd. Div. 96 (HY3): 681-702.

Ichikawa, A., and S. Ikeda (1976). Prediction of water quality by heuristic self-organization. Proc. Japan Soc. Civ. Eng. 246 (2): 47-56 (in Japanese).

Ikeda, S., S. Fugishige, and Y. Sawaragi (1976a). Nonlinear prediction model of river flow by self-organization method. Int. J. Syst. Sci. 7 (2): 165-176.

Ikeda, S., M. Ochiai, and Y. Sawaragi (1976b). Sequential GMDH algorithm and its application to river flow prediction. IEEE Trans. Syst. Man. Cybern 6 (7): 473-479.

Ivakhnenko, A.G. (1968). The group method of data handling, A rival of the method of stochastic approximation. Sov. Autom. Control 13 (3): 43-55.

Ivakhnenko, A.G. (1970). Heuristic self-organization in problems of engineering cybernetics. Automatica 6: 207-219.

Ivakhnenko, A.G. (1971). Polynomial theory of complex system. IEEE Trans. Syst. Man Cybern 1: 364-378.

Kashyap, R.L., and A.R. Rao (1973). Real time recursive prediction of river flows. Automatica 9 (2): 175-183.

Linsley, R.K., M.A. Kohler, and J.L. Paulhus (1949). Applied Hydrology. McGraw-Hill, New York.

Sage, AP. (1968). Optimum Systems Control. Prentice-Hall, Englewood Cliffs, New Jersey.

Sugawara, M. (1961). On the analysis of runoff structure about seven Japanese rivers. Jpn. J. Geophys. 2 (4): 1-76.

Tikhonov, A.N. (1963). Solution of incorrectly formulated problems and the regularization method. Dokl. Akad. Nauk. SSSR, 151: 1035-1038.

Tone-River-Dams Control Office, Kanto Regional Construction Bureau, Japan (1967). Report on the Runoff Analysis of Low Flows in the Upper Reaches of the Tone River (in Japanese).

Part Two

Control of Water Resource Systems

Introduction

The development of water resource projects requires the formulation of strategies, or operating policies, for their operation. Most operating policies are developed using long time horizons and the expected benefits of the system are assessed over the life of the system. Nevertheless, it is only when the project is operating on a day-to-day basis that its benefits are realized. Essentially, the optimal control of a water resource system implies the operation of the system so as to optimize some performance function. This optimization problem can be divided into the following parts:

- Definition of the performance function
- An internal description of the system and pertinent information about the state of the system
- Determination of the best policy from (1) and (2).

An example of a control system would be a regulator for a water distribution network that maintains water pressures within the system through the use of pumps and supplementary reservoirs. The description of the system is provided by the physical laws of fluid flow through pipes and the actual network configuration. The control system would have sensors measuring water pressure at various locations in the network. The performance function may be of the form

$$J(t) = \int \left[p_o(\underline{x}) - p(\underline{x},t)\right]^2 d\underline{x}$$

where

$p_o(\underline{x})$ is the target pressure as a function of location $\underline{x}$.

$p(\underline{x},t)$ is the measured pressure at location $\underline{x}$ and time t.

The optimal control problem can now be stated as:

From among all the admissible control functions (or policies) $u \in U$, find the one that minimizes $J(t)$, over all t, subject to the constraints of the internal dynamic system and all initial and terminal boundary conditions that may be specified.

The definition brings forward the important concept of controllability. If the

system is completely controllable, then there exists at least one policy, $u(\cdot)$, which can guide the system from its initial state to some other desired state. Controllability does not guarantee that a solution to the optimal control problem exists since the required control may not exist in the admissible set U. On the other hand, if a system is not controllable, it is not very meaningful to search for the optimal control.

One can imagine other water resources control problems besides the above example: the operation of a reservoir during a flood, the operation of wastewater treatment plants, the operation of irrigation supply canals, and so forth.

OPEN-LOOP AND CLOSED-LOOP CONTROL SYSTEMS

The example presented above represents an important class of control problems known as regulator problems. The problem is to find the appropriate control policy $u(\cdot)$ so that the controlled variable, $p(\underline{x},t)$ in the above example, tracks the reference variable $p_o(\underline{x})$, that is,

$$p(\underline{x},t) \simeq p_o(\underline{x}) \qquad t \geq t_o$$

where t_o is the time at which control begins.

There are two types of controllers that can be distinguished: open-loop and closed-loop. Open-loop controllers generate control policies on the basis of past and present values of the reference variable, as shown in Fig. 1(top). Closed-loop controllers, shown in Fig. 1(bottom), take advantage of the measurement information

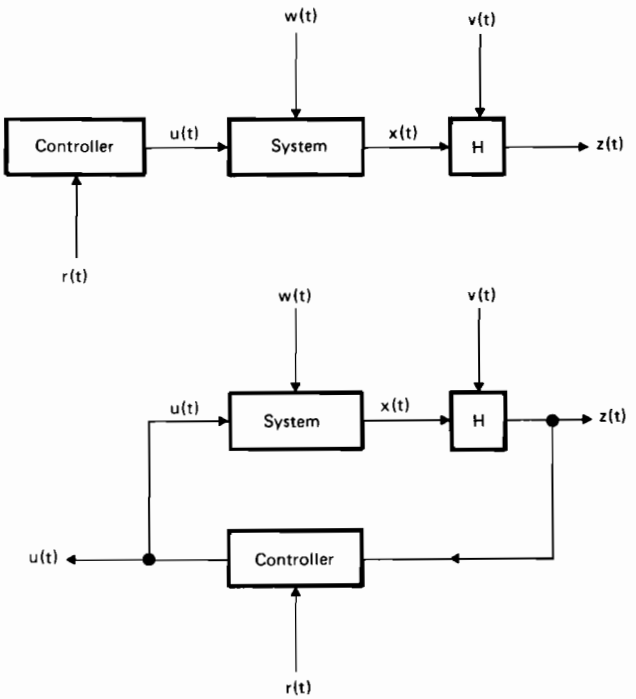


Fig. 1 (top) The open-loop control system. (bottom) The closed-loop control system.

provided by the sensors about the state of the system. The control policy becomes a function of both the reference variable and the measurements on the system. In Fig. 1, w(t) and v(t) are system disturbances and measurement error, x(t) is the state vector, H is a measurement system, z(t) are the measurements of the state variables, and r(t) is the reference variable for the controller. The system dynamics are assumed to be known.

Closed-loop controllers are much more powerful than open-loop controllers because they can accumulate information about the system during operation that will assist in developing an operating policy that compensates for disturbances.

Open-loop controllers do not have access to such information, so models in which they are used have to represent very accurately the causal links in the system to prevent it from wandering away from the desired targets.

STOCHASTIC CONTROL

In the real-time operation of hydrologic systems, it is unreasonable to assume that future states of the system are known. The control policy must be determined in the presence of unpredictable dynamics and state measurements. This is the stochastic control problem and it has not been widely applied to water resource systems.

To return to the water distribution example, a stochastic control problem would exist if the pipe friction was not perfectly known; if there were large and unexpected withdrawals, for example, for fire-fighting, at some location; or if the sensors measuring the pressure were inaccurate or incorrectly located.

The optimal, stochastic control problem is to determine the operating policy that optimizes the expected value of the performance function. The performance function is a random variable because the state of the system is uncertain.

Sorrenson (1976) points out that for stochastic systems, only closed-loop policies can give meaningful results. The performance of the open-loop control will be sensitive to the disturbances, but the control policy cannot be adjusted to compensate for them. Furthermore, this lack of causality really precludes the use of statistical black-box forecasting models for open-loop control.

In the closed-loop control situation, the control policy, u(t), is a function of the state. For linear systems with a quadratic performance function, the optimal policy u(t) is a linear feedback law, $u(t) = -K \cdot x(t)$. Part One of this volume considered the forecasting of states, x(t), when both the system equation and the measurements were subject to noise. The question now arises as to what will be the optimal control? It turns out that for linear systems corrupted by Gaussian noises and a quadratic performance function, the problem can be divided into two subproblems: find the optimal state estimator (in a least-squares sense) using, for example, a Kalman filter; and compute the control policy u(t), as if the state estimate $\hat{x}(t)$ were the true value x(t), within a deterministic controller. This is the well-known separation theorem (Aström 1970, Bryson and Ho 1975, and Casti 1977 all give good accounts of the theorem). A box diagram of the separation principle is given in Fig. 2 for closed-loop control.

Because of the importance of closed-loop control systems, the separation theorem is an important result for those situations where the assumptions hold. The theorem also points out the important relationship between the forecasting and control problems.

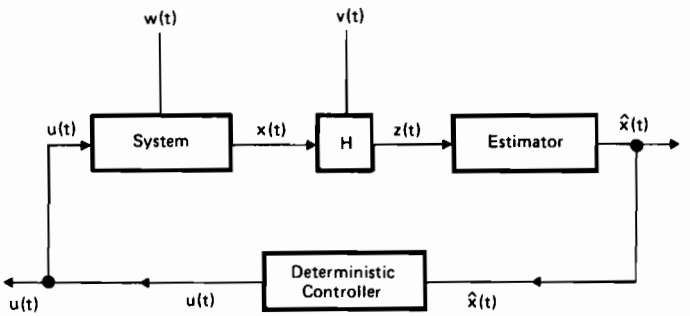


Fig. 2 Optimal closed-loop control showing the separation principle.

ADAPTIVE CONTROL

Up until now in the stochastic control problem, we have assumed that the system dynamics, the relationship between x and v,w, were known. There are many situations in which these dynamics are not known and, in Part One, adaptive identification procedures were discussed for simultaneously forecasting parameters and states.

An analogous situation exists in the control problem - the adaptive, stochastic controller. Here the problem is to develop the control policy u(t) and learn about the system (that is, identify the system) simultaneously. This problem is discussed in more detail by Szöllösi-Nagy and Wood in Chapter 12.

ON-LINE VERSUS OFF-LINE CONTROL

Much of control theory is concerned with computing the control strategy during normal system operation. With closed-loop control, this would also entail the monitoring and processing of measurement data and the determination of the controls. Such a mode of operation is called on-line or real-time control.

Another mode of operation, often called off-line control, would record the measurement data for later analysis. The computation of the control policy and its implication are separated in space and time. In many practical situations, off-line control can provide adequate results. If the time constants of the system are very long (that is, the response of the system to a control u(t) is on the order of days) as is often the case for certain water resource problems, then off-line control may be adequate. Off-line control avoids the problem of having extensive on-line computer facilities with the capability to process measurement data and determine controls very quickly. In many situations, it is just not required. Nevertheless, one still utilizes the results from the optimal stochastic control problem. One must determine the speed at which the new control for the time $t+\Delta t$ needs to be computed and ready for implementation, that is, the magnitude of Δt.

In the water distribution example presented earlier, one could imagine the need for on-line control because of the need to respond quickly to pressure fluctuations and because of the nature of the measurement system. For control of reservoir systems, off-line control may be suitable. For the problem of maintaining minimum flows for navigation or water quality maintenance, it may be adequate to have the states of the system (downstream stages, tributary discharges, reservoir levels, and so forth) measured once a day, have the data collected, analyzed, and the new control, $u(t+\Delta t)$, determined away from the controller. The new control setting

would then be implemented. The cost of such off-line control would be less than for on-line, automatic control.

LITERATURE REVIEW

There are a number of good, basic books on control theory. Good accounts of the deterministic control problem can be found in Kirk (1970), Brogan (1974), and Bryson and Ho (1975), and for stochastic control in Aoki (1967), Sage (1968), Aström (1970) and Fleming and Rishel (1975). Survey papers on the concepts and techniques of stochastic control theory have been written by Wittenmark (1975), Bar-Shalom and Tse (1976), and Sorrenson (1976).

The use of modern control theory for water resource systems has had only limited applications. Most of the work in reservoir operation has assumed a deterministic system and has concentrated on the long-term control policy, that is, a long-term strategy, often monthly, based on capacity, average demands, and prediction of monthly inflow from historical records. The problem is usually structured as an open-loop, deterministic control problem. Work representative of this approach includes Schweig and Cole (1968); Revelle et al. (1969), Gundelach and Revelle (1975); Leclerc and Marks (1973); Becker and Yeh (1974), Wilkinson (1972), Askew (1974) and Cole (1974). The problem of adjusting these controls in light of actual demands and inflows has not been fully considered, even though Mejia et al. (1974) compares the improved benefits from such policies over the "fixed" policy analysis. Jamieson and Wilkinson (1972) considered short-term reservoir control but assumed that the states and the system dynamics were known.

For deterministic control theory applied to water resource systems in general, there are papers by Buras (1967), Larson and Keckler (1969), and Jamshidi and Heidari (1975).

Some work has appeared applying stochastic control theory to water resource systems: Bather (1962), Levin (1969), Croley (1974), Takeuchi and Moreau (1974), and Strupczewski et al. (1975).

In the areas of water quality and wastewater treatment, more work has appeared in applying the concepts of modern control theory; see Dysart and Hines (1970), Winn and Moore (1973), Tamura (1974), Koivo and Phillips (1975), Gourishankar and Lawson (1975), Lin (1975), Powers and Canale (1975), Singh (1975), Young and Beck (1974), Huck and Farquhar (1974), Penumalli et al. (1976), and Williams and Hinwood (1976). For distributed parameter systems, which are much more difficult to handle, there has been some work by Tarassov et al. (1969), Davidson and Bradshaw (1970), Hullett (1974), and Özgören et al. (1975). In control for waste treatment plants, there has been work by Berthoeux et al. (1975), Fan et al. (1973), and Andrews (1974) as well as others. The more extensive use of modern control theory in water quality may be due to the fact that the systems can be described by a set of differential equations, upon which the state-space techniques are based. Surveys of recent work in water quality and wastewater treatment are given by Olsson (1977) and in Chapters 6 (by Olsson) and 11 (by Beck).

There are four papers in this part of the volume. Beck (Chapter 11) develops the concepts for forecasting and control of water quality around three case studies: in-stream quality control, state reconstruction in a wastewater treatment plant, and prediction for urban sewer flows. Chapter 12, by Szöllösi-Nagy and Wood, develops the concepts of adaptive feedback control, beyond what has been given in this introduction, using a Bayesian approach. The final two papers present applications. Maidment (Chapter 13) applies the state-space approach to reservoir control. This paper shows that one need not develop an automatic control but rather could have manual or off-line control based on state forecasts. This may be

more practical in real-world applications. In Chapter 14, Stehfest looks at the problem of real-time control of dissolved oxygen in rivers.

REFERENCES

Andrews, J.F. (1974). Dynamic models and control strategies for wastewater treatment processes. Water Res. 8: 261-289.

Aoki, M. (1967). Optimization of Stochastic Systems. Academic Press, New York.

Askew, A. (1974). Chance-constrained dynamic programming and the optimization of water resource systems. Water Resour. Res. 10(6): 1099-1106.

Aström, K.J. (1970). Introduction to Stochastic Control Theory. Academic Press, New York.

Bar-Shalom, Y., and E. Tse (1976). Concepts and methods in stochastic control, in C.T. Leondes, Ed., Control and Dynamic Systems, Advance in Theory and Application Vol. 12. Academic Press, New York.

Bather, J.A. (1962). Optimal regulation policies for finite dams. J. Soc. Ind. Appl. Math. 10: 3.

Becker, L., and W. Yeh (1974). Optimization of real time operation of a multiple-reservoir system. Water Resour. Res. 10(6): 1107-1112.

Berthoeux, P.M., W.G. Hunter, L.C. Pallesen, and C-Y. Shih (1975). Modelling sewage treatment plant input BOD data. Proc. ASCE, J. Env. Eng. Div. 101 (EE1): 127-138.

Brogan, W.L. (1974). Modern Control Theory. Quantum, New York.

Bryson, A.E., and Y-C. Ho (1975). Applied Optimal Estimation and Control. Hemisphere Publishing, Washington, D.C.

Buras, N. (1967). Dynamic programming in water resource development, in V.T. Chow, Ed., Advances in Hydrosciences Vol 3. Academic Press, New York.

Casti, J. (1977). Dynamical Systems and Their Application: Linear Theory. Academic Press, New York.

Cole, J.A. (1974). Optimal seasonal and short-term operation of a reservoir used for flood control and water supply. IAHS Publ., No. 101, 978-988.

Croley, T.E., II (1974). Sequential stochastic optimization for reservoir systems. Proc. ASCE, J. Hyd. Div. 100 (HY1): 201-219.

Davidson, B., and R.W. Bradshaw (1970). A steady state optimal design of artificially induced aeration in polluted streams by the use of Pontryagin's minimum principle. Water Resour. Res. 6(2): 389-397.

Dysart, B.C., and W.W. Hines (1970). Control of water quality in a complex natural environment. IEEE Trans. Syst. Sci. Cybern. 6(4): 322-329.

Fan, L.T., P.S. Shah, N.C. Pereira, and L.E. Erickson (1973). Dynamic analysis and optimal feedback control synthesis applied to biological waste treatment. Water Res. 7(11): 1609-1641.

Fleming, W.H., and R.W. Rishel (1975). Deterministic and Stochastic Optimum Control.

Springer Verlag, New York.

Gourishankar, V., and R.L. Lawson (1975). Optimal control of water pollution in a river system. Int. J. Syst. Sci. 6(3): 201-276.

Gundelach, J., and C. Revelle (1975). Linear decision rule in reservoir management and design, 5, A general algorithm. Water Resour. Res. 11(2): 204-207.

Huck, P.M., and G.J. Farquhar (1974). Water quality models using the Box-Jenkins method. Proc. ASCE, J. Env. Eng. Div. 100 (EE3): 733-752.

Hullett, W. (1974). Optimal estuary aeration: An application of distributed parameter control theory. Applied Mathematics and Optimization 1(1): 20-63.

Jamieson, D.G., and J.C. Wilkinson (1976). River Dee research program 3. A short-term control strategy for multi-purpose reservoir systems. Water Resour. Res. 8(4): 911-920.

Jamshidi, M., and M. Heidari (1975). Modeling and optimization of Khuzestan water resources systems. Proc. 6th IFAC Congress, Boston, Massachusetts.

Kirk, D.E. (1970). Optimal Control Theory. Prentice-Hall, Englewood Cliffs, New Jersey.

Koivo, A.J., and G.R. Phillips (1975). Optimization of dissolved oxygen concentration in polluted streams using instream aeration. Proc. 6th IFAC Congress, Boston, Massachusetts.

Larson, R.E., and W. Keckler (1969). Application of dynamic programming to the control of water resource systems. Automatica 5: 15-26.

Leclerc, G., and D.H. Marks (1973). Determination of the discharge policy for an existing reservoir network under differing objectives. Water Resour. Res. 9(5): 1155-1165.

Levin, O. (1969). Optimal control of a storage reservoir during a good season. Automatica 5: 27-34.

Lin, S.H. (1975). Feedforward water quality control by aeration of a thermally and organically polluted stream. Int. J. Syst. Sci. 6(10): 919-932.

Mejia, J.M., P. Egle, and A. Leclerc (1974). Evaluating multi-reservoir operating rules. Water Resour. Res. 10(6): 1090-1098.

Olsson, G. (1977). State of the art in sewage treatment plant control. AIChE Symp. Ser. 72(159): 52-76.

Özgören, M.K., R.W. Longman, and C.A. Cooper (1975). Optimal control of a distributed parameter time-lagged river aeration system. Trans. ASME, Ser. G., J. Dyn. Syst. Meas. Control 97(2): 164-171.

Penumalli, B.R., R.H. Flake, and E.G. Fruh (1976). Large-scale estuarine water quality matrix model. Proc. ASCE, J. Env. Eng. Div. 102 (EE1): 191-209.

Powers, W.F., and R.P. Canale (1975). Some applications of optimization techniques to water quality modelling and control. IEEE Trans, Syst. Man. Cybern. 5(3): 312-321.

Revelle, C., E. Joeres, and W. Kirby (1969). Linear decision rules in reservoir

management and design, 1. Development of the stochastic model. Water Resour. Res. 5(4): 767-777.

Sage, A.P. (1968). Optimum Systems Control. Prentice-Hall, Englewood Cliffs, New Jersey.

Schweig, Z., and J.A. Cole (1968). Optimum control of linked reservoirs. Water Resour. Res. 4(3): 479-498.

Singh, M.A. (1975). River pollution control. Int. J. Syst. Sci. 6(1): 9-21.

Sorrenson, H.W. (1976). An overview of filtering and stochastic control in dynamic systems, in C.T. Leondes, Ed., Control and Dynamic Systems, Advances in Theory and Application 12. Academic Press, New York.

Strupczewski, W.G., R.J. Kiczko, and R.J. Budzianowski (1975). General Markov model for optimal control of reservoirs. IAHS Publ., No. 115, 222-230.

Takeuchi, K., and D.H. Moreau (1974). Optimal control of multi-unit interbasin water resource control systems. Water Resour. Res. 10(3): 407-414.

Tamura, H. (1974). A discrete dynamic model with distributed transport delays and its hierarchical optimization for preserving stream quality. IEEE Trans, Syst. Man. Cybern. 4: 424-431.

Tarassov, V.J., H.J. Perlies, and B. Davidson (1969). Optimization of a class of river aeration problems by the use of multi-variable distributed parameter control theory. Water Resour. Res. 5(3): 563-573.

Wilkinson, J.C. (1972). River Dee research program, 2. Long-term control strategy for a multipurpose reservoir. Water Resour. Res. 8(4): 904-910.

Williams, B.J., and J.B. Hinwood (1976). Two-dimensional mathematical water quality model. Proc. ASCE, J. Env. Eng. Div. 102 (EE1): 149-163.

Winn, B., and J. Moore (1973). The application of optimal linear regulatory theory to a problem in water pollution. IEEE Trans., Syst. Man Cybern. 3: 450-455.

Wittenmark, B. (1975). Stochastic control methods - A survey. Int. J. Control 21: 705-730.

Young, P., and B. Beck (1974). The modelling and control of water quality in a river system. Automatica 10(5): 455-468.

11 Forecasting and Control of Water Quality

M.B. Beck

Seven or eight years ago the idea that principles of control engineering should be applied to water resources management had barely germinated in the minds of the control engineering fraternity. Equally, the idea of automation in the water industry was, with few exceptions, alien to the practising profession of water and wastewater engineering. It seemed then that the word control as used in control theory had a very different meaning from the word control in water quality control; water quality control was interpreted as meaning control by legislation and control by the more effective design of treatment plants (be they plants for water purification or for sewage treatment). Most significantly, it was not apparent to the control engineer that water quality control was being interpreted, at that time, as the more effective operation of a treatment plant.

A number of articles drawn from the recent control-oriented literature (Gourishankar and Lawson 1975, Lin 1975, Özgören et al. 1975, Powers and Canale 1975) and the environmental engineering journals (Berthoeux et al. 1975, Huck and Farquhar 1975, Penumalli et al. 1976, Williams and Hinwood 1976) show that the situation is now changing rapidly. Even so, the practice of automated water quality maintenance is in its infancy; the time is still ripe for carefully considered discussion of simple, basic concepts and problems before further exploration of the possibilities for sophisticated solutions. This paper attempts, therefore, to take a comprehensive look at the dynamic aspects of water quality modeling and control; it introduces some of the fundamental techniques and principles of control theory and indicates how they might be put to good use in the organization and operation of systems for water quality management. The discussion is set primarily against the background of several previous reviews of the state of the art in water quality management (Andrews 1974, Buhr et al. 1974, Olsson 1977, Beck 1976a).

ASPECTS OF WATER QUALITY IN WATER RESOURCE SYSTEMS

Real-time forecasting and control of water quality has received relatively little attention in comparison with analogous problems of river-flow prediction and regulation probably because flood damage prevention and drought alleviation are more obvious targets for problem solving and capital investment, the benefits of "cleaner" rivers are largely intangible, and the instrumentation available for water quality forecasting and control is at present very limited. Certainly, from the point of view of economics, it has generally been assumed that running costs of the water and wastewater industries are small, if not negligible, when compared with their capital costs. This assumption provides no incentive to develop

innovative techniques of forecasting and control and has led to a preoccupation in water quality modeling studies with steady-state analyses of long-term investment plans.

In a country such as the U.K. where water resources are used intensively, the management of river water quality, in real-time, is becoming increasingly important. Thus, as a possible motivation for studying this subject we might resolve the conflict of alternatives between the qualitative and quantitative aspects of water resources into the grossly simplified situation of Fig. 11.1. The alternative of

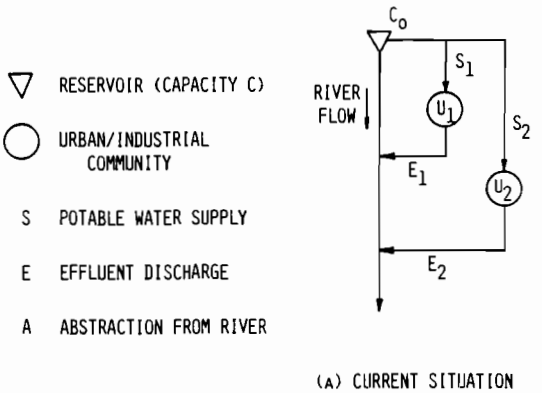
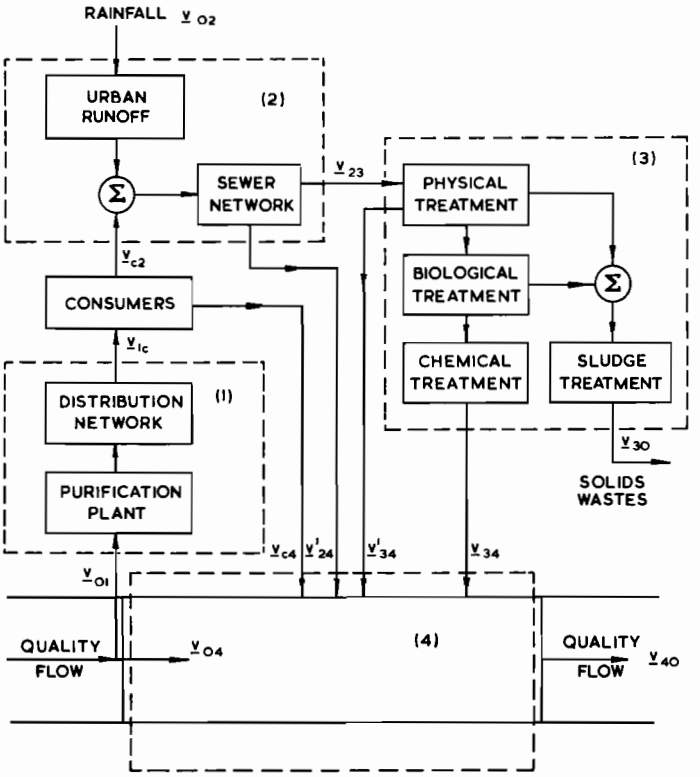


Fig. 11.1 Qualitative and quantitative alternatives for water resources management: (A) current situation; (B) future policy alternative exploiting use and reuse of river water by management of river water quality; (C) future policy alternative based on increased exploitation of clean water sources. Notational relationships: $S_1'>S_1$; $S_2'>S_2$; $E_1'>E_1$; $E_2'>E_2$.

constructing more or larger reservoirs or both (Fig. 11.1C) simply to satisfy the increased demands in the adjacent urban/industrial centers can largely be discounted. The other alternative (Fig. 11.1B) of river regulation as an integral component of river basin management (e.g., Water Resources Board 1974) implies, among other factors of real-time hydrological operations, a growing emphasis on knowledge and control of river water quality.

If we look more closely at the urban/industrial community of Fig. 11.1B (the dashed-line box) the fundamental component features of water quality may be conveniently summarized by the water quality system of Fig. 11.2 (Beck 1976a). To



SUBSYSTEMS :-

(1) POTABLE WATER ABSTRACTION, PURIFICATION, AND SUPPLY NETWORK

(2) URBAN LAND RUNOFF AND THE SEWER NETWORK

(3) WASTEWATER TREATMENT PLANT

(4) A STRETCH OF RIVER

Fig. 11.2 The water quality system; the fluxes of material $\underline{v}$ are vectors comprising volumetric flow rate and a number of water quality characteristics (from Beck 1976a).

see how the four subsystem blocks of Fig. 11.2 represent all the essential modifications, disturbances, and interactions of water quality, an analytical picture of a hypothetical river system can be constructed as in Fig. 11.3. This particular rearrangement of the subsystem blocks enables us to formulate two basic control problems of water quality. Fig. 11.4 shows thus the input/output relationship between waste discharges from urban/industrial center U_1 and the potable supply to the next downstream community U_2. First, consider the control of the treated discharge $\underline{v}_{34}$ to the river given the input disturbances $\underline{v}_{c2}$. Here the control objectives are to minimize the deviation of the (intermediate) output $\underline{v}_{34}$ about some desirable time-invariant set-point on the output, i.e., the regulator problem,

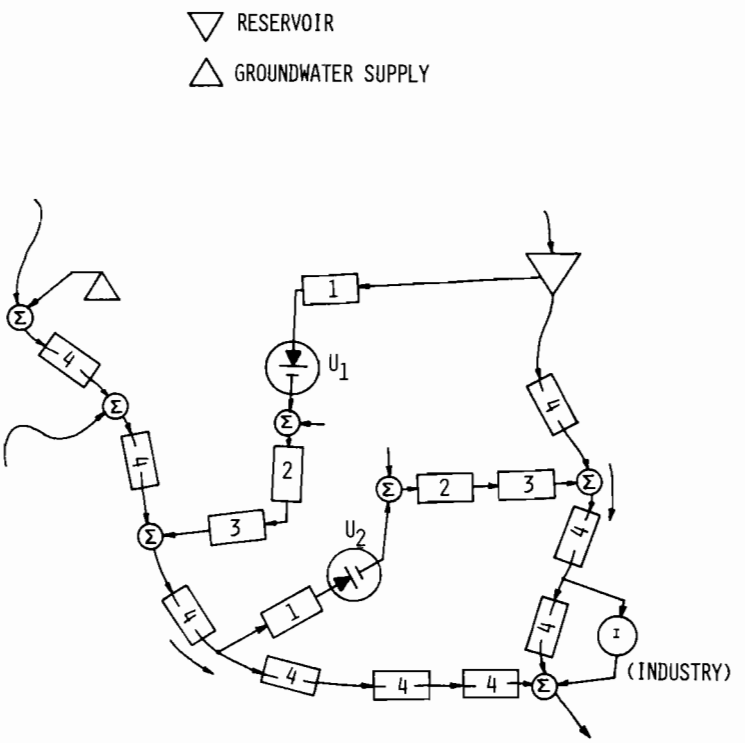


Fig. 11.3 Analytical picture of a hypothetical water resource system. U_1 and U_2 represent urban/industrial centers; the block numbers refer to the subsystems defined in Fig. 11.2.

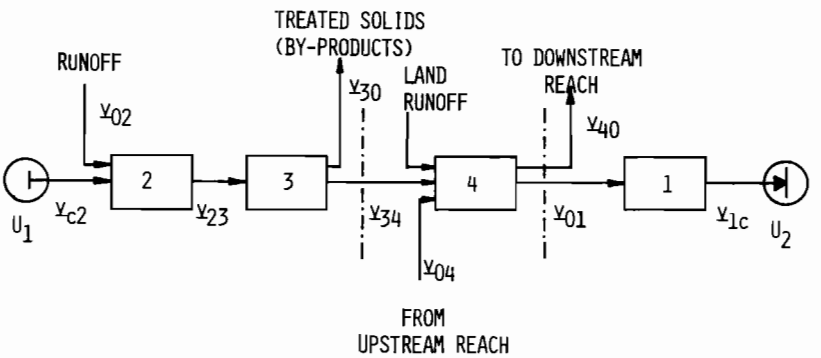


Fig. 11.4 Basic input/output flowchart for U_1 and U_2 in Fig. 11.3. The dashed lines indicate the division of the flowchart for consideration of the control problems in Fig. 11.5.

Fig. 11.5 (top). In fact, the controller should also attempt to "eliminate" load variations imposed on the river's quality resources by land runoff and upstream disturbances, $\underline{y}_{04}$. The other control problem is illustrated by Figure 11.5 (bottom). For a process (subsystem) that receives a steady load input $\underline{y}_{01}$ (i.e., constant

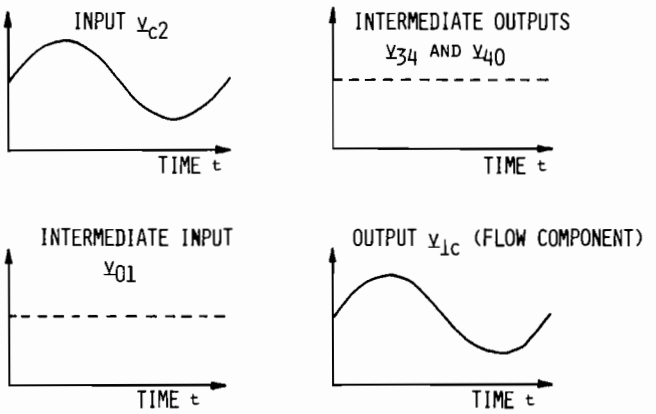


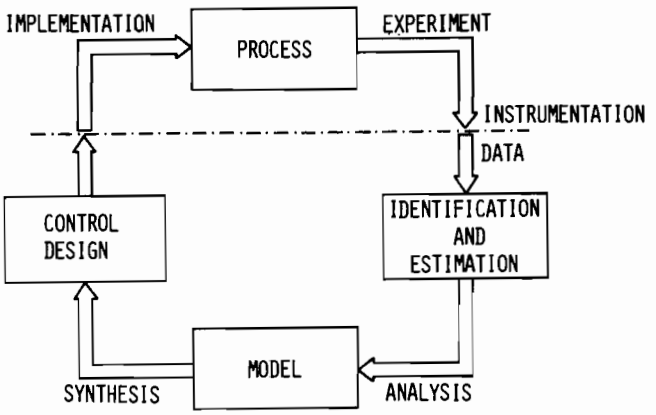
Fig. 11.5 Control problems in the water quality system: (top) in the regulator problem the controller attempts to provide a constant output given a time-varying input load; (bottom) in the servomechanism problem the controller forces the output to follow some specified time-varying performance while the input load is constant. (The dashed lines denote that conditions in the upstream reach of river and desirable conditions in the downstream reach of river may not necessarily be time-invariant; this will impose variations (of quality) on y_{34} and y_{01}.)

abstraction rate), the control objective may be to minimize the deviation of the output y_{1c} from a prespecified, time-varying, output demand function (or supply curve); this problem is commonly referred to as the servomechanism problem.

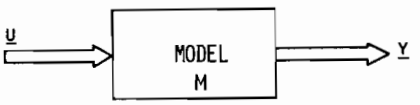
The application of effective control to a process requires sufficient flexibility in the process behavior to make the necessary control maneuvers. In other words, in the familiar setting of the flood regulation problem, a reservoir of relatively small storage capacity would be of little use as a control mechanism. It is precisely this lack of visible flexibility, or storage capacity, which makes the system of Fig. 11.4 a difficult process to control. In the following, then, our ultimate objectives are to investigate how best to utilize real-time forecasting and the control of existing ambient storage (in the sewer network and wastewater treatment plant) for the management of river water quality.

Most studies of control system design assume that a model of process dynamic behavior is available. Unfortunately this is rarely the case for the unit processes shown in Fig. 11.2; again, perhaps this is another factor contributing to the overall lack of control and forecasting applications in water quality systems. Although we shall not deal with the subject here, the analytical step entitled "Identification and Estimation" in Fig. 11.6A is an important prerequisite for control system synthesis. Notice also in Fig. 11.6A that instrumentation is both a barrier in undertaking experimental studies of process behavior for system identification and parameter estimation and a fundamental problem for control implementation. That is to say, process behavior can be controlled no more accurately than it can be observed. Detailed discussions of modeling and system identification can be found in the general texts by Aström and Eykhoff (1971) and Eykhoff (1974); more specific comments on the identification of biological process models in wastewater treatment and river water quality are presented elsewhere (Beck 1976b, Beck 1977b).

Figure 11.6B summarizes the various types of problem that one may wish to solve for



(A)



TYPE OF PROBLEM	GIVEN	TO FIND	
SIMULATION (PREDICTION)	$\underline{u}$, M	$\underline{y}$	
ANALYSIS (IDENTIFICATION)	$\underline{u}$, $\underline{y}$	M	
SYNTHESIS (CONTROL)	M, $\underline{y}$	$\underline{u}$	

(B)

Fig. 11.6 Modeling and control: (A) the iterative nature of experiment, analysis, synthesis, and implementation in practice and their relationship with process instrumentation; (B) problem definitions.

forecasting and control purposes. With respect to Case Study 3 below it can be seen that a simultaneous solution of the "simulation" and "analysis" problems leads to the idea of an adaptive, self-tuning prediction scheme. Similarly, algorithms combining the tasks of learning (modeling, analysis) and control are known collectively as adaptive controllers.

CONTROLLERS AND CONTROL SYSTEM DESIGN

In a controlled process the function of the controller can be defined as follows: a controller collects all available information (measurements) from the system being controlled and uses it to manipulate some of the system inputs, $\underline{u}$ in Fig. 11.6B, in order to bring about some desired process performance, as gauged by the

behavior of the output y in Fig. 11.6B. The controller design is usually based on two general principles, the feedforward principle and the feedback principle. See MacFarlane (1976) for a discussion of the feedback principle.

Let d(t) be a measurable disturbance variable, y(t) a control (manipulable) input variable, and y(t) the controlled output variable, where t is the independent variable of time. In order to illustrate the principles of feedforward and feedback controllers we shall transfer the analysis from the time domain to the frequency domain by introducing the Laplace transforms D(s), U(s), and Y(s) of the variables d(t), u(t), and y(t). Many standard texts on control system design are developed around an analysis of process dynamics in the frequency domain; see, e.g., Coughanowr and Koppel (1965) and Dorf (1973). The feedforward principle is concerned with the rejection of measurable disturbances. Fig. 11.7 illustrates

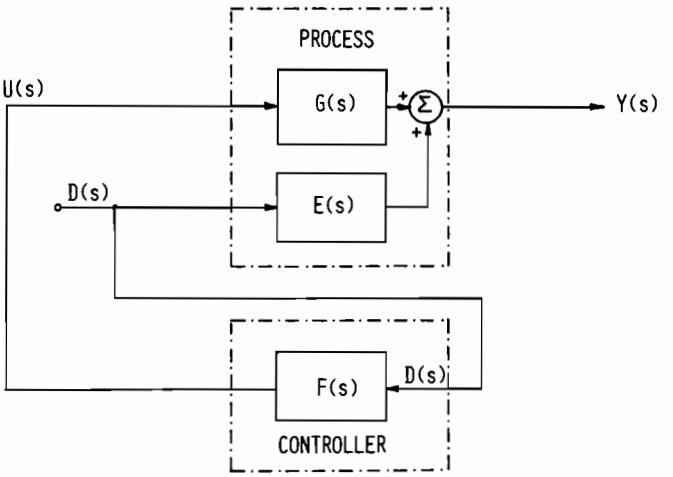


Fig. 11.7 The feedforward controller principle.

this principle, where G(s), E(s), and F(s) are transfer function relationships between the variables U(s) and Y(s), D(s) and Y(s), and D(s) and U(s), respectively. We have then,

$$Y(s) = E(s)\,D(s) + G(s)\,U(s) = (E(s) + G(s)\,F(s))D(s) \qquad (11.1)$$

and if the controller F(s) is designed so that

$$F(s) = -E(s)/G(s) \qquad (11.2)$$

we see that the effects of the disturbances D(s) on the output Y(s) are canceled. Unfortunately, the feedforward controller suffers from three major disadvantages:

- It requires accurate measurement of the disturbances for the signal fed into the controller.
- It requires an accurate model of the process dynamics, i.e., G(s) and E(S) in (11.2).
- It does not utilize a measurement of the output Y(s) and therefore cannot

take account of any inevitable misalignment of desired and actual performance of the process.

In view of the disadvantages of feedforward control - and these would appear to be especially severe with respect to the water quality system where there are poor models and poor measurement facilities - it is perhaps a little surprising to see this concept being promoted in some of the recent literature (Buhr et al. 1974, Institution of Chemical Engineers 1974).

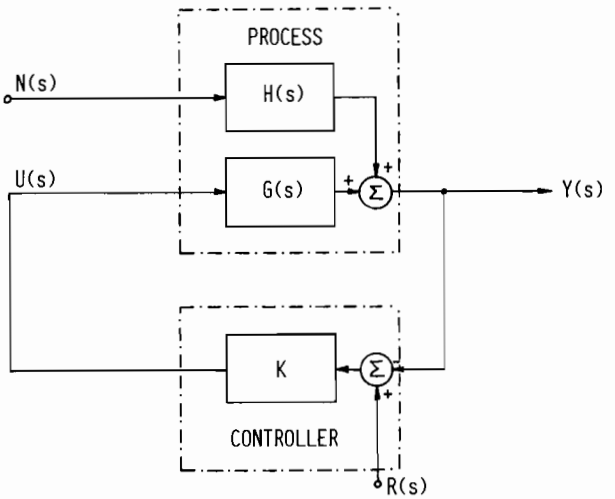


Fig. 11.8 The feedback controller principle.

The feedback controller is concerned with the rejection of unmeasurable disturbances, N(s), for example, as shown in Fig. 11.8. In this case,

$$Y(s) = G(s)U(s) + H(s)N(s) \tag{11.3}$$

and

$$U(s) = K(R(s) - Y(s)) \tag{11.4}$$

where R(s) is a variable expressing the desired performance of the ouput Y(s). Hence, on substituting (11.4) into (11.3),

$$Y(s) = KG(s)R(s) - KG(s)Y(s) + H(s)N(s) \tag{11.5}$$

or

$$(1+KG(s))Y(s) = KG(s)R(s)+H(s)N(s) \tag{11.6}$$

It is at this point that the idea of frequency response analysis becomes particularly useful in considering the properties of feedback control. If we substitute for the complex variable $s = \sigma + j\omega$ its purely imaginary component $s = j\omega$, then the frequency response of any process with transfer function T(s), say, can be determined from $T(j\omega)$. By frequency response we mean that the amplitude ratio and relative phase shift between a sinusoidal input disturbance to the process and the corresponding output sinusoidal response are given by $|T(j\omega)|$ and $\angle T(j\omega)$; ω can thus be interpreted as the frequency of the input sinusoidal signal. Hence from (11.6),

$$Y(j\omega) = (KG(j\omega)R(j\omega) + H(j\omega)N(j\omega))/(1+KG(j\omega)) \qquad (11.7)$$

where by choosing K (a simple proportional gain controller) such that $K|G(j\omega)| >> 1$, i.e., the controller has a high gain, we obtain,

$$Y(j\omega) \approx R(j\omega)+H(j\omega)N(j\omega)/(1+KG(j\omega)) \qquad (11.8)$$

And if by this choice, in particular, $K|G(j\omega)| >> |H(j\omega)|$, then (11.8) can be further approximated to,

$$Y(j\omega) \approx R(j\omega) \qquad (11.9)$$

indicating that the controller brings the actual process performance to the desired performance in spite of the unmeasurable disturbances.

The advantage of the feedback controller is that it can be designed so that its operation is relatively insensitive to any errors in the model G(s) and H(s) of the true process dynamics that are used for synthesis purposes. There is, of course, still a requirement for accurate measurement of the output variable since the control action is based on the difference (error), $\zeta(s) = R(s)-Y(s)$, between desired and actual performance. We have already mentioned this point and we shall return to it later. The major theoretical limitation on the feedback controller is the overriding desire to maintain system stability: as K is chosen so that $K|G(j\omega)|$ becomes larger and larger we find that the system, under closed-loop control, may well become unstable.

Two other features of process control might usefully be introduced here - again, the reader is referred elsewhere for a more complete discussion of these topics, e.g., Coughanowr and Koppel (1965). The simple proportional gain representation of the controller in the preceding analysis of the feedback principle can be extended to include control actions of a derivative and integral nature. In other words, in the time domain, the control variable u(t) is given as a function of the error in performance, $\zeta(t) = r(t)-y(t)$, by the relationship,

$$u(t) = K_c\left(1+\beta_I \int_0^t \zeta(t)dt + \beta_D \frac{d\zeta(t)}{dt}\right) \qquad (11.10)$$

where β_I and β_D are constants relating respectively to the integral and derivative components of the controller, and K_c is the constant previously referred to as the proportional gain. Taking Laplace transforms, such a three-term proportional, integral, derivative action (PID) controller has the following (ideal) transfer function in the frequency domain,

$$K(s) = \frac{U(s)}{\zeta(s)} = K_c(1+\beta_I/s + \beta_D s) \qquad (11.11)$$

And since the controller now contains dynamic elements, its properties influence not only the gain characteristics but also the phase shift characteristics of the closed-loop system frequency response. For this case, a high-gain controller may not be of paramount importance or even desirable; to control process behavior, it may be more important to know whether the disturbances $D(j\omega)$ and $N(j\omega)$ are composed primarily of low-frequency (e.g., diurnal variations) or high-frequency (e.g., random noise) oscillations. The choice of the three parameters K_c, β_I, and β_D of the controller will thus depend upon several interacting factors and it is difficult to make further generalizations on solutions to any given control system design

problem. The three-term controller, with its foundations in the "classical" stages of control theory development, has found and continues to find wide application in the process industries.

Case Study 1 - Control of River Dissolved Oxygen Concentration

This example deals with the control of the dissolved oxygen (DO) concentration in a river by manipulation of the rate at which an effluent is discharged to the river. Hence, the biochemical oxygen demand (BOD) load of the effluent is regulated by changing the discharge rate of the effluent instead of by manipulating its BOD concentration. The system is defined according to Fig. 11.9. Treated

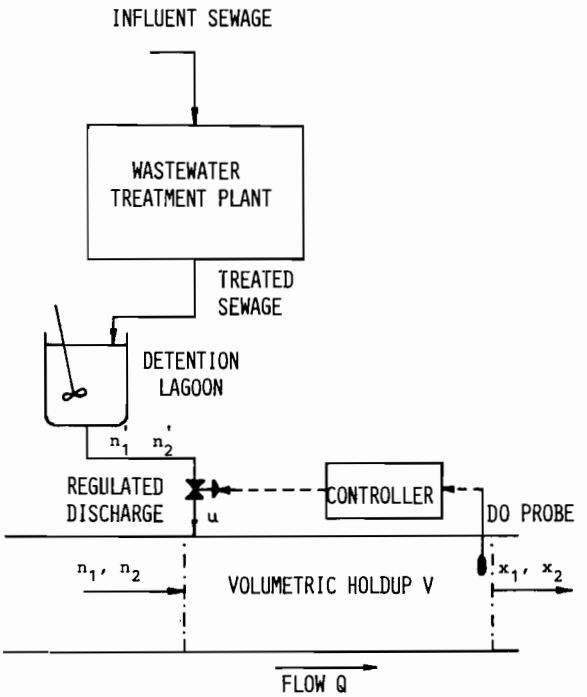


Fig. 11.9 A schematic diagram for in-stream DO control by flow detention of treated effluent; all variables are as defined for Eqs. (11.12) and (11.13).

effluent is assumed to enter and leave the detention lagoon with its DO and BOD constituents behaving as noninteracting conservative substances; the lagoon has the idealized behavior of a continuously stirred tank reactor.

The dynamic model for DO-BOD interaction in the reach of river is given by the following differential equation form (Beck and Young 1975)

$$\text{DO: } \dot{x}_1(t) = -(\alpha_1+[Q(t)+u(t)]/V)x_1(t)-\alpha_2 x_2(t)+(u(t)/V)n_1'(t)+(Q(t)/V)n_1(t) + \alpha_1 C_s(t)+s_{11}(t)+s_{12} \quad (11.12)$$

$$\text{BOD: } \dot{x}_2(t) = -(\alpha_2+\alpha_3+[Q(t)+u(t)]/V)x_2(t)+(u(t)/V)n_2'(t)+(Q(t)/V)n_2(t) + s_{21}(t) + s_{22} \tag{11.13}$$

where

$x_1(t), x_2(t)$ = downstream concentrations of DO and BOD, respectively (gm^{-3});

$n_1(t), n_2(t)$ = upstream concentrations of DO and BOD, respectively (gm^{-3});

$n_1'(t), n_2'(t)$ = concentration of DO and BOD, respectively, in the effluent discharge (gm^{-3})

$u(t)$ = regulated outflow from the detention lagoon (m^3day^{-1});

$Q(t)$ = upstream discharge of the river (m^3day^{-1});

V = (constant) mean volumetric hold-up of water in reach of river (m^3);

$C_s(t)$ = saturation concentration of DO (gm^{-3});

$s_{11}(t)$ = term representing the net production of DO by the photosynthetic/respiratory activity of algae ($gm^{-3}day^{-1}$);

s_{12} = rate of DO removal from river water by decomposition in bottom mud deposits ($gm^{-3}day^{-1}$);

$s_{21}(t)$ = term representing the net addition of a BOD load by redissolution of decomposing dead algal matter ($gm^{-3}day^{-1}$);

s_{22} = rate of BOD addition to reach of river in local surface runoff ($gm^{-3}day^{-1}$);

α_1 = reaeration rate constant for DO (day^{-1});

α_2 = BOD decay rate constant (day^{-1});

α_3 = rate constant for sedimentation of particulate BOD material (day^{-1}).

Further details of the model, in particular the form of the terms $s_{11}(t)$ and $s_{21}(t)$, are given in the source reference and also in Beck (1975).

For the design of a feedback controller we first need to determine the transfer function relationship G(s) between the control variable u(t), and the output y(t), where y(t) is the downstream DO concentration, i.e.

$$y(t) = x_1(t) \tag{11.14}$$

Equations (11.12) and (11.13) can be rearranged to give, in vector-matrix form,

$$\dot{\underline{x}}(t) = A(t)\underline{x}(t) + B(t)\underline{u}(t) + \underline{n}(t) \tag{11.15}$$

where

$$\underline{x}^T(t) = [x_1(t),\ x_2(t)];$$

$$A(t) = \begin{bmatrix} -(\alpha_1 + [Q(t)+u(t)]/V) & -\alpha_2 \\ 0 & -(\alpha_2+\alpha_3+[Q(t)+u(t)]/V) \end{bmatrix}; \quad B(t) = \begin{bmatrix} n_1'(t)/V \\ n_2'(t)/V \end{bmatrix}$$

and

$$\underline{n}(t) = \begin{bmatrix} (Q(t)/V)n_1(t)+\alpha_1 C_s(t)+s_{11}(t)+s_{12} \\ (Q(t)/V)n_2(t)+s_{21}(t)+s_{22} \end{bmatrix}$$

Here the notation $\underline{n}(t)$ is used deliberately, although now in vector form, in order to preserve the analogy with the preceding discussion of the feedback controller principle. Thus if we rewrite (11.14) as,

$$y(t) = [1 \quad 0]\underline{x}(t) = C\underline{x}(t) \tag{11.16}$$

in order to define C, then the required transfer function G(s) can be derived as,

$$G(s) = \frac{Y(s)}{U(s)} = C[sI-A]^{-1}B \tag{11.17}$$

Equation (11.17) is obtained simply by taking Laplace transforms of (11.15) and (11.16), with $\underline{n}(t) \equiv \underline{0}$, and then substituting for X(s). And similarly, to complete the problem formulation, the transfer function vector between the disturbance vector $\underline{n}(t)$ and the output y(t) can be obtained as

$$H(s) = C[sI-A]^{-1} \tag{11.18}$$

In both (11.17) and (11.18) I is the identity matrix.

On the basis of our knowledge of G(s) and H(s) a number of different controllers can be designed using a variety of synthesis techniques, e.g., Young and Beck (1974), Gourishankar and Ramar (1977), and Koivo (1977). Chapter 14 also contains a discussion about controllers. Among the simplest of these, however, is the three-term controller already introduced above. Hence, for example, by choosing a proportional + integral action controller of the form,

$$K(s) = (2.8\times10^4)(1+1/s) \tag{11.19}$$

typical simulation results for DO control are obtained as in Fig. 11.10. For this particular example, the control signal u(t) is generated as a combination of a mean, average "command" signal u_c and the appropriate manipulation of the error signal between actual and desired performance,

$$u(t) = u_c + K_c(\zeta(t) + \int_0^t \zeta(t)dt) \qquad (11.20)$$

By analogy, K_c in (11.19) is $(2.8\times10^4)(m^3day^{-1})/(gm^{-3}DO)$; in the context of the foregoing analysis of the feedback controller principle this is not necessarily a high gain but is a result of the respective units of u(t) and $\zeta(t)$. Other relevant details of the numerical values used for control system design and simulation are given in Table 11.1. These parameter values and the time-series

Table 11.1. Parameter values for control system design and simulation

Parameter	Value	Remarks
α_1	$0.17\ day^{-1}$	
α_2	$0.32\ day^{-1}$	
α_3	$0.18\ day^{-1}$	
V	$15.1 \times 10^4\ m^3$	
Q	$8.4\times10^4\ m^3day^{-1}$	Average value assumed for control system design.
u_c	$2.8\times10^4\ m^3day^{-1}$	Also substituted for u in control system design.
s_{12}	$0.5\ gm^{-3}day^{-1}$	
s_{22}	$0\ gm^{-3}day^{-1}$	
n_1'	$2.0\ gm^{-3}$	Constant average values assumed for simulation purposes.
C_s	$10.0\ gm^{-3}$	
n_2	$1.0\ gm^{-3}$	

data for Q(t), $n_1(t)$, the volume of treated sewage, and the BOD concentration in that sewage as it enters the hypothetical lagoon are all taken from an experimental study of the River Cam in eastern England (Beck and Young 1975). The variables $s_{11}(t)$ and $s_{21}(t)$ in the disturbance vector $\underline{n}(t)$ in (11.15) are generated in a deterministic fashion from a model for algal population dynamics described in Beck (1975); the peak downstream DO and BOD responses of Fig. 11.10(C) and 11.10(D) are in fact a replication of the observed effects of an algal bloom.

The major points to notice about the controller and its performance are as follows.

- The basis of the model for design purposes is the A matrix of (11.15) which determines the characteristic polynomial (see, e.g., Coughanowr and Koppel 1965) of the process dynamics; this matrix embodies nothing more complex than the classical Streeter-Phelps (1925) assumptions about DO-BOD interaction.

- Even though the controller is designed from such a simple model, its performance is still adequate. Notice that the A matrix, in particular, is

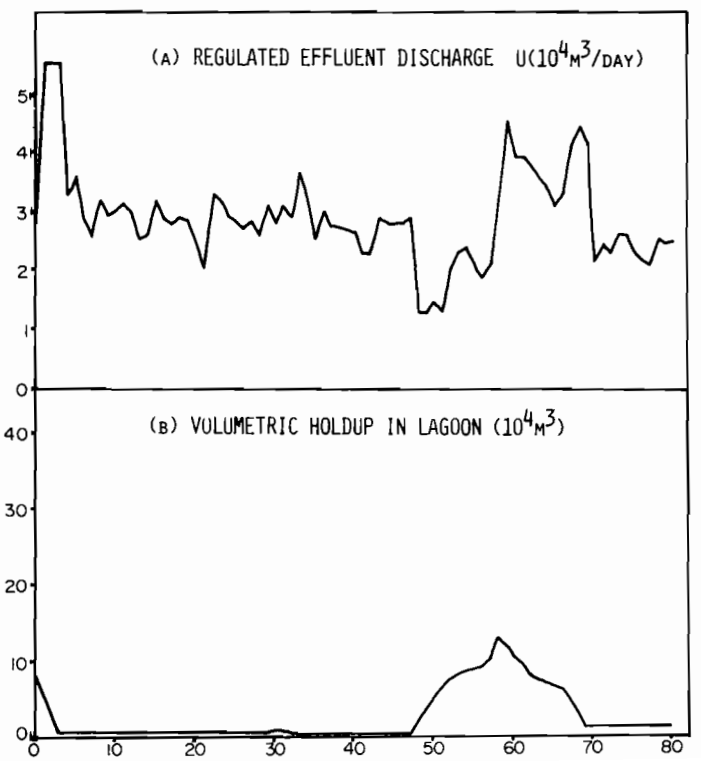
actually time varying and includes coefficients that change according to the current value of the control variable u(t). For design purposes, therefore, certain average values have to be assumed for u(t) and Q(t) as indicated in Table 11.1. The performance of the controller appears to be relatively insensitive to such parameter variations, although better designs have been demonstrated elsewhere (Young and Beck 1974).

- The analysis and simulation do not, however, include consideration of the significant diurnal variations that would result in practice from the photosynthetic/respiratory activity of plants and algae.

- The controller has been designed to deal with a regulator problem, i.e., we wish to cancel out the tendency for excessive BOD loads, as after the decay of the algal bloom, to depress the stream DO concentration below its desired set-point, namely 5.5 gm^{-3} here. When the river can receive a greater BOD level, for instance over the initial period, the controller acts so as to empty the lagoon (not exceeding a maximum discharge rate of 5.6×10^4 $m^3 day^{-1}$) but leaves a small residual volume in the lagoon when no control action is required.

- The controller is, of course, a feedback controller and requires only a measurement of the downstream DO concentration.

- Finally, in line with the analysis of (11.8), we might have chosen K(s) on the basis of making $|K(j\omega)G(j\omega)| >> |H(j\omega)|$ at those frequencies ω of which the disturbances $N(j\omega)$ are primarily composed. In fact, since in the real world it might be expected that the DO dynamics and measurements are subject to substantial high-frequency stochastic disturbances and errors, the absence



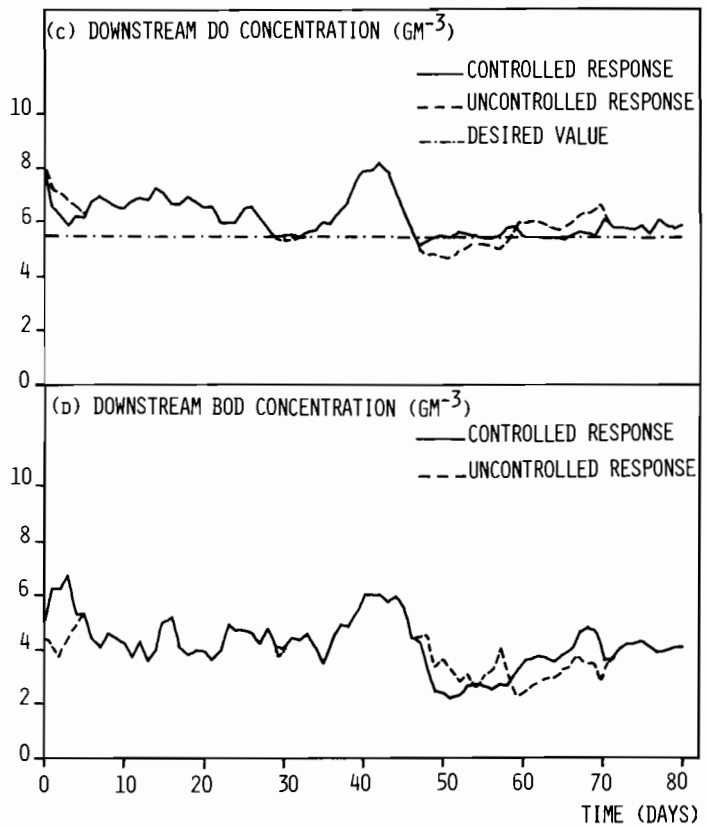


Fig. 11.10 Simulation results for in-stream DO control with a proportional and integral (PI) controller design: (A) the control variable, u, regulated effluent discharge (m^3day^{-1}); (B) volumetric hold-up in detention lagoon (m^3); (C) downstream DO concentration, x_1 (gm^{-3}) - the dashed line is the simulated uncontrolled response; (D) downstream BOD concentration, $x_2(gm^{-3})$ - the dashed line is the simulated uncontrolled response.

of derivative action in the controller is quite reasonable. While it can partially compensate for both the sluggish response of a PI controller and the "destabilizing" influence of integral action, derivative action has the drawback of amplifying high-frequency error signals fed to the controller which lead to excessive and undesirable fluctuations in the control variable $U(j\omega)$.

Although we have not discussed feedforward control at all in this example, the reader is referred to a parallel study by Whitehead (1976) for a discussion of feedforward-feedback control as it might be applied to artificial in-stream aeration. (An example of feedforward control is given in a later section entitled "Real-Time Forecasting.")

STATE ESTIMATION

A generalization of the single variable control problem to the multivariable case with several outputs $\underline{y}(t)$ leads to considerable potential complications in the synthesis of a controller. One of these additional complications is the identification of the system inputs $\underline{u}(t)$, $\underline{d}(t)$, or $\underline{n}(t)$ that dominate the behavior of $\underline{y}(t)$. It must also be decided if the control action taken to bring about the desired performance in one output is likely to degrade (to an unacceptable degree) the process performance as measured by another output variable. Such problems will not be dealt with here. Instead, this section tackles the obvious difficulties in coping with system and measurement error noise, to which we have alluded earlier, under the broad title of state estimation.

The principal components of state estimation are shown in Fig. 11.11. The concept

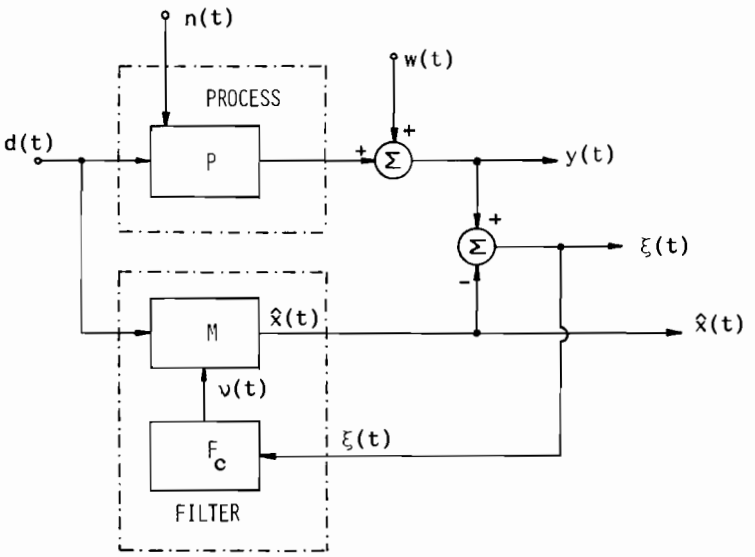


Fig. 11.11 The principal components of process state estimation and filtering; n(t) and w(t) are, respectively, system noise and random measurement errors.

of state has already been introduced implicitly in (11.15) where the state of river water quality is defined by its DO and BOD concentrations. State variable concepts are discussed fully in, for example, DeRusso et al. (1965) and Dorf (1965). Let us assume that the state of the system can only be measured in the presence of random measurement errors w(t). The filtering mechanism, which provides an estimate $\hat{x}(t)$ of the state, combines two functions: (a) the filter model M gives a simulated (predicted) response to the process input u(t); and (b) the corrective mechanism F_c manipulates the error between $\hat{x}(t)$ and y(t), the actual observation, to provide on-line updating corrections $\nu(t)$ to $\hat{x}(t)$. Precisely how the function F_c is designed and how the corrections $\nu(t)$ are presented to the model M is not of concern to us in the following. It is sufficient merely to say that F_c embodies measures of the statistical properties of n(t) and w(t) and the uncertainty in the accuracy of M as a model of P. Many other examples of filtering applications are described elsewhere, e.g., Todini and Bouillot (1976), and in Part One of this volume.

For control purposes the filter estimate $\hat{x}(t)$ would replace the function of y(t)

in the control law. Alternatively, the filter could facilitate short-term forecasts of the state of a process for the day-to-day supervisory management of, say, river basin water quality. The concept of filtering, in fact, generalizes to the idea of simultaneous state and parameter estimation, as in dynamic model structure identification (Beck and Young 1976, Beck 1979) or adaptive control schemes, and to the idea of state reconstruction.

Case Study 2 - State Reconstruction in the Activated Sludge Process of Wastewater Treatment

The activated sludge process is a biological process of wastewater treatment. The treatment is effected by a heterogeneous mixture of microorganism species that metabolize the complex organic waste substrates and decompose them into simple end-products such as carbon dioxide and water. Of considerable importance for process operation is a measure either of the biological activity of the organisms (see Olsson 1977 and Andrews et al. 1974) or of the magnitudes themselves of the various organism populations.

Consider, for example, the process of nitrification for the removal of ammonia from the influent sewage, whereby ammonia is oxidized in two stages to nitrate,

(11.21)

The two mediating species of bacteria, Nitrosomonas and Nitrobacter, operate in an aerobic environment provided by the aerator portion of the activated sludge unit (Fig. 11.12). The problem of state reconstruction, to which the filter will be

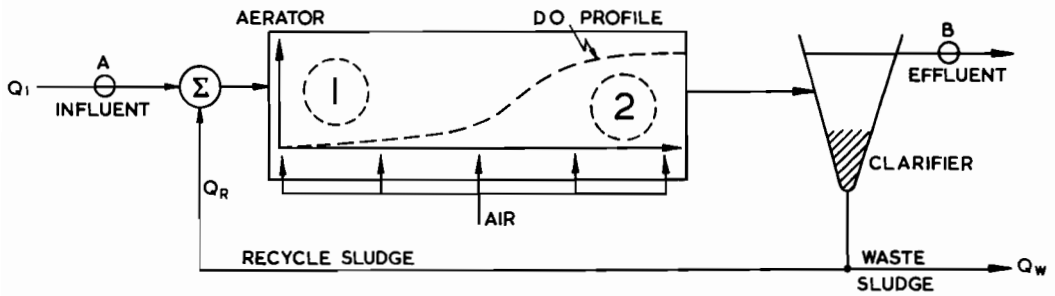


Fig. 11.12 A schematic diagram of the activated sludge process; zone 1 in the aerator is potentially an anaerobic environment; zone 2 is a predominantly aerobic environment.

applied, is as follows: given (noisy) measurements of the aerator influent/effluent substrate (ammonia, nitrite) and metabolic products (nitrite, nitrate), estimate the concentrations of Nitrosomonas and Nitrobacter in the aerator. In other words, state reconstruction is the on-line estimation of process states that are not directly measurable by instrumentation or rapid laboratory analysis. The outstanding example of such a variable that one might wish to estimate is the BOD concentration.

Taking component mass balances across the aerator gives a nonlinear dynamic model of nitrification (Poduska and Andrews 1975) with the general state-space form,

$$\dot{\underline{x}}(t) = \underline{f}\{\underline{x}(t), \underline{u}(t)\} + \underline{n}(t) \tag{11.22}$$

and with noisy observations $\underline{y}(t_k)$ being available at sampling instant t_k,

$$\underline{y}(t_k) = [I \;\vdots\; 0]\begin{bmatrix} \underline{x}_m(t_k) \\ \hdashline \underline{x}_u(t_k) \end{bmatrix} + \underline{w}(t_k) \tag{11.23}$$

Here $\underline{f}\{\cdot\}$ is a nonlinear vector function, $\underline{n}(t)$ and $\underline{w}(t_k)$ are, respectively, system disturbance and measurement noise terms; $\underline{x}_m$ is that component of the state vector for which measurements are available (ammonia, nitrite, nitrate) and $\underline{x}_u$ comprises those state variables that are not measurable, i.e., the Nitrosomonas and Nitrobacter population concentrations. Nonlinearities are present in the system dynamics because of the inclusion of a Monod (1942) characterization for bacterial growth kinetics. Nevertheless, it is possible to provide an approximate linear filter formulation for the nonlinear estimation problem of (11.22) and (11.23) such as an extended Kalman filter (EKF) (Jazwinski 1970). For this example the filtering algorithms differ slightly from those used by Moore and Weiss in Chapter 5 in two respects: the nonlinearities occur in the system dynamics and not in the observations equation, and a continuous-discrete version of the filter is used, i.e., continuous-time system dynamics with discrete-time observations. Illustrative results based on time-series data from the Norwich Sewage Works in England are reported in Beck (1977b).

REAL-TIME FORECASTING

It has already been mentioned that the concept of Kalman filtering generalizes to the notion of combined state and parameter estimation, such that, if so desired, a filter might be employed in an adaptive control and forecasting situation. Todini (1978) derived an adaptive forecasting algorithm that utilizes a linear Kalman filter in combination with a recursive instrumental variable parameter estimator (see, e.g., Young 1974). For discrete-time input/output model realizations a rather compact forecasting algorithm, termed an adaptive predictor, and otherwise known as a self-tuning predictor (Wittenmark 1974), can be implemented. This same algorithm is also applied to the river-flow forecasting problem in Chapter 9 by Ganendra. An analogous problem, as it were, in the field of water quality control is that of sewer flow prediction, for which we shall present the adaptive predictor.

Case Study 3 - Adaptive Prediction of Urban Sewer Flows

The input raw material to a wastewater treatment plant, the vector $\underline{v}_{23}$ in Fig. 11.2, exhibits large and generally poorly quantified variations with time (see Chapter 6 by Olsson). For proper operational control of the treatment plant, and hence, for control of the quality of the receiving river's water (Fig. 11.4), it would be extremely useful to have advance (short-term) predictions of the effluent flow from the sewer network, i.e., the influent flow to the plant. A particular feature of sewer-flow prediction that makes this problem somewhat different from that of river-flow prediction is the pronounced periodic nature of the sewage flow (see also Fig. 11.5 (top)). Thus, suppose that the relationship between these periodic patterns, $v_j(t_k)$, and the influent flow to the plant, $y(t_k)$, may be represented as,

$$A(q^{-1})y(t_k) = \sum_{j=1}^{m} B_j(q^{-1})v_j(t_k) + C(q^{-1})e(t_k) \tag{11.24}$$

in which q^{-1} denotes the backward shift operator,

$$q^{-1}\{y(t_k)\} = y(t_{k-1}) \text{ etc.}$$

and $e(t_k)$ is a sequence of independent, Gaussian $(0,\lambda^2)$ random variables. (As so often happens, notational conventions tend to conflict; the polynomial symbols here are not to be confused with the notation for certain matrices used earlier in this text.) $A(q^{-1})$, $B_j(q^{-1})$, $j = 1,2,\ldots,m$, and $C(q^{-1})$ are the polynomials,

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + \ldots + a_n q^{-n} \\ B_j(q^{-1}) &= b_{j0} + b_{j1} q^{-1} + \ldots + b_{jn} q^{-n} \quad ; j = 1,2,\ldots,m \\ C(q^{-1}) &= 1 + c_1 q^{-1} + \ldots + c_n q^{-n} \end{aligned} \tag{11.25}$$

Using the identity,

$$C(q^{-1}) = A(q^{-1})F(q^{-1}) + q^{-\tau}G(q^{-1}) \tag{11.26}$$

where

$$\begin{aligned} F(q^{-1}) &= 1 + f_1 q^{-1} + \ldots + f_{\tau-1} q^{-\tau+1} \\ G(q^{-1}) &= g_0 + g_1 q^{-1} + \ldots + g_{n-1} q^{-n+1} \end{aligned} \tag{11.27}$$

enbables a τ-step-ahead prediction, $\hat{y}(t_{k+\tau}|t_k)$ of $y(t_{k+\tau})$ to be made, following Aström (1970),

$$\hat{y}(t_{k+\tau}|t_k) = \frac{G(q^{-1})}{C(q^{-1})} y(t_k) + \frac{F(q^{-1})}{C(q^{-1})} \sum_{j=1}^{m} B_j(q^{-1}) v_j(t_{k+\tau}) \tag{11.28}$$

such that the variance of the prediction error,

$$\varepsilon(t_{k+\tau}) = y(t_{k+\tau}) - \hat{y}(t_{k+\tau}|t_k) \tag{11.29}$$

is minimized. The identity, (11.26), is a device for separating out the effects of (unknown) future realizations of $e(\cdot)$, namely $e(t_{k+\tau}),\ldots, e(t_{k+1})$, from the effects of current and past, and therefore known, realizations of $e(\cdot)$ on the future output behavior $y(t_{k+\tau})$. Equation (11.28) can be rearranged to give (at time t_k),

$$\begin{aligned} y(t_k) = G(q^{-1})y(t_{k-\tau}) &- (C(q^{-1})-1)\hat{y}(t_k|t_{k-\tau}) \\ &+ F(q^{-1}) \sum_{j=1}^{m} B_j(q^{-1}) v_j(t_k) + \varepsilon(t_k) \end{aligned} \tag{11.30}$$

which is more easily recognizable as a standard (predictor) model characterization suitable for, say, a recursive least-squares parameter estimation routine. The adaptive property of the self-tuning predictor is incorporated by the use of exponential weighting of past data (Wittenmark 1974), although other forms of time-varying parameter estimation (e.g., Young 1974) may indeed be more appropriate in sewer-flow prediction (Beck 1977a).

As an example of an adaptive predictor, it is found that for the Käppäla sewer system in Stockholm (Beck 1977a) the orders of the polynomials $A(q^{-1})$, $B_j(q^{-1})$, and $C(q^{-1})$ are given by $n = 1$ for a one-step-ahead predictor ($\tau = 1$). Hence, (11.30) reduces explicitly to,

$$y(t_k) = g_o y(t_{k-1}) - c_1 \hat{y}(t_{k-1}|t_{k-2}) + b_{10} v_1(t_k) + b_{11} v_1(t_{k-1}) + b_{20} v_2(t_k) + b_{21} v_2(t_{k-1}) + \epsilon(t_k) \quad . \tag{11.31}$$

where $v_1(\cdot)$ and $v_2(\cdot)$ are auxiliary variables drawn from two precomputed time series, of respectively, average weekly and daily dry-weather flow patterns. With their current estimates substituted for the parameters of (11.31), a one-step-ahead prediction can be made on the basis of (11.28), i.e.,

$$\hat{y}(t_{k+1}|t_k) = \hat{g}_o y(t_k) - \hat{c}_1 \hat{y}(t_k|t_{k-1}) + \hat{b}_{10} v_1(t_{k+1}) + \hat{b}_{11} v_1(t_k) + \hat{b}_{20} v_2(t_{k+1}) + \hat{b}_{21} v_2(t_k) \tag{11.32}$$

Typical results of the predictor are shown in Fig. 11.13.

If such predictions are to be of real benefit, we might consider the feasibility of somewhat longer-term forecasts and the use of rainfall measurements to improve prediction accuracy. The measurements of rainfall would be included in the predictor, (11.31) and (11.32), as a term(s) $v_3(t_k)$ etc. Without rainfall measurements, a four-hour-ahead predictor is found to be inadequate in its forecasts of the all-important high flows resulting from storm runoff (Beck 1977a). However, since the rainfall/influent flow dynamics of the Käppäla system exhibit a dead-time (transportation delay) of about three hours, the incorporation of rainfall measurements may well improve on such poor performance by exploiting this property of the sewer network. If there were sufficient confidence in the forecast, then it could easily be employed in a real-time control situation, although, as noted by Olsson (1977), a forecasting period of the order of eight to ten hours would ideally be required for treatment plant operations (i.e., routing and storage of flows) to be rearranged accordingly.

Nevertheless, we might generalize from our specific example in order to indicate briefly how flow forecasting can be combined with a feedforward controller. Suppose in (11.1) that the transfer functions E(s) and G(s) are given by,

$$E(s) = \gamma_1/(s+1) = Y(s)/D(s)$$
$$G(s) = \gamma_2 \exp(-\tau s)/(s+1) = Y(s)/U(s) \tag{11.33}$$

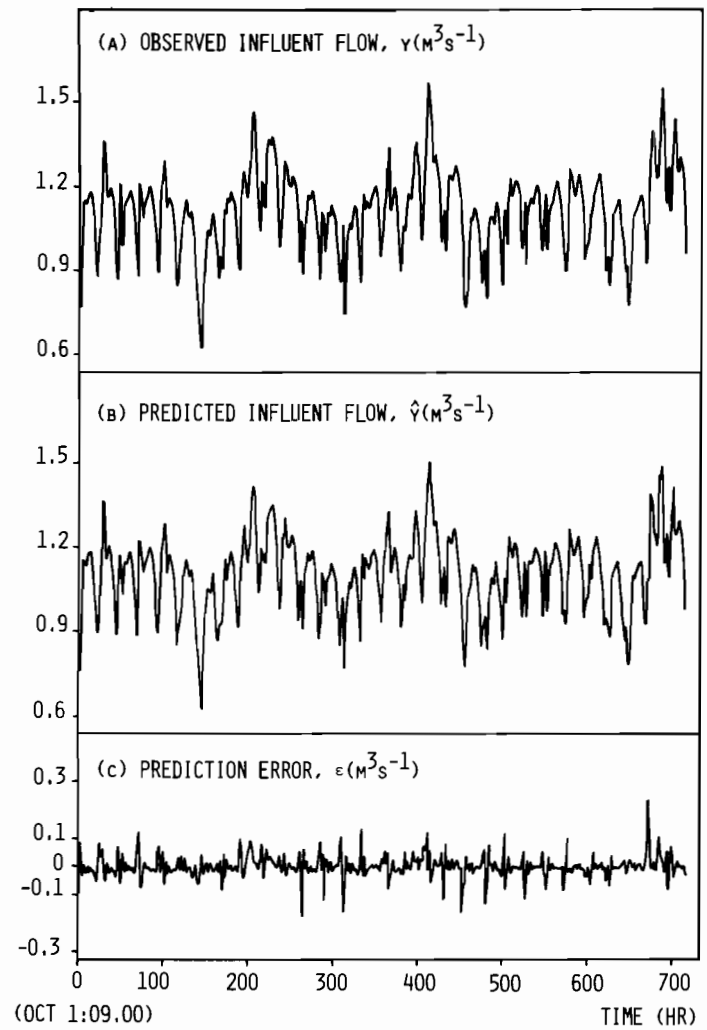


Fig. 11.13 The results of the one-step-ahead predictor for data for the month of October 1973 from the Käppäla sewer system in Stockholm: (A) observed influent flow $y(m^3sec^{-1})$; (B) predicted influent flow $\hat{y}(m^3sec^{-1})$; (C) prediction error $\varepsilon(m^3sec^{-1})$.

a physical interpretation of which is that the dynamic relationship between controlled output y(t) and control input u(t) contains a lead-time of magnitude τ. In other words, the disturbance d(t) effects a response in the output much more quickly than does the control action taken to cancel out the disturbance. Thus, from (11.2), the feedforward controller is specified by

$$F(s) = -E(s)/G(s) = -(\gamma_1/\gamma_2)\exp(\tau s) = U(s)/D(s) \qquad (11.34)$$

so that upon inverting the transforms,

$$u(t) = -(\gamma_1/\gamma_2)d(t+\tau) \qquad (11.35)$$

or, more realistically, if d(t) is not a known, deterministic function of time,

$$u(t) = -(\gamma_1/\gamma_2)\hat{d}(t+\tau|t) \qquad (11.36)$$

in which $\hat{d}(t+\tau|t)$ is a (continuous-time) prediction of $d(t+\tau)$ given all measurements up to and including the current time t. This kind of control problem arises naturally in potable water distribution networks (subsystem (1) in Fig. 11.2) where the pumping to and from service reservoirs, which may be situated a large distance from the location of demand, are scheduled in accordance with a prediction of future demand for supply (see, e.g., Fallside et al. 1975).

CONCLUSIONS

Real-time forecasting and control of river water quality is in its early stages of development. Primarily because of instrumentation problems, it is difficult to envisage a rapid translation of such techniques into the practice of river basin management. Yet the lack of instrumentation is a factor which should lend considerable impetus to the further exploration of estimation and forecasting as a means of supplementing routine operating information.

In contrast to flood prevention and drought alleviation, the management of river water quality suffers from a lack of clearly defined objectives. It is therefore, perhaps, these kinds of problems that should be solved before real-time forecasting applications can become of widespread value. Although considerable attention has been given here to the analysis, design, and illustration of automatic controllers, this is not necessarily how river water quality "control" should be interpreted for the immediate future. Rather, effective control may be carried out by a combination of well-presented routine operational data (including forecasts) and regulatory actions determined by agency managers on the basis of these data.

REFERENCES

Andrews, J.F. (1974). Dynamic models and control strategies for wastewater treatment processes. Water Res. 8: 261-289.

Andrews, J.F., H.O. Buhr, and M.K. Stenstrom (1974). Control systems for the reduction of effluent variability from the activated sludge process. Proc. International Conference on Effluent Variability from Wastewater Treatment Processes and its Control, I.A.W.P.R., Tulane/Vanderbilt Universities, New Orleans, Dec.

Aström, K.J. (1970). Introduction to Stochastic Control Theory. Academic Press, New York.

Aström, K.J., and P. Eykhoff (1971). System identification - A survey. Automatica 7: 123-162.

Beck, M.B. (1975). The identification of algal population dynamics in a freshwater stream, in G.C. Vansteenkiste, Ed., Computer Simulation of Water Resource Systems. North-Holland, Amsterdam, pp. 483-496.

Beck, M.B. (1976a). Dynamic Modelling and Control Applications in Water Quality Maintenance. Water Res. 10: 575-595.

Beck, M.B. (1976b). Identification and parameter estimation of biological process models, in G.C. Vansteenkiste, Ed., System Simulation in Water Resources. North Holland, Amsterdam, pp. 19-43.

Beck, M.B. (1977a). The identification and adaptive prediction of urban sewer flows. Int. J. Control 25: 425-440.

Beck, M.B. (1977b). Problems of river water quality modelling and control: A review, in H. Akashi, Ed., Preprints IFAC Symposium on Environmental Systems Planning, Design, and Control. Pergamon, Oxford, pp. 341-349.

Beck, M.B. (1977c). Model structure identification from experimental data, in E. Halfon, Ed., Theoretical Systems Ecology. Academic Press, New York, pp. 259-289.

Beck, M.B., and P.C. Young (1975). A dynamic model for DO-BOD relationships in a non-tidal stream. Water Res. 9: 769-776.

Beck, M.B., and P.C. Young (1976). Systematic identification of DO-BOD model structure. Proc. ASCE, J. Env. Eng. Div. 102: 909-927.

Berthoeux, P.M., W.G. Hunter, L.C. Pallesen, and C-Y. Shih (1975). Modelling sewage treatment plant input BOD data. Proc. ASCE, J. Env. Eng. Div. 101 (EE1): 127-138.

Buhr, H.O., J.F. Andrews, and T.M. Keinath, Eds. (1974). Research needs for automation of wastewater treatment systems. Proc. U.S. Environmental Protection Agency/Clemson University Workshop, Clemson, South Carolina.

Coughanowr, D.R., and L.B. Koppel (1965). Process Systems Analysis and Control. McGraw-Hill, New York.

DeRusso, P.M., R.J. Roy, and C.M. Close (1965). State Variables for Engineers. Wiley, New York.

Dorf, R.C. (1965). Time-domain Analysis and Design of Control Systems. Addison-Wesley, Reading, Massachusetts.

Dorf, R.C. (1973). Modern Control Systems. Addison-Wesley, Reading, Massachusetts.

Eykhoff, P. (1974). System Identification - Parameter and State Estimation. Wiley, London.

Fallside, F., P.F. Perry, and P.D. Rice (1975). On-line prediction of consumption for water supply network control. Proceedings 6th IFAC Congress, Boston, Massachusetts.

Gourishankar, V., and R.L. Lawson (1975). Optimal control of water pollution in a river system. Int. J. Syst. Sci. 6(3): 201-276.

Gourishankar, V., and K. Ramar (1977). Design of a water quality controller by pole assignment. IEEE Trans. Syst. Man Cybern. 7(2): 109-112.

Huck, P.M., and G.J. Farquhar (1975). Water quality models using the Box-Jenkins method. Proc. ASCE, J. Env. Eng. Div. 100(EE3): 733-752.

Institution of Chemical Engineers (1974). Water Borne Waste. Supplement to The Chemical Engineer, Oct.

Jazwinski, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Koivo, H.N. (1977). Robust decentralized control of water quality in a river system, in H. Akashi, Ed., Preprints IFAC Symposium on Environmental Systems Planning, Design, and Control. Pergamon, Oxford, pp. 365-368.

Lin, S.H. (1975). Feedforward water quality control by aeration of a thermally and organically polluted stream. Int. J. Syst. Sci. 6(10): 919-932.

MacFarlane, A.G.J. (1976). Feedback (1976 Institute of Measurement and Control Thomson Lecture). Measurement and Control 9: 446-462.

Monod, J. (1942). Recherche sur la croissance des cultures bacteriennes. Hermann, Paris.

Olsson, G. (1977). State of the art in sewage treatment plant control. AIChE Symp. Ser. 72(159): 52-76.

Özgören, M.K., R.W. Longman, and C.A. Cooper (1975). Optimal control of a distributed parameter time-lagged river aeration system. Trans. ASME, Ser. G., J. Dyn. Syst. Meas. Control 97(2): 164-171.

Penumalli, B.R., R.H. Flake, and E.G. Fruh (1976). Large-scale estuarine water quality matrix model. Proc. ASCE, J. Env. Eng. Div. 102(EE1): 191-209.

Poduska, R.A., and J.F. Andrews (1975). Dynamics of nitrification in the activated sludge process. J. Water Pollut. Control Fed. 47: 2599-2619.

Powers, W.F., and R.P. Canale (1975). Some applications of optimization techniques to water quality modelling and control. IEEE, Trans. Syst. Man Cybern. 5(3): 312-321.

Streeter, H.W., and E.B. Phelps (1925). A study of the pollution and natural purification of the Ohio River. Bulletin No. 146, U.S. Public Health Service.

Todini, E. (1978). Mutually interactive state/parameter (MISP) estimation in hydrological applications, in G.C. Vansteenkiste, Ed., Modeling, Identification, and Control in Environmental Systems. North-Holland, Amsterdam, pp. 699-713.

Todini, E., and D. Bouillot (1976). A rainfall-runoff Kalman filter model, in G.C. Vansteenkiste, Ed., System Simulation in Water Resources. North-Holland, Amsterdam, pp. 69-81.

Water Resources Board (1974). Water Resources in England and Wales (Publication No. 22). Her Majesty's Stationery Office, London.

Whitehead, P.G. (1976). A control and systems analysis of artificial in-stream aeration, in Lecture Notes in Computer Science 40. Springer-Verlag, Berlin, pp. 207-228.

Williams, B.J., and J.B. Hinwood (1976). Two-dimensional mathematical water quality model. Proc. ASCE, J. Env. Eng. Div. 102 (EE1): 149-163.

Wittenmark, B. (1974). A self-tuning predictor. IEEE Trans. Autom. Control: 19(6): 848-851.

Young, P.C. (1974). A recursive approach to time-series analysis. Bull. Inst. Math. Appl. 209-224.

Young, P.C., and M.B. Beck (1974). The modelling and control of water quality in a river system. Automatica 10(5): 455-468.

12 Bayesian Strategies for Controlling Dynamic Water Resources

A. Szöllösi-Nagy and E.F. Wood

In most real-time application studies one of the main problems is to find an optimal control policy when there is little or unreliable knowledge about system structure or parameters. The system should be able to adapt to variations in structure or parameters and we wish to construct control policies that take these changes into account. A policy that considers these uncertainties will be defined as a Bayesian control policy. Loosely speaking, the Bayesian approach to the optimal control problems requires the assumption of an a priori probability distribution function for the unknown parameters. These distribution functions are updated by Bayes rule, given current information on the system's behavior.

This paper surveys some techniques which seem to be applicable to the adaptive real-time control of water resources systems. First, the basis of adaptive estimation and control for nonlinear dynamics systems is described. Weighting coefficients for the nonlinear case are derived in which the learning nature of the algorithms is described through a recursive Bayesian relation. Adaptive estimation and control for linear systems is discussed, and weighting coefficients for this case are derived.

ADAPTIVE ESTIMATION AND CONTROL FOR NONLINEAR DYNAMIC SYSTEMS

Assume that we are given a discrete nonlinear stochastic dynamic system described by

$$\underline{x}(t+1) = \underline{f}(\underline{x}(t), \underline{u}(t), \underline{w}(t)), \quad t = 0, 1, \ldots, N-1$$

where $\underline{x}(\cdot)$ is a vector of state variables, $\underline{u}(\cdot)$ is a vector of control variables and $w(\cdot)$ is a zero-mean, Gaussian white noise (GWN) sequence, known as systems noise, with covariance

$$E[w(t) \cdot w^T(\tau)] = Q(t)\delta_{t\tau}$$

The corresponding measurement (or output) equation is

$$\underline{z}(t) = \underline{h}(x(t), \underline{v}(t))$$

where $v(\cdot)$ is a zero-mean GWN sequence known as the measurement error with covariance

$$E[v(t)v^T(\tau)] = R(t)\delta_{t\tau}$$

Moreover, the noise processes are assumed to be independent of each other.

In filtering problems (see Chapter 2 and the Introduction to Part One), we saw that the optimal filtered estimate of the state is given by

$$\underline{\hat{x}}(t|t) = E[\underline{x}(t)|Z_t] \ , \tag{12.1}$$

the conditional expectation of the state given the measurement sequence Z_t. In other words

$$\underline{\hat{x}}(t|t) = \int_X \underline{x}(t)p(\underline{x}(t)|Z_t)dx(t) \ , \tag{12.2}$$

where X is the finite-dimensional state space and $p(\underline{x}(t)|Z_t)$ is the conditional pdf of the states. There are several, in most cases approximate, ways to evaluate (12.2) for nonlinear systems: extended Kalman filtering (Jazwinski 1966), Gaussian sum approximations (Aslpach and Sorrenson 1972) or using spline functions (Lainiotis 1974).

We assume that the system is characterized by a vector of unknown but constant discrete parameters $\underline{\theta}$ belonging to a finite dimensional parameter space Θ. [The case of a continuous parameter space is treated by Lainiotis (1974)]. The estimated states depend on this parameter vector. Equation (12.2) can be written as

$$\begin{aligned} \underline{\hat{x}}(t|t) &= \int_X \underline{x}(t) \int_\Theta p(\underline{x}(t), \underline{\theta}|Z_t)d\underline{\theta}\ d\underline{x}(t) \\ &= \int_X \underline{x}(t) \int_\Theta p(\underline{x}(t)|\underline{\theta}, Z_t)p(\underline{\theta}|Z_t)d\underline{\theta}\ d\underline{x}(t) \end{aligned} \tag{12.3}$$

Interchanging the order of integration, which is permissible as long as the integrand is absolutely integrable, and defining the $\underline{\theta}$-conditional estimate,

$$\begin{aligned} \underline{\hat{x}}(t|t,\underline{\theta}) &\triangleq E[\underline{x}(t)|\underline{\theta}, Z_t] \\ &= \int_X \underline{x}(t)\ p\ (\underline{x}(t)|\underline{\Theta}, Z_t)d\underline{x}(t) \end{aligned} \tag{12.4}$$

the optimal adaptive estimate is given by Magill (1965) as

$$\underline{\hat{x}}(t|t) = \int_\Theta \underline{\hat{x}}(t|t,\underline{\theta})\ p\ (\underline{\theta}|Z_t)d\underline{\theta} \tag{12.5}$$

i.e., it is formed by taking the complete set of $\underline{\theta}$-conditional estimates, weighting each with the conditional probability of the appropriate parameter vector, given

the measurement sequence Z_t, and integrating over the space of all possible parameter values. It is assumed that $\underline{\theta}$ has an a priori probability density function $p(\underline{\theta})$, i.e., the weighting coefficient $p(\underline{\theta}|Z_t)$ is the a posteriori pdf of the parameter vector given the measurement sequence Z_t. The concept is illustrated in Fig. 12.1 for discrete parameters. In this case, the parameter

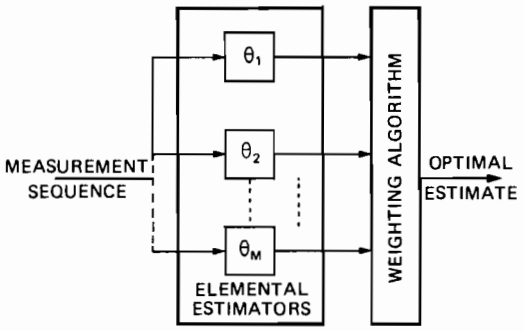


Fig. 12.1 Discrete parameter adaptive estimation.

space is spanned by M discrete values, $\theta_1,\ldots,\theta_M$, and the integration in (12.5) is replaced by a sum over the space. The a priori discrete pdf $p(\underline{\theta})$ for $\underline{\theta}$ is an M-vector with the components

$$0 \leq p_i(\theta) = \text{Prob}[\theta=\theta_i] \leq 1, \quad i=1,\ldots,M$$

satisfying

$$\sum_{i=1}^{M} p_i(\theta) = 1 \tag{12.6}$$

The optimal estimate is given now by

$$\underline{\hat{x}}(t|t) = \sum_{i=1}^{M} \underline{\hat{x}}(t|t,\theta_i)p_i(\underline{\theta}|Z_t) \tag{12.7}$$

Equation (12.5) immediately follows from the smoothing property of conditional expectations (Doob 1953):

$$E[\underline{x}|Z] = E[E(\underline{x}|Z)|Z] \tag{12.8}$$

By similar arguments, one can easily define the adaptive control, which is essentially a feedback type of control (Sworder 1966), for equivalent systems, as follows

$$\underline{u}(t) = \int_{\Theta} \underline{u}(t|\underline{\theta})\ p(\underline{\theta}|Z_t)d\underline{\theta}\ , \tag{12.9}$$

where the $\underline{\theta}$-conditional control is given by

$$\underline{u}(t|\underline{\theta}) = \underline{L}[\underline{\hat{x}}(t|t),\underline{\theta}] \tag{12.10}$$

In order to determine the optimal adaptive estimate/control, one needs to know the weighting coefficients, and, for control, the solution of the corresponding stochastic control problems. Therefore, the problems of obtaining optimal control policies are generally much harder for adaptive systems than for purely stochastic systems.

DERIVATION OF THE WEIGHTING COEFFICIENTS FOR THE NONLINEAR CASE

The conditional pdf of the parameter vector $\underline{\theta}$, given the measurement sequence Z_t, is defined by

$$p(\underline{\theta}|Z_t) = \frac{p(\underline{\theta},Z_t)}{p(Z_t)}$$

$$= \frac{p(\underline{\theta},Z_{t-1},\underline{z}(t))}{p(Z_{t-1},\underline{z}(t))} ,$$

or, according to the chain rule of conditional probabilities, as

$$p(\underline{\theta}|Z_t) = \frac{p(\underline{z}(t)|\underline{\theta},Z_{t-1})p(\underline{\theta}|Z_{t-1})p(Z_{t-1})}{p(Z_{t-1}, \underline{z}(t))p(\underline{z}(t))} \tag{12.11}$$

Since

$$p(Z_{t-1}|\underline{z}(t)) = \int_{\Theta} p(\underline{\theta},Z_{t-1}|\underline{z}(t))d\underline{\theta}$$

and

$$p(\underline{\theta},Z_{t-1}|\underline{z}(t)) = \frac{p(\underline{\theta},Z_{t-1},\underline{z}(t))}{p(\underline{z}(t))}$$

we get for the marginal distribution

$$p(Z_{t-1}|\underline{z}(t)) = \frac{1}{p(\underline{z}(t))} \int_{\Theta} p(\underline{\theta},Z_{t-1},\underline{z}(t))d\underline{\theta}$$

Substituting this into the denominator of (12.11) and considering that

$$p(\underline{\theta},Z_{t-1},\underline{z}(t)) = p(\underline{z}(t)|\underline{\theta},Z_{t-1})p(\underline{\theta}|Z_{t-1})p(Z_{t-1})$$

we have

$$p(\underline{\theta}|Z_t) = \frac{p(\underline{z}(t)|\underline{\theta},Z_{t-1})}{\int_{\Theta} p(\underline{z}(t)|\underline{\theta},Z_{t-1})p(\underline{\theta}|Z_{t-1})d\underline{\theta}} \tag{12.12}$$

which is in fact a recursive Bayesian algorithm for the calculation of the weighting coefficients with the initial condition $p(\underline{\theta}|Z_0) = p(\underline{\theta})$.

For the case when the parameter space Θ is discrete and consists of M elements, the a priori pdf is

$$p(\underline{\theta}) = \sum_{i=1}^{M} p(\theta_i)\delta(\theta - \theta_i)$$

and the a posteriori pdf is given by

$$p(\underline{\theta}|Z_t) = \sum_{i=1}^{M} p(\theta_i|Z_t)\delta(\theta - \theta_i)$$

The recursive algorithm becomes

$$p(\theta_i|Z_t) = \frac{\Lambda(t|\theta_i)}{\sum_{j=1}^{M} \Lambda(t|\theta_j)p(\theta_j|Z_{t-1})} p(\theta_i|Z_{t-1}) \qquad (12.13)$$

where the likelihood function $\Lambda(t|\theta_i)$ stands for $p(\underline{z}(t)|\theta_i, Z_{t-1})$.

Remark

If the assumption of time-invariant parameters is relaxed, i.e., when

$$\underline{\theta}(t) = \underline{g}(\underline{\theta}(t-1)) \ , \ \forall \underline{\theta}(t) \in \Theta$$

the parameters are related by a nonsingular, nonlinear mapping $\underline{g}(\cdot)$, then we can proceed along the same lines as outlined above. This means that the recursive functional relation for the a posteriori pdf is

$$p(\underline{\theta}(t)|Z_t) = \frac{p(\underline{z}(t)|\underline{\theta}(t), Z_{t-1})}{\int_{\Theta} p(\underline{z}(t)\ \underline{\theta}(t), Z_{t-1})p(\underline{\theta}(t)\ Z_{t-1})d\underline{\theta}(t)} p(\underline{\theta}(t)|Z_{t-1})$$

though we need additional steps to make the projection $p(\underline{\theta}(t-1)|Z_{t-1}) \rightarrow p(\underline{\theta}(t)|Z_{t-1})$. This projection is given by

$$p(\underline{\theta}(t)|Z_{t-1}) = \int_{\Theta} p(\underline{\theta}(t-1)|Z_{t-1})p(\underline{\theta}(t)|\underline{\theta}(t-1), Z_{t-1})d\underline{\theta}(t-1),$$

where the conditional pdf $p(\underline{\theta}(t)|\underline{\theta}(t-1), Z_{t-1})$ must be evaluated. This evaluation is given by

$$p(\underline{\theta}(t)|\underline{\theta}(t-1), Z_{t-1}) = \frac{p(\underline{\theta}(t), \underline{\theta}(t-1)|Z_{t-1})}{p(\underline{\theta}(t-1)|Z_{t-1})}$$

where the denominator is computed as in (12.13). From the theory of derived distributions

$$p(\underline{\theta}(t),\underline{\theta}(t-1)|Z_{t-1}) = p(\underline{\theta}(t),\underline{g}^{-1}(\underline{\theta}(t))|Z_{t-1})J ,$$

where

$$J = \frac{\partial \underline{g}^{-1}(\underline{\theta}(t))}{\partial\, \underline{\theta}(t)}$$

is the Jacobian. The evaluation of $p(\underline{\theta}(t), \underline{g}^{-1}(\underline{\theta}(t))|Z_{t-1})$ can be accomplished analytically by knowing $\underline{g}(\cdot)$. That is, we can construct a recursive Bayesian algorithm even when the parameters are time varying. For the case when one knows that the parameters have changed but does not know about the direction of this change in the parameter space, that is, $\underline{g}(\underline{\theta}(t-1))$ cannot be formulated, Peterka (1976a) has derived a recursive algorithm.

ADAPTIVE ESTIMATION AND CONTROL FOR LINEAR DYNAMIC SYSTEMS

Consider now the stochastic LQP-control problem when the parameters in the linear system model are unknown; the state and measurement equations become

$$\underline{x}(t+1) = \Phi(t+1,t;\underline{\theta})\underline{x}(t) + \Gamma(t,\underline{\theta})\underline{u}(t) + \underline{w}(t) \tag{12.14}$$

$$\underline{z}(t) = H(t,\underline{\theta})\underline{x}(t) + \underline{v}(t) , \tag{12.15}$$

where $\underline{\theta}$ is a q-vector of unknown parameters in Φ, Γ, H, Q, R, $\underline{x}(0)$ belonging to a finite-dimensional parameter space Θ. It is assumed that $\underline{\theta}$ is time invariant and has an *a priori* pdf $p(\underline{\theta})$. Of course, when the model parameters are exactly known, the problem is that of linear stochastic optimal control and can be solved by taking advantage of the separation theorem discussed in the Introduction to Part Two. For uncertain parameters, however, the problem becomes much more complicated since the joint estimation of the uncertain states or parameters becomes nonlinear even in the case of linear system dynamics. This means that the linear adaptive control problem, in fact, is a nonlinear stochastic control problem. Therefore, one cannot expect that the separation property discussed in the Introduction to Part Two will apply here. Nevertheless, it will be shown that a kind of separation property can be found even in this case which makes the implementation of a feedback scheme relatively easy.

Again, we are seeking a feedback control policy such that for each t=0,1,...,N-1, the sequence $\{\underline{u}(t), \ldots,\underline{u}(N-1)\}$ will minimize the performance measure

$$J(t) = \min_{u(t),\ldots,\underline{u}(N-1)} E\Big[||\underline{x}(N)||^2_A + \sum_{\tau=t}^{N-1} \big[||\underline{x}(\tau)||^2_{B(t)} + ||\underline{u}(\tau)||^2_{C(t)}\big]\Big|Z_t\Big] \tag{12.16}$$

i.e., J(t) is the optimal expected cost-to-go, conditioned by Z_t. Using the smoothing property of conditional expectations (12.8), we have

$$J(t) = \min_{\underline{u}(t),\ldots,\underline{u}(N-1)} E\Big[E(||\underline{x}(N)||_A^2 + \sum_{\tau=t}^{N-1} [||\underline{x}(\tau)||_{B(t)}^2 + ||\underline{u}(\tau)||_{C(t)}^2] | \underline{\theta}, Z_t) | Z_t\Big] \tag{12.17}$$

If we could find a closed-form solution for $\underline{u}(t)$, then we would have the optimal feedback solutions; but as Deshpande *et al.* (1973) pointed out, this is not feasible. Assuming, however, that the minimization and the first expected value operation can be interchanged [for the conditions of the interchange see Meier *et al.* (1971)] we can arrive at a solution. We have

$$\tilde{J}^*(t) = E\left[\min_{\underline{u}(t),\ldots,\underline{u}(N-1)} E(||\underline{x}(N)||_A^2 + \sum_{\tau=t}^{N-1} [||\underline{x}(\tau)||_{B(t)}^2 + ||\underline{u}(\tau)||_{C(t)}^2] \mid \underline{\theta}, Z_t) | Z_t\right] \tag{12.18}$$

which is the weighted average of the optimal expected cost-to-go, (12.16); the weights are the *a posteriori* pdf of $\underline{\theta}$. For each $\underline{\theta} \in \Theta$ we perform the minimization in (12.18) subject to (12.14) and (12.15); the minimization is, in fact, a stochastic LQP-control problem. For each $\underline{\theta} \in \Theta$ we obtain a corresponding optimal control $\underline{u}^*(t|\underline{\theta})$, which is a $\underline{\theta}$-conditional control

$$\underline{u}^*(t|\underline{\theta}) = -L(t,\underline{\theta})\hat{\underline{x}}(t|t,\underline{\theta}) \ , \tag{12.19}$$

where the conditional feedback control gain matrix $L(t,\underline{\theta})$ is computed as in the deterministic control problem and $\hat{\underline{x}}(t|t,\underline{\theta})$ is computed via the Kalman filter; of course the Φ, Γ, H, Q,R matrices are $\underline{\theta}$-dependent in this case.

The optimal adaptive control is then generated using (12.9), as long as the *a posteriori* pdf is known.

DERIVATION OF THE WEIGHTING COEFFICIENTS FOR THE LINEAR CASE

The weighting coefficients, i.e., the *a posteriori* pdf for the parameter vectors, for the linear case may be obtained from the general Bayesian recursive algorithm (12.12). Owing to the special linear structure, however, we can express the likelihood function $p(\underline{z}(t)|\underline{\theta},Z_{t-1})$ in terms of the corresponding innovation sequences ν. It is known from the innovation theory (Kailath 1968) that if we are given a stochastic process $\{\underline{z}(t) : t \in T\}$ as a GWN process, $\underline{z}(\cdot)$ can be calculated from $\nu(\cdot)$ by a causal (i.e., non-anticipative) and causally invertible transformation. The point is that $\nu(\cdot)$ and $\underline{z}(\cdot)$ contain the same "statistical information" since we can go back and forth in real-time from one process to the other, but, of course, $\nu(\cdot)$ will generally be a much simpler process than $\underline{z}(\cdot)$. Moreover, since the values of $\nu(\cdot)$ at different time instants are statistically independent of each other, each observation $\nu(t)$ brings "new information" only, unlike the observation z(t) which is, in general, statistically related to past values of $\underline{z}(\cdot)$. $\nu(\cdot)$ is thus called "new information" or the "innovation" process of $\underline{z}(\cdot)$.

Therefore, the $\underline{\theta}$-conditional innovation is defined as

$$\nu(t,\underline{\theta}) = \underline{z}(t) - \underline{\hat{z}}(t|t-1,\underline{\theta})$$

$$= \underline{z}(t) - E\{H(t,\underline{\theta})\underline{x}(t) + \underline{v}(t)|Z_{t-1}\}$$

$$= \underline{z}(t) - H(t,\underline{\theta})\underline{\hat{x}}(t|t-1,\underline{\theta})$$

since it is part of the measured output that contains some information not previously available. So, we can replace $\underline{z}(\cdot)$ by $\nu(\cdot)$ and according to the theory of derived distributions, we have

$$p(\underline{\theta},\underline{z}(t)|Z_{t-1}) = p(\underline{\theta},\nu(t,\underline{\theta})|Z_{t-1})J ,$$

where due to the linearity, the Jacobian is equal to

$$J = \left[\frac{\partial(\underline{z}(t) - H(t,\underline{\theta})\underline{x}(t|t-1,\underline{\theta}))}{\partial\underline{z}(t)}\right] = I ,$$

the identity matrix. Since the innovation process is a GWN process and is independent of the previous measurements Z_{t-1}, we have on the one hand

$$p(\underline{\theta},\underline{z}(t)|Z_{t-1}) = p(\underline{\theta}|Z_{t-1})p(\nu(t,\underline{\theta}))$$

and on the other

$$p(\underline{\theta},\underline{z}(t)|Z_{t-1}) = p(\underline{z}(t),\underline{\theta}|Z_{t-1})$$

$$= p(\underline{z}(t)|\underline{\theta},Z_{t-1})p(\underline{\theta}|Z_{t-1})$$

Combining the above two expressions, we get for the likelihood function

$$p(\underline{z}(t)|\underline{\theta},Z_{t-1}) = p(\nu(t,\underline{\theta})) = p_{\nu}(\underline{z}(t) - H(t,\underline{\theta})\underline{x}(t|t-1,\underline{\theta}))$$

It is easy to see that the innovations form a zero-mean GWN process with covariance $P_{\nu}(t,\underline{\theta})$, $\nu(t,\underline{\theta}) \sim N(0,P_{\nu}(t,\underline{\theta}))$. The covariance matrix can be readily determined from an equivalent representation of

$$\nu(t,\underline{\theta}) = -H(t,\underline{\theta})\tilde{\underline{x}}(t|t-1,\underline{\theta}) + \underline{v}(t) ,$$

where

$$\tilde{\underline{x}}(t|t-1,\underline{\theta}) = \underline{x}(t) - \hat{\underline{x}}(t|t-1,\underline{\theta})$$

is the <u>one-step-ahead prediction error</u> since

$$P_{\nu}(t,\underline{\theta}) = \text{cov}[\nu(t,\underline{\theta}),\nu(t,\underline{\theta})] = E\{\nu(t,\underline{\theta})\nu^T(t,\underline{\theta})\}$$

$$= E\{[-H(t,\underline{\theta})\tilde{\underline{x}}(t|t-1,\underline{\theta}) + \underline{v}(t)]\ [\cdot]^T\}$$

$$= E\{H(t,\underline{\theta})\tilde{\underline{x}}(t|t-1,\underline{\theta})\tilde{\underline{x}}^T(t|t-1,\underline{\theta})H^T(t,\underline{\theta})\}$$

$$+ E\{\underline{v}(t)\underline{v}^T(t)\}$$

$$= H(t,\underline{\theta})P(t|t-1,\underline{\theta})H^T(t,\underline{\theta}) + R(t,\underline{\theta})$$

where

$$P(t|t-1,\underline{\theta}) = E[\tilde{\underline{x}}(t|t-1,\underline{\theta})\tilde{\underline{x}}^T(t|t-1,\underline{\theta})]$$

is the $\underline{\theta}$-conditional error covariance matrix.

Summing up, we have the following recursive relation for calculating the weighting coefficients

$$p(\underline{\theta}|Z_t) = \frac{|P_{\nu}(t,\underline{\theta})|^{-1/2}\exp[-1/2||\nu(t,\underline{\theta})||^2_{P_{\nu}^{-1}(t,\underline{\theta})}]p(\underline{\theta}|Z_{t-1})}{\int_{\theta} |P_{\nu}(t,\underline{\theta})|^{-1/2}\exp[-1/2||\nu(t,\underline{\theta})||^2_{P_{\nu}^{-1}(t,\underline{\theta})}]p(\underline{\theta}|Z_{t-1})d\theta}$$

where

$$P_{\nu}(t,\underline{\theta}) = H(t,\underline{\theta})P(t|t-1,\underline{\theta})H^T(t,\underline{\theta}) + R(t,\underline{\theta})$$

$$\nu(t,\underline{\theta}) = \underline{z}(t) - H(t,\underline{\theta})\hat{\underline{x}}(t|t-1,\underline{\theta})$$

and the initial conditions are

$$p(\underline{\theta}|Z_0) = p(\underline{\theta}) \text{ , for the } \underline{\text{a posteriori}} \text{ pdf, and}$$

$$\Lambda(0|\underline{\theta}) = 1 \text{ , for the likelihood function}$$

Remark

The state estimation error covariance matrix $P(t|t)$, which is useful for the on-line evaluation of the estimation performance, is given by Lainiotis (1974) as

$$P(t|t) = \int_{\Theta} \{P(t|t,\underline{\theta}) + [\hat{\underline{x}}(t|t,\underline{\theta}) - \hat{\underline{x}}(t|t)][\hat{\underline{x}}(t|t,\underline{\theta})$$

$$- \hat{\underline{x}}(t|t)]^T\} \cdot p(\underline{\theta}|Z_t)d\underline{\theta}$$

The block diagram of the optimal adaptive estimation algorithm is depicted in Fig. 12.2.

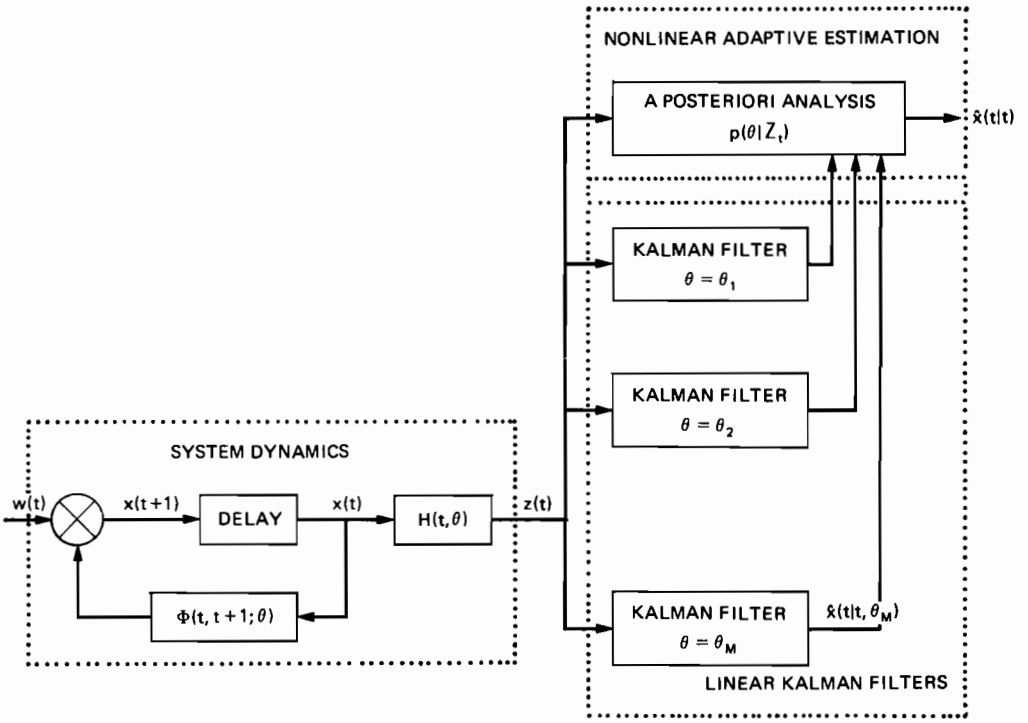


Fig. 12.2 Block diagram of the nonlinear adaptive sequential estimation algorithm for a discrete linear stochastic system with discrete parameters.

The adaptive control can be obtained by utilizing the above equations and is nonlinear. Since for each $\underline{\theta} \in \Theta$ we have to solve a $\underline{\theta}$-conditional stochastic LQP problem and then synthesize the optimal adaptive control, there must be a kind of separation even in this case. The algorithm divides the adaptive control into a linear nonadaptive part, i.e., the set of $\underline{\theta}$-conditional linear controls, and a nonlinear part, i.e., the a posteriori pdfs that incorporate the learning nature of the adaptive control. We mention here that for nonlinear system dynamics Lainiotis's (1974) partition theorem can be applied.

Remark

The notation of structural adaptation can also be imbedded into the above algorithms. We need only fix an upper bound $\tilde{n}$ of the system dimensionality and augment the parameter vector θ up to $\tilde{n}$. Then the adaptive algorithms automatically give an estimate for the (changing) system order. In other words, it gives zero elements for the "superfluous" parameter values.

DISCUSSION AND SUMMARY

There are several pertinent questions related to the use of the Bayesian approach:

- How do we choose the a priori pdf $p(\underline{\theta})$ for the uncertain parameters $\underline{\theta}$?

- Under what conditions does the sequence of a posteriori pdf $p(\underline{\theta}|Z_t)$ computed by Bayes rule converge to the true distribution as $t \to \infty$?

- Do the different choices of a priori pdf converge eventually to the same distribution?

- What is the effect of a particular choice of an a priori pdf on the performance measure of the system?

The first question has been a source of various criticisms of the Bayesian approach. So far, no rational procedure has been proposed to choose the a priori pdf, although it is advocated that pdfs of self-reproducing type, or in other words, conjugate distributions, should be chosen mainly for algorithmic reasons (Peterka 1976b). The second question is intimately related to the third and, apparently, both of them can be answered satisfactorily by utilizing some martingale convergence theorems (Doob 1953). To answer the fourth question, Aoki (1967) derived lower bounds and Hawkes and Moore (1976) derived upper bounds to show the effect of the controller's ignorance on the performance measure.

As far as the numerical calculation of the a posteriori pdfs is concerned, one can use the iterative algorithm developed by Sengbush and Lainiotis (1969) based on the quantization of the parameter space. If the parameter space is of high dimension, a Monte Carlo technique might be appropriate to evaluate the weighting coefficients.

REFERENCES

Alspach, D.L., and H.W. Sorrenson (1972). Non-linear Bayesian estimation using gauss sums. Automatica 7: 465-479.

Aoki, M. (1967). Optimization of Stochastic Systems. Academic Press, New York.

Deshpande, J.G., T.N. Upadhyay, and D.G. Lainiotis (1973). Adaptive control of linear stochastic systems. Automatica 9: 107-115.

Doob, J. (1953). Stochastic Processes. Wiley, New York.

Jazwinski, A. (1966). Filtering for nonlinear dynamic systems. IEEE Trans. Autom. Control 11: 765.

Kailath, T. (1968). An innovations approach to least-squares estimation, Part 1: Linear filtering in additive white noise. IEEE Trans. Autom. Control 13: 646-655.

Lainiotis, D.G. (1971). Optimal adaptive estimation: Structure and parameter estimation. IEEE Trans. Autom. Control 16: 260-270.

Lainiotis, D.G. (1974). Partitioned estimation algorithms, I. Nonlinear estimation. D.G. Lainiotis, Ed., in Estimation Theory. Elsevier, New York.

Magill, D.T. (1965). Optimal adaptive estimation of sampled stochastic processes. IEEE Trans. Autom. Control 10(4): 434-439.

Meier, L., R.E. Larson, and A.J. Tether (1971). Dynamic programming for stochastic control of discrete systems. IEEE Trans. Autom. Control 16(6): 767-775.

Peterka, V. (1976a). Subjective probability approach to real-time identification. Fourth IFAC Symp. on Identification and System Parameter Estimation, Tblisi, USSR.

Peterka, V. (1976b). Experience accumulation for decision making in multivariate time series. RM-76-70, International Institute for Applied Systems Analysis, Laxenburg, Austria.

Sengbush, R.L., and D.G. Lainiotis (1969). Simplified parameter quantization procedure for adaptive estimation. IEEE Trans. Autom. Control (Corresp.) 14: 424-425.

Smallwood, R.D. (1968). A decision analysis of model selection. IEEE Trans. Syst. Sci. Cybern. 4(3): 333-342.

Sworder, D. (1966). Optimal Adaptive Control Systems. Academic Press, New York.

13 Stochastic State Space Approach to Reservoir Control

David R. Maidment

The development of reservoirs has a substantial impact on the natural environment. Unique ecological sites, fertile agricultural land, and human communities are often irretrievably altered or destroyed in the process. As a result, the need for these developments is being increasingly questioned in many countries. Critical decisions must be made concerning the nature and extent of such developments in the future.

Central to these decisions is the question of the forecasting and control, or operation, of the reservoirs, since it is only when the project is operating on a day-to-day or real-time basis that its objectives can be realized. The importance of this question can be seen in emergency situations, such as floods, as well as in operation under normal conditions where hydroelectric power, irrigation, water supply, and recreation are provided.

In practice, the development of strategies for reservoir operation is a very complex procedure because of the long time horizon involved, the large number of variables to be considered, and the many uncertainties affecting the inputs and outputs of the reservoirs.

Reservoir control policies have been developed for time scales ranging from minutes to years (Erschler et al. 1974, Fiering and Jackson 1971). Shorter time scales are appropriate for specialized studies of the real-time operation of the system where the sizes of the components are fixed. Longer time scales are appropriate for assessing the expected long-term benefits of the system, considering the size of the components as variables whose optimal value is to be found. On all except perhaps the shortest time scales, the magnitude of the future inflows to the reservoirs must be considered as unknown. Although these inflows may be forecast, in reality there is always some uncertainty about these forecasts, and the degree of uncertainty increases with the length of time over which the inflows are forecast.

The development of reservoir control policies based on monthly time periods has often been found to be appropriate (Schweig and Cole 1968, Askew et al. 1971, Butcher 1971, Torabi and Mobasheri 1973, Croley 1974, Mawer and Thorn 1974, Su and Deininger 1974, Takeuchi and Moreau 1974, McKerchar 1975, Loucks and Dorfman 1975, Colorni and Fronza 1976). From such a time scale it is possible to assess the long-term benefits of the system. Operating policies defined in terms of the average rate of release over a month are actually used for very large reservoirs, such as the Great Lakes of North America; for smaller reservoirs monthly release

policies may form the framework within which the policies for daily operation can be determined, as, for example, in the California Central Valley Project (Becker and Yeh 1974).

Most reservoirs are operated for a number of purposes - for hydroelectric power, irrigation, municipal water supply, and recreation; and this operation may be directed towards several objectives such as national economic development, environmental quality, and social welfare. The risk of emptying the reservoir or of failing to meet goals specified about the releases must be considered, as must the nonlinear relationships inherent in most reservoir systems, particularly in hydroelectric power systems.

Since the problems are so complex, it is not surprising that the use of mathematical programming to optimize the operating policies is really in its infancy as far as its application by the authorities involved in reservoir operation. Current policies have been developed mostly by simulation (U.S. Army Corps of Engineers 1971, Beard 1973, Singh et al. 1975). However, the number of variables which may be optimized in a simulation study is very limited. There is, therefore, a definite need to develop new methodologies based on mathematical programming for reservoir operation, which can account for the complexities inherent in actual systems, and in which a larger number of variables can be optimized.

The problem addressed by the proposed methodology is the optimization of the operating policy for a reservoir for monthly time periods over a year, the operating policy being specified as an average rate of release of water from the reservoir in each month. This rate of release of water is expressed as a function of the storage in the reservoir at the beginning of the month, and it can also be expressed as a function of the inflow into the reservoir in the previous month if the monthly inflows are highly correlated.

The inflows to the system in each month are considered as stochastic variables, i.e., they may be described by probability distributions whose parameters change with time; the sequence of values of average monthly storage volumes also constitutes a stochastic process. When the monthly inflows are considered to be independently distributed from month to month, the sequence of values of storage is a univariate Markov process, which means that the value of storage in the current month is related only to the value of storage which occurred in the immediately preceding month, and not to storage values further back in time. In the case where the inflows in adjacent months are assumed to be jointly distributed, the sequence of storage values and the sequence of inflow values form a bivariate Markov process. Since the statistical parameters of the inflows' probability distributions vary from month to month but not from year to year, these Markov processes are periodic, with a period of one year.

The determination of the operating policy for the reservoir is a Markov decision process that is solved using a successive approximations algorithm of dynamic programming (Howard 1960, White 1963, Su and Deininger 1972). Two policies are developed, one considering independent inflows and one considering serially correlated inflows.

FORMULATION OF MATHEMATICAL MODELS

The variables involved in the formulation of the mathematical model of the system are shown in Fig. 13.1. The time horizon is divided into K stages, $k = 1,2,\ldots,K$. The volume of storage at the beginning of stage k is denoted by $x(k)$. During stage k, the volumes of inflow and release are denoted by $q(k)$ and $u(k)$, respectively. Using the principle of conservation of mass:

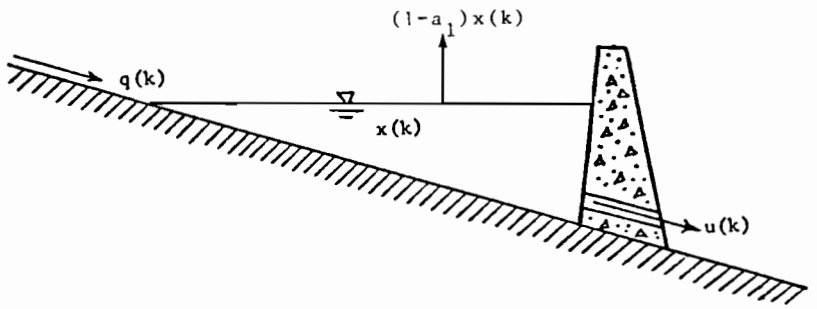


Fig. 13.1 A storage reservoir system.

$$x(k+1) = a_1 x(k) - u(k) + q(k) \tag{13.1}$$

where a_1 is a coefficient dependent on x(k), which accounts for seepage, evaporation, and spillway losses. By assuming that q(k) is an independently distributed random variable with known mean $\mu(k)$ and variance $\sigma(k)$, and introducing a random variable w(k), which has the same distribution but with zero mean and unit standard deviation, we may rewrite (13.1) as follows:

$$x(k+1) = a_1 x(k) - u(k) + \mu(k) + \sigma(k)w(k) \ . \tag{13.2}$$

Within the dynamic programming procedure, x(k) is the state variable, and u(k) is the decision variable at each stage. Equation (13.2) is the state transformation equation needed for the dynamic programming when the inflows are independently distributed.

Alternatively, q(k) can be assumed to be correlated with the inflow in the previous month, q(k-1). If the inflows follow a normal distribution, the relation between q(k) and q(k-1), and their respective parameters, can be formulated as:

$$\frac{q(k) - \mu(k)}{\sigma(k)} = \rho(k) \left[\frac{q(k-1) - \mu(k-1)}{\sigma(k-1)}\right] + \left[1-\rho^2(k)\right]^{\frac{1}{2}} w(k) \tag{13.3}$$

in which $\rho(k)$ is the correlation coefficient between q(k) and q(k-1), and w(k) is a normally distributed random variable with zero mean and unit standard deviation. By defining $x_1(k+1)$ as the normalized flow in month k, $[q(k) - \mu(k)]/\sigma(k)$, and $x_1(k)$ as the corresponding normalized flow in the previous month, (13.3) can be rewritten as:

$$x_1(k+1) = \rho(k)x_1(k) + \left[1 - \rho^2(k)\right]^{\frac{1}{2}} w(k) \ . \tag{13.4}$$

In the dynamic programming procedure, two state variables are now needed, $x_1(k)$, and the volume of storage at the beginning of stage k, which is denoted by $x_2(k)$ in this case to avoid confusion with the independent model. It may be shown, by combining (13.1), (13.3), and (13.4), that the state transformation equation for the dynamic programming with serially correlated inflows can be written in a vector-matrix format as:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} \rho(k) & 0 \\ \sigma(k)\rho(k) & a_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} u(k) \\ \mu(k) \end{bmatrix} + \begin{bmatrix} [1 - \rho^2(k)]^{\frac{1}{2}} \\ \sigma(k)[1 - \rho^2(k)]^{\frac{1}{2}} \end{bmatrix} w(k) \ . \qquad (13.5)$$

It may be noted that (13.2) and (13.5) are special cases of the state-space Gauss-Markov sequence model formulated by Meditch (1969) for the theory of stochastic optimal control. Although the sequel of the present discussion does not rely heavily on the results of this theory, which has been developed for rather specialized systems, the conceptual approach has some similarities. Most notable among these is the concept of the spreading, or propagation, of the probability distribution of the flows as time passes within each stage. In (13.4), the inflow for the current month, $x_1(k+1)$, is the parameter being forecast, and the growth in the variance of the forecast as time passes is a central theme of stochastic estimation and control theory.

OPTIMIZATION BY STOCHASTIC DYNAMIC PROGRAMMING

The optimization algorithm employs the usual method of dynamic programming. The time horizon of K stages is analyzed stage by stage, and each stage uses the results of the analysis of the immediately preceding stage. To illustrate the procedure, consider the optimization for the independent inflow model. At each stage, the feasible range of the input state variable, x(k), which is $x_{min}(k)$ to $x_{max}(k)$, is divided into Q_k discrete intervals, $x_i(k)$, $i = 1,2,\ldots,Q_k$, of size $\Delta x(k)$; and likewise the feasible range of x(k+1), $x_{min}(k+1)$ to $x_{max}(k+1)$, is divided into Q_{k+1} intervals, $x_j(k+1)$, $j = 1,2,\ldots,Q_{k+1}$ of size $\Delta x(k+1)$. The objective of the optimization is to choose the best value for u(k), the release volume in stage k, for each of the discrete intervals of x(k), in each of the K stages.

From a given $x_i(k)$, the "stochastic transformation" of the system over the stage resulting from a release u(k), is described by (13.2), and is shown in Fig. (13.2). In this figure, the "expected transformation" is that which would occur if the mean inflow, $\mu(k)$, occurred, and the output state probability distribution shown in this figure results from the random nature of q(k). From (13.2) it may be seen that the mean and standard deviations of this probability distribution are $a_1x(k) - u(k) + \mu(k)$, and $\sigma(k)$, respectively. The probabilities of occurrence, $P_{ij}(k)$, of the output state intervals, $x_j(k+1)$, can then be found from the appropriate probability distribution function.

It is apparent that there may be stochastic transformations whose resulting output state probability distribution lies partially outside the admissible range of x(k+1). This undesirable behavior can be limited only on a probabilistic or "chance constraint" basis. For each stochastic transformation, the probability that $x(k+1) < x_{min}(k+1)$, and the probability that $x(k+1) > x_{max}(k+1)$ can be found. If upper bounds are assigned to these probabilities then the alternative under consideration is rejected if either of the probabilities is above its upper bounds.

If the alternative, u(k), under consideration from input state interval $x_i(k)$, satisfies these chance constraints, its value is then estimated. To do this, a function $F_{k+1}[x_j(k+1)]$, $j = 1,2,\ldots,Q_{k+1}$ is needed to specify the expected net

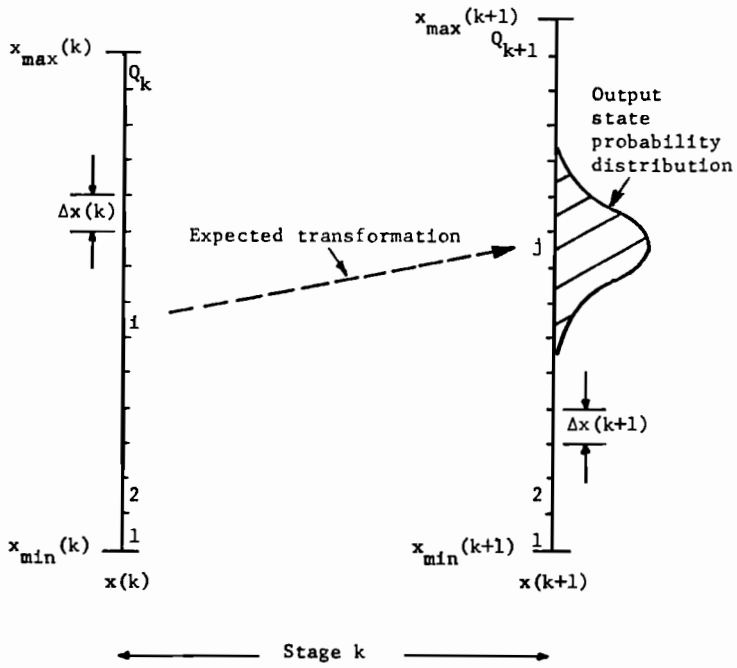


Fig. 13.2 Stochastic transformation.

benefits of future operation of the system for the remaining stages, k+1,k+2,...,K, beginning from the state interval $k_j(k+1)$. If we assume that such a function is available from the optimization that has already been carried out for those stages, the value $F_k[x_i(k)]$ of an analogous function for input state interval $x_i(k)$ in stage k can be found as an expectation over the values of all probable state transformations to the output state intervals $x_j(k+1)$, which could occur from $x_i(k)$ with decision u(k). The value of each transformation is computed as the sum of the value of operation over the stage, $R_k[x_i(k),x_j(k+1),u(k)]$, and the value of $F_{k+1}[x_j(k+1)]$. The function of $R_k[x_i(k),x_j(k+1),u(k)]$ is typically the net benefit expressed in monetary terms of the irrigation water, hydroelectric power, etc. which are supplied during the stage. Mathematically, this computation for $F_k[x_i(k)]$ can be formulated as

$$F_k[x_i(k)] = \sum_{j=1}^{Q_{k+1}} p_{ij}(k)\{R_k[x_i(k),x_j(k+1),u(k)] + F_{k+1}[x_j(k+1)]\} \qquad (13.6)$$

Equation (13.6) is the basic recursive equation for the dynamic programming. The aim of the optimization is to find the optimal value of $F_k[x_i(k)]$ by varying u(k) within its feasible range. The optimization of this recursive equation is carried out for all input state intervals, $x_i(k)$, $i = 1,2,...,Q_k$. This procedure is summarized in Fig. (13.3).

The analysis of one year of operation by monthly stages constitutes an iteration of the optimization procedure. The stages are treated in reverse order beginning with the last stage K. For the first iteration, the value function for the output state intervals, $F_{k+1}[x(K+1)]$, must be arbitrarily assumed. Since these output state intervals are the input state intervals for stage 1, the expected value, or "gain,"

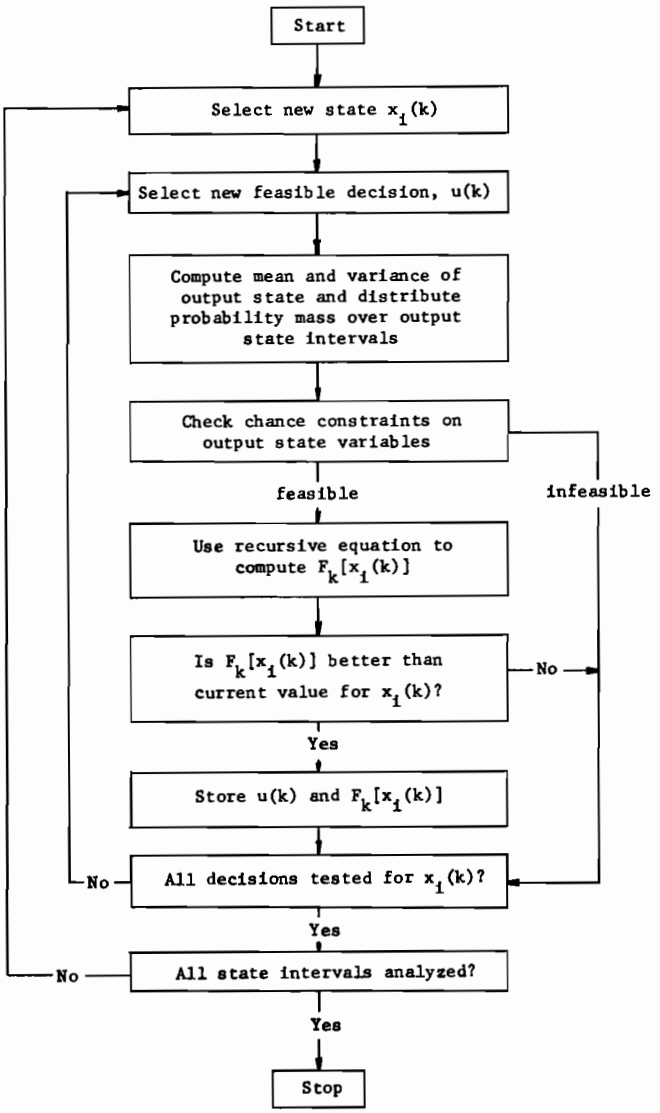


Fig. 13.3 Flowchart of recursive equation optimization.

g_i, of a year's operation, beginning and ending in the same state, can be calculated by

$$g_i = F_1[x_i(1)] - F_{K+1}[x_i(K+1)] \quad ; \quad i = 1,2,\ldots,Q_1 \tag{13.7}$$

For subsequent iterations, the value function, $F_{K+1}[\underline{x}(K+1)]$, can be reinitialized by subtracting the value of the lowest storage state, $F_1[x_1(1)]$:

$$F_{K+1}[x_i(K+1)] = F_1[x_i(1)] - F_1[x_1(1)] \quad ; \quad i = 1,2,\ldots,Q_1 \tag{13.8}$$

After three or four iterations, the values of the individual state gains, g_i, converge to a single value, g, called the gain of the system, which represents the expected annual benefits from the system if it is operated using the optimal policy. The "relative" value function found using (13.8) also stabilizes; the stable function represents the present worth over the long term of beginning operation in a specified storage relative to beginning operation in the lowest storage state interval. A general flowchart of the algorithm is shown in Fig. (13.4).

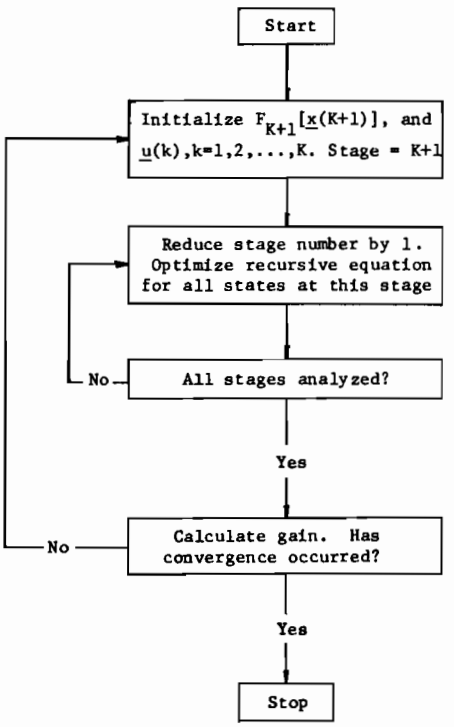


Fig. 13.4 General flowchart of the algorithm for computation of the gain of the system.

In principle, the optimization procedure for the serially correlated inflow model is similar to that described for the independent inflow model. In practice, however, it is more complex. That the random variables $x_1(k+1)$ and $x_2(k+1)$ are linearly dependent can be demonstrated from (13.5). The linear independence of the random variables means that the computation of the recursive equation equivalent to (13.6) involves a line integration of the probability and value functions in two dimensions. Since both functions are defined using a discrete grid of the variables $x_1(k+1)$ and $x_2(k+1)$, and the line of integration is at an angle to this grid, the resultant numerical method for the line integration is somewhat complicated.

The computer time requirements for dynamic programming algorithms can be formulated using the method proposed by Chow et al. (1975). In this method, the sequence of steps within the outer loop of the flowchart in Fig. (13.3) has been termed a "unit operation," and the computer time requirement, T_e, can be estimated as the time for one such unit operation, T_a, multiplied by the number of unit operations performed in the optimization:

$$T_e = T_a M \sum_{k=1}^{K} \left(\prod_{i=1}^{n} Q_{i,k} \prod_{j=1}^{p} P_{j,k} \right) \qquad (13.9)$$

in which M is the number of iterations involved in optimization, K is the number of stages, and $Q_{i,k}$ and $P_{j,k}$ are the number of feasible values that state variable i, i = 1,2,...,n, and decision variable, j,j = 1,2,...,p, can take at stage k in each iteration or in the optimization procedure. The value of T_a depends on the type of computer, the method of coding, the kind of compiler, the nature of the constraints, and other factors and so must be empirically determined.

AN APPLICATION

Description of the Physical System

The methodology is applied to the Watasheamu dam and reservoir which have been proposed for construction by the U.S. Bureau of Reclamation in Nevada. The objective of the analysis is to derive, for this reservoir, an operating policy that specifies the volume of water to be released from the reservoir in each month of the year. Two policies are developed: the first, using the independent inflow model, specifies the release as a function of the storage available at the beginning of the month; and the second, coming from the serially correlated inflow model, considers in addition to the available storage, the average rate of inflow that occurred in the previous month. This particular system was chosen because it is relatively simple (it has only one reservoir) and also because the required data are readily available (Butcher and Fordham 1970).

The region is arid. The average rainfall at Reno, Nevada is 9 in. (230 mm); in the watershed of the Watasheamu reservoir, the average annual rainfall is about 30 in. (760 mm). Much of this precipitation falls as snow. In the warmer months, from June to October, the streamflow is supplied by snowmelt and the average monthly streamflows in this period are highly correlated. In the other months of the year, this is not so. A stream gauging station is located near the site of the proposed dam and 28 yrs of streamflow records are available. The statistical parameters of these data are summarized in Table 13.1 and the data are assumed to be normally distributed. The watershed area is approximately 280 square miles (725 km^2) and the Watasheamu river has an average annual flow rate of 357 ft^3/sec (10 m^3/sec or 259 KAF per year; 1 KAF = 1,000 acre-feet = 1.23×10^6 m^3).

The physical features of the proposed Watasheamu dam are shown in Fig. (13.5). The dam will be of earth and rock-fill construction and will be approximately 300 ft (90 m) high. There are three ways by which water will be released from the dam. Two of these are controlled outlet works that draw water from the reservoir through the intake structures shown in the figure. The main outlet works at elevation 5,075 ft (1547 m) release water through the powerhouse up to the capacity of the power generation equipment, which is considered as 50 MW. When the level of the water in the reservoir falls below 5,188 ft (1,582 m), power will no longer be generated and water released through the main outlet works will bypass the power-house, as will any flow beyond the capacity of the generation facilities.

Water may also be released through auxiliary outlet works whose intake is at elevation 5,060 ft (1,543 m). Such releases cannot be used to produce power, however. The 30 KAF (3.7×10^7 m^3) storage below elevation 5,159 ft (1573 m) in the reservoir has been identified as "inactive" storage by the U.S. Bureau of Reclamation. It is considered undesirable that the reservoir fall below this level. The capacity of the reservoir at the spillway crest elevation will be 160 KAF (1.97×10^8 m^3), so that the net active storage capacity will be 130 KAF (1.60×10^8 m^3).

Table 13.1. Statistical parameters of streamflow data for the Watasheamu Reservoir watershed

	Jan	Feb	Mar	Apr	May	June
Mean	9.948	10.908	14.630	35.420	67.794	58.406
Standard deviation	6.865	7.553	5.909	13.120	23.760	27.720
Serial correlation	0.754	0.452	0.237	0.355	0.500	0.729
	July	Aug	Sept	Oct	Nov	Dec
Mean	23.086	8.251	5.205	4.858	8.078	12.365
Standard deviation	16.670	4.511	2.336	1.534	11.460	16.090
Serial correlation	0.934	0.920	0.955	0.802	0.211	0.669

Note: The mean and standard deviation data are in KAF/month. The serial correlation coefficient is that between the month shown and the previous month.

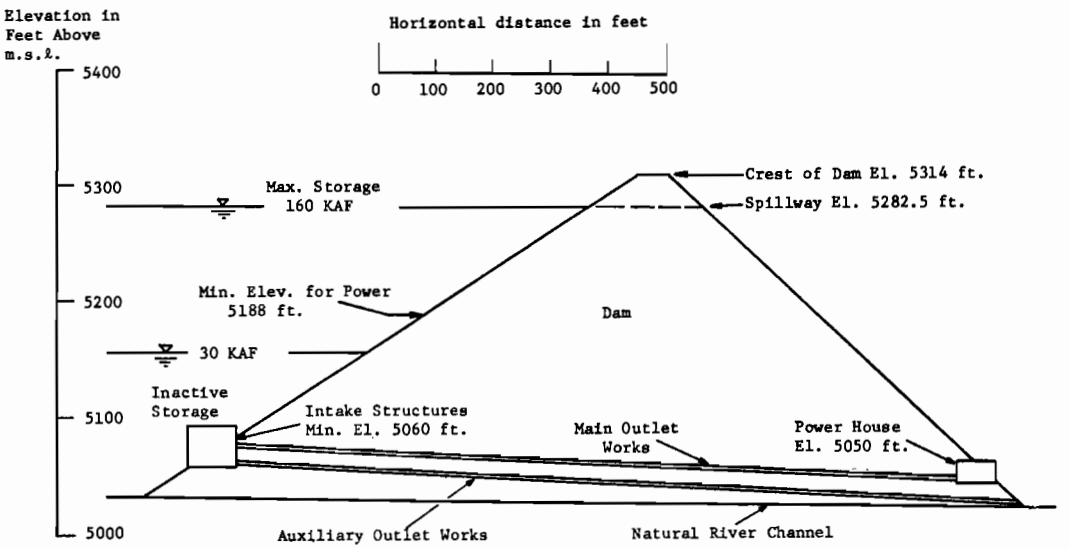


Fig. 13.5 Watasheamu dam.

The Mathematical Model

In the mathematical models formulated as (13.2) and (13.5), the loss coefficient of the storage, a_1, must be determined. To account for evaporation, which is the major cause of loss of water from the reservoir, the depth of evaporation from a water surface in each month is multiplied by the water surface area. The resulting data for a_1 are fitted with a polynomial function in each month using storage volume as the dependent variable.

In the dynamic programming, the storage state variable is divided into discrete intervals of the size of 1 unit of normalized flow. In general, the intervals should be no larger than two standard deviations of the probability distribution that is applied to them in the computation of the recursive equation.

The reservoir will be operated for hydroelectric power, the supply of irrigation water, and flood control. Flood control is allowed for by reserving storage at the top of the reservoir in the appropriate months. The other two purposes are incorporated into the objective function, $R_k[x_i(k),x_j(k+1),u(k)]$, which may be expressed in dollars in a simplified form as follows:

$$R_k[x_i(k),x_j(k+1),u(k)] = 5.453\text{x}10^{-3}u(k)\bar{h} + 2.5\min\{u(k);u_{mi}(k)\} \qquad (13.10)$$

in which $u(k)$ is the rate of release in acre-feet per month, $\bar{h}$ is the average elevation of the reservoir in the month, and $u_{mi}(k)$ is the maximum volume of irrigation water needed in month k. The first term of (13.10), $5.453\text{x}10^{-3}u(k)\bar{h}$, is the value of power generation assuming all power can be sold at a price of 0.71 cents per kW-hour, and a power conversion efficiency of 0.75. The second term of this equation is the value of water for irrigation if irrigation water can be sold for \$2.50 per acre-foot up to a maximum volume $u_{mi}(k)$. The values of $u_{mi}(k)$ in KAF per month are as follows: May, 25; June, 30; July to September, 45; October, 20; and zero in the remaining months. The power function is actually more complex than that expressed in (13.10) because there is no power generated if $\bar{h}$ is less than a specified minimum level, and the maximum rate of generation is limited by the capacity of the generating equipment. Therefore, the objective function is nonlinear, discontinuous, and changes from stage to stage.

The gain, g, which is to be maximized is the expected annual benefits from the system:

$$g = \underset{x_i(k)}{E}\left\{\sum_{k=1}^{12} R_k[x_i(k),x_j(k+1),u(k)]\right. \qquad (13.11)$$

$$i = 1,2,\ldots,Q_k$$

$$k = 1,2,\ldots,12$$

The months or stages are numbered, as are the calendar months; stage 1 is January, etc.

Results of Optimization

The results of optimization include the policy for operation of the reservoir, together with the average annual net benefits and expected response of the reservoir with this policy. Two computer programs were written in FORTRAN, one program for the independent inflow model, and the other for the serially correlated inflow model. Maidment (1976) describes these programs and gives their computer code. The results presented here are primarily obtained using the independent inflow model, although a number of comparisons are made between these results and those obtained from the serially correlated inflow model.

The optimal policy for operation using the independent inflow model is shown in Table 13.2, from which several observations about the form of the policy can be

Table 13.2. Optimal policy for operation using the independent inflow model

Storage in KAF	Release from Reservoir (KAF/month)											
	Jan	Feb	Mar	Apr	May	June	July	Aug	Sept	Oct	Nov	Dec
30	0.0	0.0	0.0	0.0	13.75	0.0	0.0	0.0	6.25	7.5	0.0	0.0
40	0.0	0.0	0.0	0.0	25.0	0.0	0.0	0.0	14.0	16.25	0.0	0.0
50	0.0	0.0	0.0	0.0	25.0	12.5	0.0	0.0	15.25	20.0	0.0	0.0
60	0.0	0.0	0.0	0.0	25.0	23.75	4.0	0.0	24.0	20.0	5.0	5.0
70	0.0	0.0	0.0	0.0	25.0	30.0	4.0	7.5	44.75	20.0	5.0	15.0
80	1.0	0.0	0.0	0.0	25.0	30.0	14.75	16.5	45.0	20.0	14.75	25.0
90	8.5	3.0	0.0	0.0	25.0	30.0	24.75	27.25	45.0	20.0	24.75	32.5
100		10.0	4.0	1.0	27.5	30.0	33.75	37.25	45.0	20.0	34.75	
110			11.25	11.25	37.75	30.0	43.75	45.0	45.0	20.0	44.75	
120				18.75	48.75	30.0	45.0	45.0	45.0	20.0	52.25	
130					58.75	31.0	45.0	45.0	45.0	21.5		
140					68.75	41.25	45.0	45.0	45.0	27.5		
150					78.75	51.25	45.0	45.0	45.0	37.5		
160												

NOTE: The blank spaces are reserved for flood control.

made. As is to be expected, the optimal release in each month increases with the volume of available storage. The major releases are made in the irrigation season, from May to October, and these releases are controlled by the irrigation requirements, since, over a wide range of available storage, the optimal release is equal to the maximum volume required for irrigation, $u_{mi}(k)$, specified in the previous section. By comparing the mean inflows, given by Table 13.1, with the optimal policy, it can be seen that large releases are made in months of high inflow, particularly in May and June. In October and November large releases are also made because then the reservoir must be drawn down for flood control.

The optimal policy for the serially correlated inflow model comprises tables similar to Table 13.2 for each month of the year, in which the optimal release is specified as a function of available storage and as a function of the inflow in the previous month. These tables are presented in Maidment (1976) and some general comments are made here. The form of the policy is similar to that described previously for the independent inflow model. The optimal release increases with available storage, and also with the rate of inflow in the previous month.

If the optimal releases of the two policies for any specified available storage are compared, it appears that the release from the independent inflow model is approximately the average of the optimal releases for the various flow states in the serially correlated inflow model. When a low inflow has occurred, the serially correlated release is low; conversely, this release is higher than the average when a high inflow has occurred; over the middle of the flow range, the releases are similar.

The gain, or average annual benefits, from the operation converges after three or four iterations of optimization. Under the same constraint conditions, the value of the gain for the independent inflow model is \$773,560,000 and for the serially correlated model \$777,506,000, a difference of about 0.5%.

The serial correlation coefficients of the inflows are high during the irrigation season, and because of this it might have been expected that the gain of the serially correlated model would be significantly greater than that of the independent model. The fact that they turn out to be similar might be explained in several ways. One explanation might be that since the independent model's release policy does not usually differ greatly from the serially correlated model's policy over the range of inflows most likely to occur, the independent policy probably realizes most of the available benefits from the operation, particularly those in the irrigation season. Alternatively, it could be postulated that the objective function is not sufficiently complex to properly evaluate the benefits from the system and, therefore, to properly discriminate between the two models. For practical purposes, however, it appears that the independent model is to be recommended over the serially correlated inflow model because it is much simpler to use, and in this example, little benefit from operation seems to be sacrificed in doing so.

The computer execution times for the two programs are approximately 30 sec for the independent inflow model, and 105 sec for the serially correlated inflow model. The corresponding values of T_a, found using (13.9), are 3.54×10^{-3} sec, and 5.77×10^{-3} sec. Thus, about half of the increase in computer time incurred by using the serially correlated inflow model is due to the increased complexity of evaluating a single decision, and the other half of this increase is due to the larger number of decisions examined since there are two state variables involved instead of just one.

These computer time requirements may be compared with those involved in applying dynamic programming to the same problem using sets of deterministic inflow sequences. From the data presented by Chow et al. (1975), it is estimated that an

equivalent deterministic analysis would require a T_a of about 3×10^{-4} sec. Therefore, the computer time required for the independent inflow model is about the same as that for the analysis of 10 deterministic sequences. This is a favorable comparison for the present methodology because the number of deterministic sequences required to obtain meaningful results is usually considerably more than 10.

CONCLUSIONS

The principal advantage of the stochastic state variable dynamic programming is that the stochastic nature of the water resource system inputs is embedded directly within the optimization procedure without requiring computationally expensive system simulations or discrete probability distributions. The policy found from the optimization allows for the feedback of changing conditions in the system, since it is a function of the state variables and the stage, instead of being a function of just the stage alone. There is no restriction on the form of the objective function, as is required by some optimization procedures. There are few difficulties with the imposition of constraints on the state and decision variables, whether they are deterministic or probabilistic constraints.

A disadvantage of the proposed methodology is the large amount of computer time and memory requirements. Although these requirements are not excessive when compared with those for comparable analyses by other methods, they probably restrict the application of the methodology to systems involving, at most, four state variables at the current level of computer technology.

Although the procedure of application of the serially correlated inflow model is more complex than that of the independent inflow model, the average annual benefit found from the optimization using the serially correlated inflow model is only slightly greater than the average annual benefit found using the independent inflow model. For practical purposes, therefore, it is recommended that the independent inflow model be adopted.

To develop real-time operating policies where the releases must usually be made on a daily or even hourly basis, a suballocation methodology would have to be formulated for each month to disaggregate the average monthly release determined from the present methodology into the releases required at the shorter time intervals. The development of such a suballocation methodology is a challenging area for further research.

REFERENCES

Askew, A.J., W.W-G. Yeh, and W.A. Hall (1971). Use of Monte Carlo techniques in the design and operation of a multipurpose reservoir system. Water Resour. Res. 7(4): 819-826.

Beard, L.R. (1973). Status of water resources systems analysis. Proc. ASCE, J. Hyd. Div. 99 (HY4): 559-565.

Becker, L., and W.W-G. Yeh (1974). Optimization of real time operation of a multiple-reservoir system. Water Resour. Res. 10(6): 1107-1112.

Butcher, W.S. (1971). Stochastic dynamic programming for optimum reservoir operation. Water Resour. Bull. 7(1): 115-123.

Butcher, W.S., and J.W. Fordham (1970). Optimum reservoir operation using stochastic dynamic programming. Publ. No. 9, Center for Water Resources Research, University of Nevada System, Reno, Sept.

Chow, V.T., D.R. Maidment, and G.W. Tauxe (1975). Computer time and memory requirements for DP and DDDP in water resource systems analysis. Water Resour. Res. 11(5): 621-628.

Colorni, A., and G. Fronza (1976). Reservoir management via reliability programming. Water Resour. Res. 12(1): 85-88.

Croley, T.E., II (1974). Sequential stochastic optimization for reservoir systems. Proc. ASCE, J. Hyd. Div. 100 (HY1): 201-219.

Erschler, J., F. Roubellat, and J.P. Vernhes (1974). Automation of a hydroelectric power station using variable structure control systems. Automatica 10(1): 31-36.

Fiering, M.B., and B.B. Jackson (1971). Synthetic stream flows. Water Resour. Monogr. 1, Am. Geophys. Union, Washington, D.C.

Howard, R.A. (1960). Dynamic Programming and Markov Processes. M.I.T. Press, Cambridge, Massachusetts.

Loucks, D.P., and P.J. Dorfman (1975). An evaluation of some linear decision rules in chance-constrained models for reservoir planning and operation. Water Resour. Res. 11(6): 777-782.

McKerchar, A.I. (1975). Optimal monthly operation of interconnected hydroelectric power storages. J. Hydrol. 25: 137-158.

Maidment, D.R. (1976). Stochastic state variable dynamic programming for water resources systems analysis. Ph.D. Thesis, Univ. of Illinois at Urbana-Champaign. University Microfilms, Ann Arbor, Michigan.

Mawer, P.A., and D. Thorn (1974). Improved dynamic programming procedures and their practical application to water resource systems. Water Resour. Res. 10(2): 183-190.

Meditch, J.S. (1969). Stochastic Optimal Linear Estimation and Control. McGraw-Hill, New York.

Schweig, Z., and J.A. Cole (1968). Optimum control of linked reservoirs. Water Resour. Res. 4(3): 479-498.

Singh, K.P., J.B. Stall, C. Lonnquist, and R.H. Smith (1975). Analysis of the operation of Lake Shelbyville and Carlyle Lake to maximize agricultural and recreation benefits. Contract report, Illinois State Water Survey, Urbana, Illinois, June.

Su, S.Y., and R.A. Deininger (1972). Generalization of White's method of successive approximations to a periodic Markovian decision process. Oper. Res. 20(2): 318-327.

Su, S.Y., and R.A. Deininger (1974). Modeling the regulation of the Great Lakes under uncertainty of future water supplies. Water Resour. Res. 10(1): 11-25.

Takeuchi, K., and D.H. Moreau (1974). Optimal control of multiunit interbasin water resource systems. Water Resour. Res. 10(3): 407-414.

Torabi, M., and F. Mobasheri (1973). A stochastic dynamic programming model for the optimum operation of a multiple purpose reservoir. Water Resour. Res. 9(6): 1089-1099.

U.S. Army Corps of Engineers (1971). Runoff evaluation and stream-flow simulation by computer. North Pacific Div., U.S. Army Corps of Engineers, Portland, Oregon, May.

White, D.J. (1963). Dynamic programming, Markov chains and the method of successive approximation. J. Math. Anal. Appl. 6: 373-376.

14 Some Remarks on Real-Time Control of Dissolved Oxygen in Rivers

H. Stehfest

The goal of this contribution is to show approaches to the problem of real-time control of river quality on three levels of sophistication. The control problem consists of operating certain river quality control facilities in such a way that desirable water quality is maintained. We will confine our considerations here to oxygen concentration in the river as an indicator of water quality. Application of the described methodology to other river quality aspects is straightforward. Other assumptions made for the sake of simplicity are: steady-state hydrology and temperature, the use of artificial aerators and wastewater treatment plants as the only control facilities, and the very simple Streeter-Phelps model to describe the water quality dynamics.

The Streeter-Phelps model has the following form

$$\begin{aligned} \frac{\partial b}{\partial t} + v\,\frac{\partial b}{\partial \ell} - D\,\frac{\partial^2 b}{\partial \ell^2} &= -k_1 b + u_b \\ \frac{\partial c}{\partial t} + v\,\frac{\partial c}{\partial \ell} - D\,\frac{\partial^2 c}{\partial \ell^2} &= -k_1 b + k_2(c_s - c) + u_c\,, \end{aligned} \tag{14.1}$$

where

t = time,

ℓ = distance along the river,

$v(\ell)$ = stream velocity,

$D(\ell)$ = diffusion coefficient,

$b(\ell,t)$ = biological oxygen demand (BOD),

$c(\ell,t)$ = dissolved oxygen (DO),

$k_1(\ell)$ = BOD degradation coefficient,

$k_2(\ell)$ = reaeration coefficient,

c_s = oxygen saturation concentration,

$u_b(\ell,t)$ = BOD input,

$u_c(\ell,t)$ = DO input.

Hence, the system to be controlled is a linear, distributed parameter system for which many inputs have to be adjusted simultaneously so that the outputs behave in a desirable way.

An optimal control problem for such a system is hard to solve, so the problem is usually simplified in some way. To begin with, one can neglect longitudinal dispersion. This implies that the system can be transformed into a lumped parameter system with flow time τ as the independent variable. An optimal control problem becomes, through this transformation,

$$\min_{u\in U} J = \int_0^T g(x(\tau), u(\tau))\, d\tau$$

subject to (14.1), which can be rewritten as

$$\frac{dx}{d\tau} = Fx + Gu$$

with

$$x = (b,\ c_s - c)^T,\quad u = (u_b, u_c)^T,\quad F = \begin{pmatrix} -k_1 & 0 \\ +k_1 & -k_2 \end{pmatrix},\quad G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

U is the set of admissible controls and T the flow time of the river section considered. Even this simplified optimal control problem can only be solved easily in particular cases. One such case is for a quadratic performance index, i.e.,

$$J = \int_0^T (x^T\ u^T) A\ (x^T\ u^T)^T\, d\tau\ .$$

The simplicity of the linear-quadratic optimal control problem has enticed many authors into using a quadratic performance index even in cases such as design studies, where its justification may be questioned. For example, in papers on installment of artificial aerators, a performance index of the form

$$J = \int_0^T (\alpha(c - c_d)^2 + \beta u_c^2)\, d\tau$$

has been used, where c_d is the desired DO level and α and β are weighting factors. The first term is to account for the environmental damage, and the second one stands for the costs associated with the artificial aeration. This index is not very realistic because exceeding the desired oxygen concentration ought to be considered beneficial, and the aeration required for a certain oxygenation rate should depend on the oxygen concentration.

If the design problem has been solved with a realistic performance index, the problem of maintaining the nominal oxygen concentration despite various disturbances is a so-called second variation problem, which has exactly the desirable linear-quadratic structure just described (Bryson and Ho 1975). The problem consists in minimizing the second variation of the performance index (the first one vanishes anyway), which can be written as

$$\delta^2 J = \int_0^T (\delta x^T \; \delta u^T) \begin{pmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{pmatrix} (\delta x^T \; \delta u^T)^T \, d\tau$$

where H is the Hamiltonian:

$$H = g + \lambda(Fx + Gu) \; ,$$

whose second derivatives are identical to the second derivatives of the cost function g. The variations δx and δu obey the equation

$$\frac{d\delta x}{d\tau} = F\delta x + G\delta u,$$

where $\delta x(0)$ is specified.

The solution of this problem can be written in the form

$$\delta u(\tau) = K(\tau) \; \delta x(\tau)$$

where the gain matrix $K(\tau)$ is

$$K(\tau) = - g_{uu}^{-1} \; (P + g_{xu}) \; ,$$

where P is the solution of a Riccati equation and the derivatives of g are evaluated along the nominal direction of motion. The expression for δu has the form of a continuous feedback law, and it is obviously optimal if a deviation from the nominal direction of motion occurs later than at $\tau=0$.

If we go back to the original coordinates t and ℓ, we see that the feedback control law is time invariant and that no exchange of information between different points along the river is necessary (local control):

$$\delta u(\ell,t) = k(\ell) \; \delta x(\ell,t) \; .$$

The implementation of this control, which requires discreteness in space, is shown in Fig. 14.1 for one reach.

In the case of artificial aeration, we have u_c as the only controllable input, and the control law is

$$\delta u_c(\ell,t) = (k'(\ell)k''(\ell))(\delta b(\ell,t)\ \delta c(\ell,t))^T ,$$

where k' and k" are computed by solving a 2 x 2 matrix differential equation, a Riccati equation.

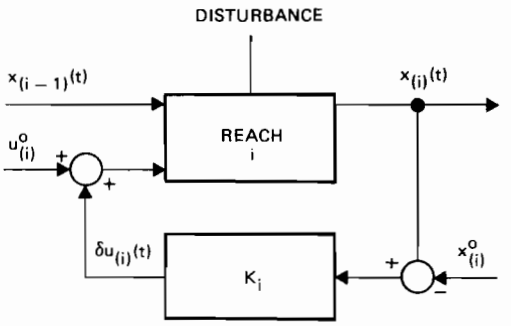


Fig. 14.1 Structure of the local feedback control scheme resulting from the second variation problem (from Rinaldi et al. 1978).

A possibility for refining the plug flow model which led to the simple lumped parameter, second variation problem, is to divide the river into reaches that are considered to be continuously stirred (completely mixed) tank reactors (CSTR model; Young and Beck 1974). This model, which is similar to the Nash model of hydrology, can be derived as a low-frequency approximation to the exact distributed-parameter system. The low-frequency approximation to the transfer function of a series of n river reaches for BOD is

$$M(s) = (m(s))^n$$

where

$$m(s) = \frac{m(0)}{1 + \dfrac{L/n}{v\left(1 + \dfrac{2k_1 D}{v^2}\right)} s}$$

is the transfer function of a single reach and L is the total length of the river section considered. (The oxygen concentration needs a slightly different treatment, since it is coupled with the BOD, but it is not necessary to go into these details in order to explain the idea behind the CSTR model.) On the other hand, the transfer function of a series of n connected, completely mixed tanks in which the BOD reaction goes on is

$$M'(s) = (m'(s))^n$$

where

$$m'(s) = \frac{m'(0)}{1 + \dfrac{L/n}{v(1 + k_1 \frac{L/n}{v})} s} .$$

L/(nv) is the residence time of the water in the reaches. Comparison of the two-reach transfer functions shows that they become identical if the reach length L/n is chosen such that L/n = 2D/v (in this case the gains are identical). Since 2D/v, too, is usually small and the number of reaches not very large, the time constant of the serially connected tanks is smaller than the time constant of the low-frequency dispersion model. If this is the case, in order to have identical wave propagation velocities, an additional time delay Δs has to be introduced. The reach transfer function becomes

$$m''(s) = \frac{m'(0)\, e^{-\Delta s}}{1 + \frac{L/n}{v(1+k_1 \frac{L/n}{v})}\, s} \quad .$$

If we use a CSTR model with a suitable reach length, the quality control problem is still of the same degree of complexity as the general problem made discrete in space. As suggested by the structure of the solution of the second variation problem, we can try to control the system locally such that the nominal water quality is maintained. This would correspond to the most traditional approach of process control engineering for multivariable systems, namely, to associate each output with one and only one control variable through a feedback controller (Fig. 14.2). The most critical problem with this approach is the input-output assignment.

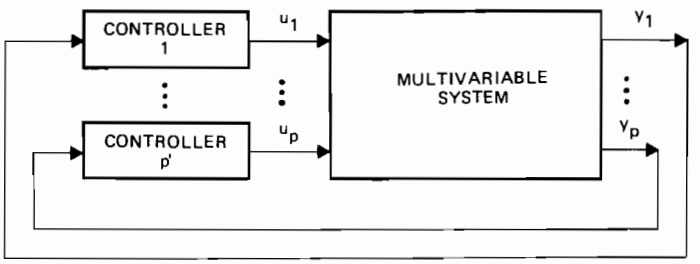


Fig. 14.2 The classical multivariable control system (from Rinaldi et al. 1978).

This is often solved in a heuristic way by associating with each input variable that control variable which has the strongest influence on it. This approach would satisfy the condition of local control if the reaches are sufficiently long. If the reaches are too short, input-output connections that overlap one or several reaches are preferable (Fig. 14.3).

Assuming that we have solved the assignment problem in the form of local feedbacks, we have to design the appropriate local controllers. The most simple approach to this problem is to solve a control problem for each reach under the assumption that all other reaches are at their nominal quality. In principle, such an approach can lead to very undesirable behavior of the whole system, but if the input-output assignment is reasonable, satisfactory results can be expected. In particular, we can be sure that the total system is stable if each feedback system is stable, because the subsystems are in a cascade. If we adopt this decomposition into individual feedback control problems, we find that the design procedure is remarkably different from that of the second variation problem previously discussed. Here we have to solve several simple feedback control problems which are completely independent of each other, while in the second variation problem, the control laws were coupled through the Riccati equation.

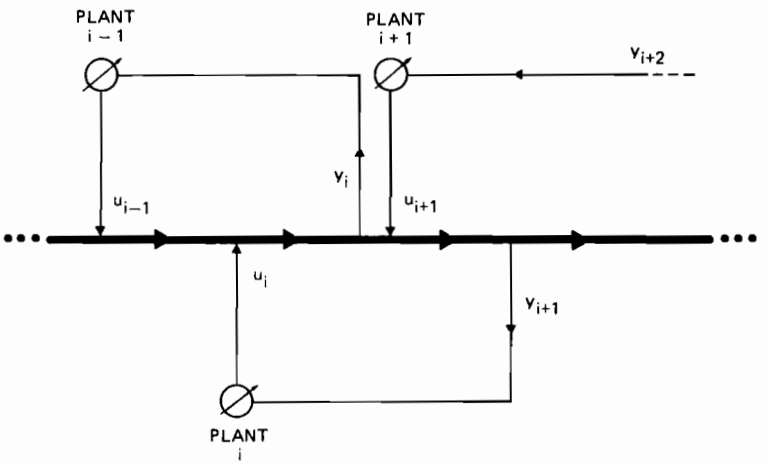


Fig. 14.3 Overlapping control structure along a river (from Rinaldi et al. 1978).

The criterion for the local control problems could again be the minimization of some performance index. Another, more pragmatic criterion is stability, which means that we try to design the controller such that the resulting system enjoys a satisfactory dynamic behavior (without being optimal). In the case of the river reaches, the desired change of the dynamic behavior is, of course, the speeding up of the naturally quite slow disappearance of disturbances. With linear systems, this problem is equivalent to designing the controller such that the closed-loop system has a prescribed set of eigenvalues; this problem is called pole assignment.

If we consider a linear feedback state controller, i.e., a feedback of the type

$$\delta u(t) = K\delta x(t) \ ,$$

the dynamics of the closed-loop system are governed by

$$\frac{d\delta x}{dt} = (F - GK)\delta x \ ,$$

and the problem of pole assignment consists of determining K such that the equation

$$\det(F - GK - \lambda I) = 0$$

has prescribed solutions. It can be shown that this can be done for any set of poles if and only if the original system is completely controllable, i.e., if the controllability matrix

$$C = (G \ FG \ . \ . \ . \ F^{n-1}G)$$

is of rank n. Hence, inspection of the controllability properties of the system

yields information about the possibility of pole assignment. If the poles cannot be fixed arbitrarily, a change in the proposed input-output assignment may provide the solution.

The calculation of the matrix K is quite complex for multi-input systems, but for single-input systems the solution can often be obtained by hand. If the control is achieved by varying the efficiency of a treatment plant on the reach, the dynamics of a reach are described in the CSTR model by

$$\frac{d\delta b}{dt} = -\left(k_1 + \frac{v}{L/n}\right)\delta b + \delta u$$

$$\frac{d\delta c}{dt} = -k_1\delta b - \left(k_2 + \frac{v}{L/n}\right)\delta c \ ,$$

and the controllability matrix is consequently

$$C = \begin{pmatrix} 1 & -\left(k_1 + \frac{v}{L/n}\right) \\ 0 & -k_1 \end{pmatrix},$$

which is obviously of rank 2. The row vector K, which represents the control law as a function of the prescribed poles λ_1 and λ_2, reads

$$K^T = \begin{pmatrix} -\left(k_1 + \frac{v}{L/n} + k_2 + \frac{v}{L/n} + \lambda_1 + \lambda_2\right) \\ -\frac{1}{k_1}\left(k_2 + \frac{v}{L/n} + \lambda_1\right)\left(k_2 + \frac{v}{L/n} + \lambda_2\right) \end{pmatrix} .$$

The implementation of the control law is somewhat problematic because the BOD measurements are available only at a time lag of 2 or more days. The solution suggested by linear system theory is to use a linear state reconstructor in series with the controller. The poles of this enlarged system can be fixed arbitrarily if and only if the original system is completely controllable and completely observable. Another way to solve the problem of BOD measurement delay is to measure a variable which is strongly correlated with the BOD. Such substitute measures could be total organic carbon (TOC) or chemical oxygen demand (COD).

Instead of using a proportional feedback we could also use a dynamic feedback containing, for example, an integrator for the deviation of oxygen concentration from the reference value. This would require that the system be at the reference value under steady-state conditions, no matter what the uncontrollable input is. Looking at the integral of the oxygen deviations as a state variable of the open-loop system, one can again make use of the relationship between controllability and pole assignability. Because of the troubles with BOD measurements, one can also use a single industrial proportional and integral (PI) regulator working on the oxygen deviation (see Young and Beck 1974). Of course, with a single PI regulator the poles of the closed-loop system cannot be fixed at will because one has fewer

parameters to adjust than poles.

For very large disturbances, like accidental releases of huge amounts of pollutants, the control schemes discussed so far are not sufficient. As far as the model is concerned, diffusion is not considered adequately. Technical constraints for control actions have to be observed, and coordinated action of the control facilities is necessary. We briefly discuss, in conclusion, an approach to the optimal control of an emergency situation caused by an accidental release of pollutants (Tamura 1974). The computational effort will be much bigger in this case than for the control schemes discussed so far.

The model to be used is the so-called distributed-lag model. It is a generalization of the CSTR model, which can be explained best in the frequency domain. Instead of introducing a pure additional time delay Δs between two adjacent reaches, we can introduce a more general function $\Phi(s)$ in order to describe the behavior of the river better:

$$m(s) = \frac{m(0)\,\Phi(s)}{1 + \dfrac{L/n}{v(1 + k_1 \frac{L/n}{v})}\,s} \quad .$$

If the CSTR river model is interpreted as a sequence of channels and pools, where the channels act as pure time delays, then in the distributed-lag model the channels are allowed to have a more general transfer function. The value of the delay Δs in the CSTR model could be deduced from the diffusion coefficient, flow velocity, and reach length, but the function $\Phi(s)$ (or the corresponding impulse response function) can be estimated only from actual process observations.

The problem is made discrete in time, as well as space, whereby the distributed-parameter optimal control problem is transformed into a mathematical programming problem. The performance index chosen is, in discrete form,

$$J = \sum_{t=0}^{T-1} \sum_{i=1}^{n} (\alpha_i(b_i(t+1) - b_i^o)^2 + \beta_i(c_i(t+1) - c_i^o)^2 + \gamma_i(\varepsilon_i(t) - \varepsilon^o)^2).$$

The superscript ^{o} indicates steady-state conditions. As in the index for the aerator allocation given on p. 232, the first two terms describe losses due to quality deterioration, and the third one reflects the costs of treatment as a function of treatment efficiency ε_i. The use of this quadratic performance index is justified here, because deviations from the nominal values in this emergency situation can be expected to be in one direction only, and because the costs associated with a certain treatment efficiency are independent of the river state.

The constraints are the equations of the distributed-lag model and the inequality constraints for the efficiencies of the treatment plants. The model equations are of the form

$$b_i(t + 1) = a_i b_i(t) + \frac{Q_{i-1}}{v_i} \sum_{j=0}^{\theta} \phi_{j-1}(j) b_{i-1}(t - j) + (1 - \varepsilon_i(t))W_i$$

and there is an analogous equation for $c_i(t + 1)$. The initial conditions are the nominal conditions everywhere except in the reach in which the accidental release occurred. W_i is the waste water production in reach i and Q_i is the volume of river flow.

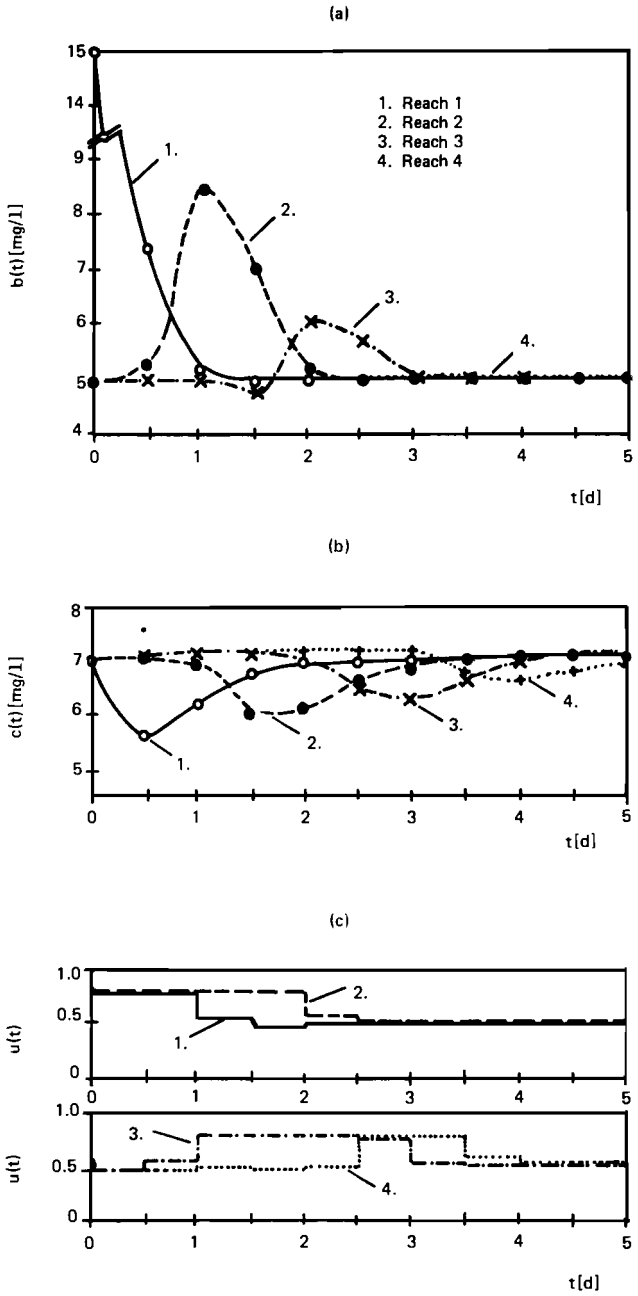


Fig. 14.4 Optimal emergency control. (a) BOD, (b) DO, (c) treatment efficiency and BOD discharge (from Rinaldi et al. 1978).

The summation in the equation can be looked upon as the discrete form of the convolution of $b_{i-1}(t)$ with the impulse response of the above-mentioned channels. The constraints for the efficiency are of the type

$$0 \leq \epsilon_i(t) \leq \bar{\epsilon}$$

The resulting mathematical programming problem is convex and has a separable objection function and linear constraints. Various techniques could be used for its solution. For the example given, a particularly efficient method based on Lagrange duality was applied which makes use of the special structure of the problem (Tamura 1974).

The results of an optimal emergency control along 4 reaches of the river Cam in England are shown in Fig. 14.4. One sees that in the fourth reach almost nothing is felt of the accident in the first reach. The solution shows clearly the anticipatory action in the downstream reaches: the efficiency of the treatment plants is increased before the water quality in the receiving reach deteriorates. This kind of control could not have occurred with the previously discussed local feedback schemes. On the other hand, it is an indication that the performance index chosen may not be as reasonable as originally thought: the optimal control may create positive deviations from the nominal water quality just before the pollutant peak arrives, and these deviations ought to be considered beneficial in the performance index. But the algorithm applied could also be used for a problem with a more realistic performance index, since any reasonable index will be convex.

ACKNOWLEDGEMENT

This contribution is based on a chapter of the book, Modeling and Control of River Quality, by S. Rinaldi, R. Soncini-Sessa, H. Stehfest, and H. Tamura. The author is indebted to his colleagues for the stimulating and fruitful collaboration during the writing of this book.

REFERENCES

Bryson, A.E., and Y.C. Ho (1975). Applied Optimal Control. Hemisphere Publishing Corp., Washington, D.C.

Rinaldi, S., R. Soncini-Sessa, H. Stehfest, and H. Tamura (1978). Modeling and Control of River Quality. McGraw-Hill Series in Water Resources and Environmental Engineering. McGraw-Hill, New York.

Tamura, H. (1974). A discrete dynamic model with distributed transport delays and its hierarchical optimization for preserving stream quality. IEEE Trans. Syst. Man Cybern. 4: 424-431.

Young, P., and B. Beck (1974). The modelling and control of water quality in a river system. Automatica 10: 455-468.

Part Three

Operational Experience

Introduction

In Parts One and Two of this volume, the focus of the material has been on recent developments in state-space modeling, either in the forecasting or the control domain. During the workshop, an important consideration was the relevance of the proposed techniques to operational experiences in the field. In almost every paper, case studies were presented to illustrate the proposed techniques, but the step from illustrative examples to operation situations is formidable.

It appears that there is a movement within the operational hydrology community to develop real-time forecasting models based on state estimation procedures. It is also important to recognize that the hydrologic forecasts are often a hierarchy of forecasts culminating with a river forecast. It is reasonable to construct extended forecasts of main river flows based on forecasts of tributary flows. These forecasts utilize subcatchment overland flow forecasts and forecasts of point and areal precipitation. By forecasting precipitation (either by radar, satellite, or subjective probability assessments), the operational forecast length can often be significantly extended.

The United States National Weather Service has been making subjective probability assessments of precipitation for a number of years and is considering conducting experimental studies into subjective quantitative precipitation forecasting. The paper by Murphy (Chapter 15) is an important contribution to the evaluation of this technique. Murphy links the subjective assessments with quantitative hydrologic forecasting models, which represents an important area for future development.

River basin development is becoming ever more complex because the demand on basin yields is increasing. In many river basins, the easily accessible or easily constructed reservoirs have been built and new reservoir sites are not available or their use restricted for environmental reasons. Since demands on existing water resources are increasing and since increased yield may not be obtained by expansion, improved operation is required if the increased demands are to be met. The improvement of operation was the focus of the workshop.

Collinge et al. (1967) reported that waste due to inefficient reservoir operation could account for up to 22 percent of the reservoir storage capacity. Waste included unnecessary releases due to the inability to forecast tributary inflow downstream of the reservoir, as well as excess releases in anticipation of floods.

The majority of the work reported in this volume has focused on forecasting techniques since it is well known that forecasting is a prerequisite for optimal control. Better forecasting of flood situations has the most dramatic impact on operational management systems. In the papers by Cole (Chapter 16), Bobinski et al. (Chapter 17) and Curtis and Smith (Chapter 18), flood forecasting was a prime operational consideration, as it was with many of the techniques described earlier. Although these papers do not utilize state estimation models, it is important for researchers to appreciate the procedures being used and the field operation methodology.

The United States National Weather Service (NWS) probably has the most extensive operational river forecast system in the world. As described in Chapter 18, through eleven regional forecast centers in the continental United States, they regularly issue river forecasts and flood warnings for 2,500 communities, and make seasonal water forecasts. The NWS system has followed the trend in deterministic hydrology by moving from index-based catchment models to conceptual catchment models. Currently, the Office of Hydrology of the NWS is conducting research on the application of state estimation models, of the type discussed earlier in this volume, to their forecasting.

As in the United States NWS, the forecasting techniques in Poland, described in Chapter 17 by Bobinski et al., and the models used on the River Dee, described by Cole in Chapter 16, are based on deterministic, conceptual hydrologic models. The operational procedures tend to have heuristic methods to tune the models during operations. One problem with these procedures is that they do not efficiently utilize the real-time information. The structure of the conceptual models is for long-term simulation, whereas the Kalman filter approach focuses on the forecasting/updating formulation. Other operational models could be cited, but Chapters 16, 17, and 18, describing techniques in the United States, Poland, and the United Kingdom, are typical of the operational state-of-the-art forecasting procedures.

None of the proposed state estimation models requires a large use of the computer. All of the models have been developed around recursive algorithms which avoid the storage of past histories of rainfall and streamflows. Furthermore, most of the models have a simple structure and require only small computers. The implication for operational forecasting is the possibility of a system using many local forecast centers, each with a minicomputer, as opposed to the regional center with a much larger computer system.

In the United States, an increasingly large percentage of flood damage is occurring on small tributaries rather than on main rivers, implying that it may be desirable to have local forecasting centers for many tributaries. How small a computer is needed for the state estimation models? At the workshop, E. Todini made forecasts for the Ombrone River in Italy using an IBM-5100 desk-top calculator. This machine can also be used as a remote terminal to receive data (precipitation and streamflow) from remote gauging stations and can send forecasts to a regional computer center.

FUTURE DIRECTIONS IN OPERATION FORECASTING/CONTROL

There is a growing awareness that river basin management may be improved if some of the techniques proposed in this volume are indeed applied. One result of the IIASA workshop was an attempt by Cole to apply the state estimation models to forecasting and control of the River Dee. The results (Institute of Hydrology, forthcoming) indicate that the models may be very well suited for operational procedures. Further work on this case study is being done as well as work in applying these models and other real-time control models to parts of the Thames

River basin in the United Kingdom. As stated earlier, the U.S. National Weather Service is supporting research in the application of state estimation techniques to operational forecasting and it is hoped that the results can be compared with the results obtained from their current methods.

Plans for applying real-time control (which will also utilize real-time forecasts) have been discussed for the Zagyva/Tarna Rivers in Hungary (Salamin and Beck 1976), the Ohre River in Czechoslovakia, and parts of the River Thames in the United Kingdom (Institute of Hydrology, forthcoming). Furthermore, the proposed Hydrological Operational Multipurpose System (HOMS) of the World Meteorological Organization (WMO 1977), includes a real-time forecasting system and has identified six river basins where feasibility studies that include hydrologic forecasting may be carried out.

REFERENCES

Collinge, V.K., R. Fisher, and R.B. Bussell (1967). The operation of regulating reservoirs. International Association of Scientific Hydrology, Gardia Publ. 71, pp. 683-695.

Salamin, A., and M.B. Beck (1976). Some control problems in the management of a Hungarian river basin. Report CAMS/76/6b, Dept. of Engineering, University of Cambridge, Cambridge, England.

World Meteorological Organization (1977). Proposal for a plan for the development of HOMS. Geneva, Switzerland, Jan.

15 Subjective Quantification of Uncertainty in Real-Time Weather Forecasts in the United States

Allan H. Murphy

The real-time weather forecasts that are currently issued to the public and to specific users by the National Weather Service (NWS) in the U.S. are subjective forecasts. The process of formulating these forecasts is a judgmental process that involves the intuitive assimilation of information from a variety of sources by NWS forecasters. It should be noted that two sources of "objective" information play important roles in the formulation of these forecasts: the basic output (e.g., analyses, predictions) associated with numerical weather prediction (NWP) models, and the guidance forecasts produced by model output statistics (MOS) procedures. MOS procedures and forecasts are described in some detail by Klein and Glahn (1974) and Glahn and Klein (1975). Thus, while public weather forecasts are subjective forecasts, they are based in part upon objective analyses and forecasts.

It is generally agreed that weather forecasts expressed in probabilistic terms offer at least two important advantages over traditional categorical or deterministic forecasts. First, a probabilistic mode of expression provides forecasters with a means of quantitatively describing the uncertainty inherent in their forecasts. Second, probabilistic forecasts provide potential users of forecasts with information needed to make rational decisions in uncertain situations. Since real-time weather forecasts are subjective forecasts, it is important to determine whether forecasters can subjectively quantify their uncertainty in a reliable and skillful manner. MOS guidance forecasts for certain variables are also expressed in probabilistic terms (Klein and Glahn 1974, Glahn and Klein 1975).

The purpose of this paper is to summarize some of the results of operational and experimental programs in the U.S. in which NWS forecasters have formulated subjective probability forecasts. We briefly describe the NWS operational (point) precipitation probability forecasting program and present some representative results of this program. Some of the results of two recent experiments in which point and area precipitation probability forecasts were formulated by NWS forecasters are described and compared, and we summarize the results of two experiments in which NWS forecasters expressed the uncertainty in temperature forecasts in terms of credible intervals. A recent experiment involving the quantification of uncertainty in tornado forecasts is also briefly described.

After several years of experimentation, the NWS initiated a nationwide probability of precipitation (PoP) forecasting program in 1965. Under the PoP program, precipitation probabilities are routinely appended to public weather forecasts, which are generally issued 4 times a day by NWS forecasters. These probabilities are valid for either 3 consecutive 12-hour periods (e.g., "today", "tonight," and

"tomorrow") or a 6-hour period followed by two 12-hour periods (e.g., "this afternoon," "tonight," and "tomorrow"). A typical PoP forecast might state that "the precipitation probabilities for Denver (Colorado) today, tonight, and tomorrow are 30 percent, 20 percent, and 40 percent, respectively".

PoP forecasts are subjective forecasts. Specifically, a PoP forecast expresses the forecaster's "degree of belief" that a measurable amount of precipitation (i.e., $\geq$ 0.01 in.) will occur during a specified six-hour or twelve-hour period at a particular point in the forecast area (generally the official raingauge). It is generally assumed that the probability of precipitation is the same at every point in the area. In this case, a PoP forecast also represents an average point probability forecast. Forecasters assimilate information from a variety of sources in the process of formulating these (and other) forecasts. In this regard, a recent survey of NWS forecasters (Murphy and Winkler 1973; see also Murphy and Winkler 1974b) provides some information concerning the process used by the forecasters in formulating PoP forecasts. The results of this survey indicate that, while the relative importance and order of examination of various information sources differ from forecaster to forecaster and even from occasion to occasion for a particular forecaster, the output of numerical models plays an important role in the formulation of these forecasts - directly in the form of numerical prognoses and other NWP products and indirectly in the form of objective precipitation probability forecasts produced by an MOS procedure.

During the 10 years since the initiation of the PoP program, several million precipitation probability forecasts have been formulated and issued to the public in the U.S. These forecasts have been subjected to rather extensive and detailed evaluations, both on a national basis (e.g., see Sadowski and Cobb 1974) and on a regional basis (e.g., see Cummings 1974; Hughes 1976). In this paper we shall briefly examine the reliability and skill of a representative sample of PoP forecasts. Reliability refers to the degree of correspondence between forecast probabilities and observed relative frequencies over a collection of forecasts. Skill, on the other hand, refers to the "accuracy" of the forecasts relative to some standard of comparison such as climatology, where accuracy is defined as the average degree of correspondence between individual forecasts and observations over a collection of forecasts.

First we examine the 12 hour PoP forecasts formulated by forecasters in the Southern Region of the NWS during the period from April 1973 through March 1974 (see Cummings 1974). The Southern Region comprises most of the eastern two-thirds of the southern half of the U.S. (ten states). The reliability of these forecasts is depicted in Fig. 15.1, in which the forecast probabilities are plotted against the observed relative frequencies for (a) the first period (0-12 hours), (b) the second period (12-24 hours), and (c) the third period (24-36 hours). Each reliability diagram is based on approximately 38,650 forecasts. In formulating these forecasts, the forecasters were permitted to assign the probabilities 0 percent, 2 percent, 5 percent, 20 percent, ..., 90 percent, and >95 percent to the occurrence of measurable precipitation. The plotted points, then, indicate the observed relative frequencies of precipitation when these probability values were used. For example, the relative frequency of occurrence of precipitation on those occasions on which the forecasters assessed a probability of 40 percent for the first period was 39.6 percent (see Fig. 15.1a). We have also entered the number of forecasts next to each point on three diagrams; thus, forecasts of 40 percent for the first period were issued on 2,001 occasions, and precipitation actually occurred on 792 (39.6 percent). The diagonal 45°-line in each diagram represents perfect reliability, in the sense that the relative frequencies exactly equal the probabilities.

Examination of Fig. 15.1 reveals that these forecasts were quite reliable - the observed relative frequencies corresponded quite closely to the forecast probabilities - for all three periods or lead times. The first-period forecasts,

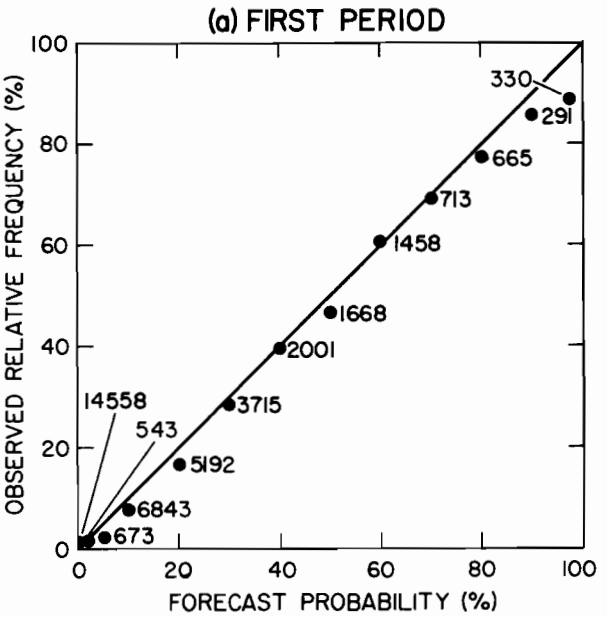


Fig. 15.1a

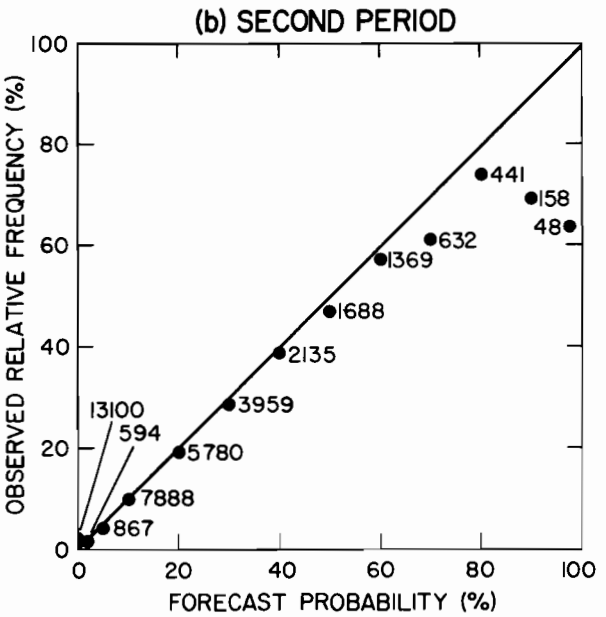


Fig. 15.1b

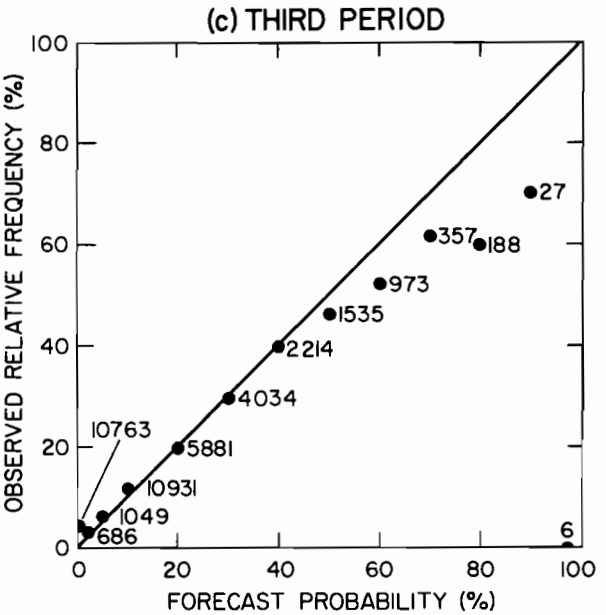


Fig. 15.1 The forecast probabilities versus the observed relative frequencies for the precipitation probability forecasts issued by NWS forecasters in the Southern Region during the period from April 1973 through March 1974 (see Cummings 1974). (a) First-period (0-12 hr) forecasts. (b) Second-period (12-24 hr) forecasts. (c) Third-period (24-36 hr) forecasts.

in particular, were very reliable. A tendency did exist for the forecasters to overforecast at the upper end of the probability scale, and this tendency increased as lead time increased. That is, the probability values were greater than the relative frequencies for the 80 percent, 90 percent, and >95 percent forecasts in the first period, and the tendency to overforecast appeared to extend down to the 70 percent and 50 percent forecasts in the second and third periods, respectively. In addition, it should be noted that the frequency of use of probability values at the upper end of the scale is considerably less than that at the lower end of the scale, and this difference also increased as lead time increased. These results, of course, simply reflect the current state of the art in precipitation forecasting and are due, at least in part, to the fact that it is not possible for forecasters to assign high probabilities frequently and in a completely reliable manner to an event that has an average (regionwide) climatological point probability of occurrence of approximately 0.17. Nevertheless, the reliability of these subjective precipitation probability forecasts is excellent for the first-period forecasts, and it is still very good for the second-period and third-period forecasts except for those forecasts associated with the highest probability values. Clearly, the forecasters were able to distinguish between those occasions on which, for example, probabilities of 30 percent and 40 percent should have been assigned to the occurrence of precipitation (for all three periods).

While reliability is an important attribute of precipitation probability forecasts, it is not the only or necessarily even the most important attribute. After all, forecasts based solely upon the long-term climatological probability of precipitation would also be quite reliable. Such forecasts would not, however, be very

accurate. The Brier score (Brier 1950) is an evaluation measure widely used in meteorology to assess the accuracy of probabilistic forecasts. This score is simply the mean square error of the (probabilistic) forecasts, in which an individual observation assumes a value of one if precipitation occurs and zero otherwise. The skill of the forecasts can then be determined by computing the percent improvement of the Brier score for the forecasts over the Brier score for the relevant climatological probabilities. The results of such a computation for the Southern Region PoP forecasts indicate that these forecasts represented a 27.5 percent, 16.9 percent, and 8.3 percent improvement over climatology for the first, second, and third periods, respectively (see Cummings 1974). As expected, skill decreased as lead time increased. Nevertheless, these results indicate that the PoP forecasts were both reliable and skillful.

Space limitations preclude a more detailed examination of these (and other) operational precipitation probability forecasts. For a recent evaluation of another set of subjective PoP forecasts, including a discussion of the forecasts formulated by individual forecasters, see Murphy and Winkler (1977a). A paper containing a comprehensive review of the PoP forecasting program is in preparation.

PRECIPITATION PROBABILITIES: EXPERIMENTAL POINT AND AREA FORECASTS

The Experiments

The purposes of these experiments, which were conducted in the NWS Forecast Office (WSFO) in St. Louis, Missouri, and in the NWS Office (WSO) in Rapid City, South Dakota, were to investigate the ability of forecasters to differentiate among different points in a forecast area with regard to the likelihood of occurrence of measurable precipitation, and the relative ability of forecasters to make point and area (including areal coverage) precipitation probability forecasts.

Precipitation probabilities have been included in weather forecasts in the U.S. on a regular basis for more than 10 years, and the evidence available suggests that these probabilities are considered to be an important and integral part of the NWS's public weather forecasting program. However, many aspects of precipitation probability forecasting are in need of further detailed investigation. For example, many users are interested in the occurrence of (measurable) precipitation at points in the forecast area for which the probability of precipitation may be quite different from what it is at the official raingauge. The present practice of issuing a single point probability forecast for an entire area clearly does not satisfy the requirements of such users, unless the point probability is indeed the same at every point in the area.

If the forecast area of concern is defined in terms of a finite set of K points, then the average point probability is $\bar{p} = \Sigma p_i/K$, where p_i represents the point probability of precipitation at point i. In the special case in which the point probability is uniform over the forecast area, $\bar{p}$ is equal to the common value of the point probabilities. However, since $\bar{p}$ is a summary measure, it does not, in general, contain all of the information provided by the set of individual point probabilities. When we are concerned with the forecast area as a whole rather than with the individual points making up the area, then an area probability, which is defined as the probability that precipitation will occur somewhere in the forecast area, may be of interest. The area probability, a, and δ_i, an indicator variable equal to 1 if precipitation occurs at point i and 0 otherwise, have the following relation: $a = P(\Sigma\delta_i > 0)$. Since the occurrence of precipitation at a point implies the occurrence of precipitation in the area, the area probability must be at least as large as each point probability, which implies that $a \geq \max p_i$. Another measure relating to the entire area is the expected areal coverage, e, and this measure can be defined either conditionally or unconditionally. The

unconditional expected areal coverage, e_u, is the expected proportion of the forecast area over which precipitation will occur. Thus, $e_u = E(\Sigma\delta_i/K) = \Sigma E(\delta_i)/K = \Sigma p_i/K = \bar{p}$. The conditional expected areal coverage, on the other hand, is the expected proportion of the forecast area over which precipitation will occur given that precipitation occurs at one or more points in the area. This measure, which we shall denote by e_c, is necessarily greater than or equal to e_u (and, as a result, $e_c \geq \bar{p}$). For further discussion of these (and other) relationships, see Epstein (1966), Curtiss (1968), Winkler and Murphy (1976), and Murphy and Winkler (1977b).

The subjects in the experiments were 14 forecasters in the St. Louis WSFO and 9 forecasters in the Rapid City WSO. Each time they were on forecasting duty during the periods of the experiments (November 1972 - March 1973 in St. Louis and June - September 1974 in Rapid City), the forecasters made point and area precipitation probability forecasts for the St. Louis and Rapid City areas. In particular, the forecasters were asked for an average point probability of measurable precipitation for the entire forecast area; point probabilities of measurable precipitation at specific points in the forecast area (5 points in the St. Louis area and 4 in the Rapid City area); an area probability of measurable precipitation for the forecast area; and the unconditional (conditional) expected areal coverage of the forecast area by measurable precipitation (unconditional for St. Louis and conditional for Rapid City). The forecast areas were defined by 20 and 10 points (i.e., raingauges), respectively, in the St. Louis and Rapid City experiments. In St. Louis, the forecasts were made for 3 different 12-hour periods (on the day shift, for "tonight," "tomorrow," and "tomorrow night"; on the midnight shift, for "today," "tonight," and "tomorrow"), while at Rapid City the forecasts were made for only one 12-hour period (on the day shift, for "tonight"; on the midnight shift, for "today"). In total, 771 and 222 sets of forecasts were made in the St. Louis and Rapid City experiments, respectively.

Some Results

A brief summary and comparison of part of the results from the St. Louis and Rapid City experiments is presented here. For a more detailed description, see Winkler and Murphy (1976) and Murphy and Winkler (1977b), respectively.

Point probability forecasts for specific points. The averages of the probability forecasts, and certain functions of the forecasts, for both the St. Louis and Rapid City experiments are presented in Table 15.1. With regard to the individual point probabilities, the average values of the point probabilities are very close for the 5 points at St. Louis and not so close for the 4 points at Rapid City. The points which, on the average, were assigned higher probabilities at Rapid City are located in portions of the forecast area in which topographic features would be expected, a priori, to increase the frequency of occurrence of measurable precipitation. A sample s_p^2 (see definition in Table 15.1) was computed for each set of point probabilities in the experiments. The average values of s_p^2 indicate that differences among the five point probabilities in the St. Louis experiment were infrequent in the individual sets of forecasts (as well as on the average), while such differences were considerably more frequent among the four points in the Rapid City experiment. In this regard, a closer examination of the forecasts in the St. Louis experiment reveals that for 619 (80.3 percent) of the 771 forecasts, $p_1=p_2=p_3=p_4=p_5$. On the other hand, $p_1=p_2=p_3=p_4$ in the Rapid City experiment for only 38 (17.1 percent) of the 222 forecasts.

The relative frequencies of measurable precipitation at the points, or raingauges, defining the forecast areas in the experiments are presented in Table 15.2. The first entries in each list are the relative frequencies at the raingauges for which point forecasts were made in the experiments. These values ranged from 0.200 to 0.257 (0.223 to 0.257 for the five point-forecast gauges) in the St. Louis experiment and from 0.045 to 0.171 (0.104 to 0.167 for the four point-forecast gauges) in the

Table 15.1. Averages and selected function of the point probability forecasts

Probabilities and functions of probabilities	St. Louis experiment (n=771)	Rapid City experiment (n=222)
p_1	0.226	0.083
p_2	0.227	0.123
p_3	0.233	0.205
p_4	0.222	0.172
p_5	0.231	
$\bar{p}_f$	0.229	0.119
m	0.244	0.213
s_p^2	0.001	0.007
a	0.240	0.356
e	0.229	0.274

NOTE: $p_1,\ldots,p_5$ are the point probabilities, $\bar{p}_f$ is the forecast average point probability, $m = \max\{p_i\}$ is the largest point probability, $s_p^2 = \Sigma_i(p_i-\bar{p}_c)^2/4$ (St. Louis) or $\Sigma_i(p_i-\bar{p}_c)^2/3$ (Rapid City) is the sample variance of the point probabilities, a is the area probability, e is the expected areal coverage, and $\bar{p}_c$ is the sample mean of the point probabilities. $e = e_u$ for the St. Louis experiment and $e = e_c$ (n=90) for the Rapid City experiment.

Rapid City experiment. Thus, some variation among the points in terms of relative frequency of precipitation existed in the St. Louis experiment, but the amount of variation was not great. This result suggests that the lack of variation among the forecasters' point probability forecasts was justified, to a considerable degree, by the "weather situations" that occurred during the period of the experiment. In the Rapid City experiment, on the other hand, the variation among the relative frequencies was similar in magnitude to the variation among the forecasters' average point probabilities. Moreover, a modest degree of correspondence appears to exist between the (average) point probabilities and the relative frequencies in this experiment. The gauge-to-gauge variation can also be investigated, and these data indicate that the observations were the same at all 20 (10) points for 77.7 percent (60.0 percent) of the periods at St. Louis (Rapid City); as expected, more variation among points was observed at Rapid City than at St. Louis.

Table 15.2. Relative frequency of occurrence of measurable precipitation at the rain gauges defining the forecast areas

Rain-gauge number	St. Louis experiment (n=771)	Rapid City experiment (n=222)
1/11	0.226/0.238	0.117
2/12	0.223/0.226	0.126
3/13	0.242/0.234	0.167
4/14	0.257/0.234	0.104
5/15	0.238/0.208	0.167
6/16	0.249/0.238	0.153
7/17	0.200/0.211	0.045
8/18	0.208/0.223	0.171
9/19	0.230/0.223	0.045
10/20	0.226/0.215	0.113
Average	0.227	0.122

An evaluation of the forecasts in light of the observations can provide information concerning the extent to which the forecasters were successful in differentiating among the raingauges for which individual point probabilities were assigned. Since the forecasters at St. Louis were not able to make such distinctions very often, owing in large measure to the lack of variation among the observations, their success in differentiating among the 5 points was at best very limited. At Rapid City, on the other hand, a higher average value of the point probability corresponded to a higher relative frequency at three of the four raingauges. Moreover, an examination of the average probabilities as a function of the observations revealed that the average probability assigned to gauges receiving rain was more than three times as large as the average probability assigned to gauges not receiving rain (see Murphy and Winkler, 1977b). Thus, at Rapid City (but not at St. Louis), the forecasters were able to differentiate with considerable success among different points in the forecast area in terms of the probability of precipitation.

The reliability of the point probability forecasts can be examined by plotting the forecast probabilities versus the observed relative frequencies, and a reliability curve for the entire sample of point probability forecasts in each experiment is presented in Fig. 15.2. At both St. Louis and Rapid City, the point probabilities tended to be slightly too high for low probability values (i.e., the forecasters tended to overforecast). For higher probabilities, the St. Louis forecasts tended to be slightly too low, while the subsamples of forecasts at Rapid City for these higher values were too small to reach a definite conclusion concerning overforecasting. Overall, the reliability of the individual point probability forecasts

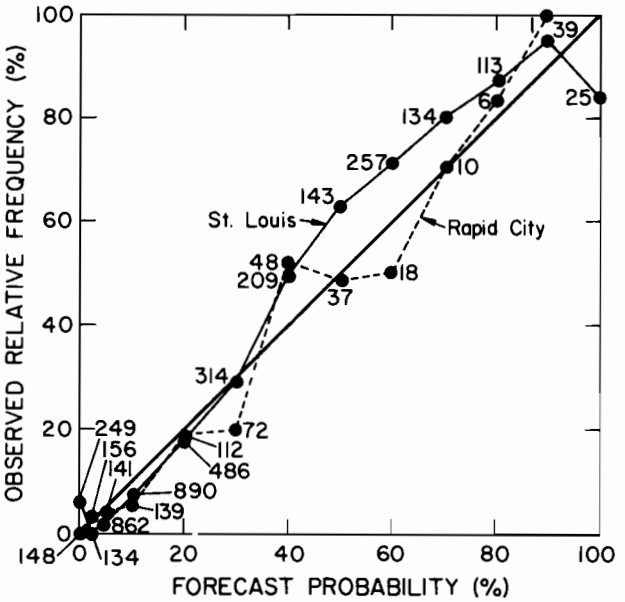


Fig. 15.2 The reliability diagram for the individual point probability forecasts formulated by NWS forecasters during the St. Louis and Rapid City experiments.

was quite good, particularly when the relatively small sample sizes are taken into consideration.

Average point probability forecasts. A comparison of the average point probabilities with the average relative frequencies of precipitation indicates that, on the average, these quantities corresponded quite closely in both experiments (cf. Tables 15.1 and 15.2). The average point probabilities are plotted against the average relative frequencies in Fig. 15.3. This figure provides further evidence that the average point probabilities in both experiments were very reliable; except for high probabilities involving very few forecasts, the reliability curves for St. Louis and Rapid City lie quite close to the 45° line (a tendency did exist for the average probability at St. Louis to be too high for higher probability values).

Area probability forecasts. From Table 15.1, the average area probability forecast for St. Louis (Rapid City) was 0.240 (0.356), while the relative frequency of precipitation "somewhere in the area" was 0.332 (0.405). On the average, then, the area probability was too low in both experiments; the underestimation at St. Louis was almost twice as great as that at Rapid City. The reliability of the area probability forecasts is indicated in Fig. 15.4. This figure reveals that the area probability tended to be less than the relative frequency throughout the range of probability values at St. Louis and for probability values less than or equal to 30 percent at Rapid City. Clearly, the forecasters underforecast at St. Louis, while the overall agreement with the 45° line was quite good at Rapid City, especially when the small number of observations for each probability value is taken into consideration.

The consistency of the point and area probabilities is also of concern. From Table 15.1, the average value of the area probability, a, was -0.004 (0.143) larger than the average value of m, the largest of the point probabilities. Thus, on the

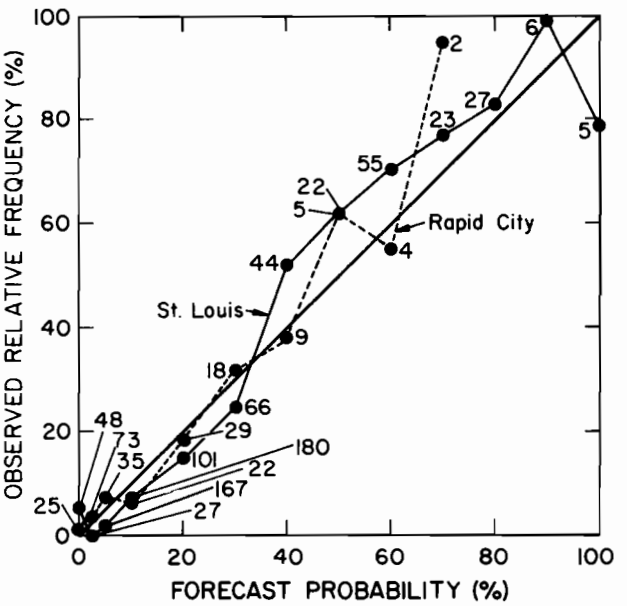


Fig. 15.3 The reliability diagram for the average point probability forecasts formulated by NWS forecasters during the St. Louis and Rapid City experiments.

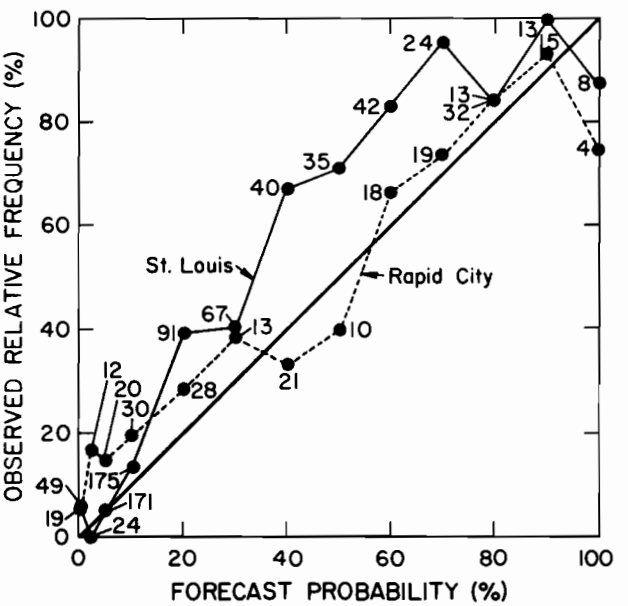


Fig. 15.4 The reliability diagram for the area probability forecasts formulated by NWS forecasters during the St. Louis and Rapid City experiments.

average, the area probability at St. Louis violated the relationship that the area probability must be at least as large as the largest point probability. In terms of individual forecasts, a < m on 7.6 percent (2.3 percent), a = m on 89.4 percent (23.4 percent), and a > m on 3.0 percent (74.3 percent) of the forecasting occasions at St. Louis (Rapid City). Theoretically, only under the condition that if precipitation occurs anywhere in the area, it must occur at the point corresponding to the largest point probability, can a equal m. Thus, the point and area probabilities at St. Louis (Rapid City) were inconsistent or only marginally consistent on 97.0 percent (25.7 percent) of the occasions. When these results are considered in conjunction with the results related to the reliability of the point and area forecasts (see Figs. 15.2 and 15.4), it seems reasonable to conclude that the area probabilities at St. Louis were inconsistent because they were too low. At Rapid City, on the other hand, the point and area probability forecasts appear to be quite reliable and consistent.

Expected areal coverage forecasts. The final forecast made on each occasion was an expected areal coverage forecast, an unconditional forecast (e_u) at St. Louis and a conditional forecast (e_c) at Rapid City. The average value of e_u (e_c) was 0.229 (0.274) (see Table 15.1) and the average observed unconditional (conditional) areal coverage was 0.227 (0.301). Thus, the overall correspondence between average forecast and average observation was quite close in both experiments. Moreover, it should be noted that the observed unconditional areal coverage in the Rapid City experiment was only 0.122, which suggests that the forecasters considered the full impact of the conditional nature of the expected areal coverage forecasts appropriately in this experiment.

As previously indicated, the unconditional (conditional) expected areal coverage should be equal to (greater than) the average point probability, $\bar{p}_f$. Note that, on the average, $e-\bar{p} = 0.000$ in the St. Louis experiment and $e_c-\bar{p}_f = 0.055$ ($n = 90$ for e_c and p_f) in the Rapid City experiment (see Table 15.1). Thus, the expected areal coverage forecasts in both experiments were consistent with the average point probability forecasts.

EXPERIMENTAL CREDIBLE INTERVAL TEMPERATURE FORECASTS

The Experiments

The purposes of these experiments, which were conducted in the WSFOs in Denver, Colorado, and Milwaukee, Wisconsin, were to investigate the ability of forecasters to express the uncertainty in their temperature forecasts in probabilistic terms and to compare two approaches (variable-width intervals and fixed-width intervals) to credible interval temperature forecasting. A credible interval represents an interval of potential values of the variable of concern together with a probability (expressing the forecaster's degree of belief) that the actual value of the variable will fall in the interval. These intervals appear to provide a promising format for probability forecasts of continuous variables such as temperature (see Peterson et al. 1972).

Two forecasters at Denver and 3 forecasters at Milwaukee worked within the framework of variable-width, fixed-probability forecasts, using 50 percent and 75 percent central credible intervals. To obtain these intervals, each forecaster was asked to make a total of five "indifference judgments" at equal odds. The first indifference judgment determines the median of the forecaster's probability distribution, while the remaining judgments determine the 25th, 12-1/2th, 75th, and 87-1/2th percentiles of the forecaster's distribution. Once these five indifference judgments are made, the 50 percent central credible interval is the interval from the 25th percentile to the 75th percentile, and the 75 percent central credible interval is the interval from the 12-1/2th percentile to the 87-1/2th

percentile. The other two forecasters in each experiment worked within the framework of fixed-width, variable-probability forecasts, using intervals of width 5°F and 9°F. First, the median of the forecaster's distribution was determined, just as in the case of the variable-width forecasts. Then, the forecaster was asked to determine probabilities for intervals of width 5°F and 9°F centered at the median. All intervals were assumed to include their end points, and all temperatures were recorded to the nearest degree.

As indicated above, the subjects in the experiments were 4 forecasters in the Denver WSFO and 5 forecasters in the Milwaukee WSFO. During the respective periods of the experiments (August 1972 - March 1973 in Denver and September 1974 - July 1975 in Milwaukee), the forecasters made credible interval temperature forecasts of maximum (high) and minimum (low) temperatures. In the Denver experiment, the forecasts were for "tonight's low" and "tomorrow's high" on the day shift and for "today's high" and "tonight's low" on the midnight shift; in the Milwaukee experiment, the forecasts were for "tonight's low," "tomorrow's high," and "tomorrow night's low" on the day shift and for "today's high," "tonight's low," and "tomorrow's high" on the midnight shift. Thus, the Denver experiment involved 12-hour and 24-hour forecasts, while the Milwaukee experiment involved 12-hour, 24-hour, and 36-hour forecasts. Overall, 127 and 233 sets of forecasts were made in the Denver and Milwaukee experiments, respectively.

Some Results

Some of the results of the experiments are summarized and compared here. The results of the Denver experiment are described in detail by Murphy and Winkler (1974a) and of the Milwaukee experiment by Winkler and Murphy (1977).

The first task on each forecasting occasion for all of the participants in both experiments was to determine a median. A comparison of the median temperatures (MTs) with the corresponding observed temperatures (OTs) is presented in Table 15.3a. The MTs would be completely reliable if the percentage of the time that MT > OT equaled the percentage of the time that MT < OT. The percentages presented in Table 15.3a indicate that a tendency existed for the MTs to underestimate the OTs, and this tendency was slightly greater in the Milwaukee experiment than in the Denver experiment. These results are further supported by the average differences between MT and OT, which are negative for both experiments, with the difference at Milwaukee slightly exceeding that at Denver. On the other hand, the average absolute differences between MT and OT are essentially equal in the two experiments. For comparative purposes, the official temperature forecast (FT) issued to the public was recorded on each occasion, and the average absolute differences between FT and OT presented in Table 15.3a indicate that the medians determined by the forecasters for the purposes of the experiments were, on the average, comparable to the official forecasts as point forecasts of high and low temperatures.

Climatological median temperatures (CTs) provide a convenient standard with which to compare MT as a point forecast. The climatological forecasts considered here are median high and low temperatures based upon historical data for the 5-yr periods immediately preceding the respective experiments, and they were computed on a monthly basis. These forecasts were analyzed in the same manner as the forecasters' assessed medians, and the results are presented in Table 15.3b. As in the case of MT, CT exhibited a slight tendency to underestimate OT (CT < OT more often than CT > OT). Thus, the forecasters' tendency to underestimate may be due in part to above normal temperatures during the experimental periods. In any event, these tendencies are not strong, and the MTs appear to be quite reliable point forecasts. It is also of interest to note that the average value of $|CT - OT|$ is approximately twice as large as the average value of $|MT - OT|$ in both experiments.

Table 15.3. Forecasters' medians and climatological medians

(a) Forecasters' medians

Experiment	Number of forecasts	Percentages			Average (°F)		
		MT>OT	MT=OT	MT<OT	MT-OT	|MT-OT|	|FT-OT|
Denver	254	39.4	12.6	48.0	-0.5	3.8	3.9
Milwaukee	699	37.2	9.4	53.4	-0.8	3.7	3.7

(b) Climatological medians

Experiment	Number of forecasts	Percentages			Average (°F)	
		CT>OT	CT=OT	CT<OT	CT-OT	|CT-OT|
Denver	254	39.4	3.1	57.5	0.6	8.9
Milwaukee	699	40.6	4.6	54.8	-1.3	7.1

NOTE: MT stands for median temperature, OT for observed temperature, FT for forecast temperature, and CT for climatological temperature.

The results presented in Table 15.4a indicate that the variable-width forecasts were quite reliable, in the sense that the observed relative frequencies below, in, and above the variable-width intervals are close to the probabilities of the intervals (0.25, 0.50, and 0.25, respectively, for the 50 percent intervals; 0.125, 0.750, and 0.125, respectively, for the 75 percent intervals). Note that the relative frequencies in the intervals were slightly too low in the Denver experiment, while these frequencies were slightly too high in the Milwaukee experiment. The fact that the relative frequencies above the intervals were greater than the relative frequencies below the intervals in both experiments lends further support to the above-mentioned result that the MTs tended to underestimate the OTs. Table 15.4a also indicates that the widths of the 50 percent and 75 percent intervals were slightly narrower, on the average, in the Milwaukee experiment than in the Denver experiment.

Climatology can also be used as a standard of comparison for interval forecasts, and climatological variable-width interval forecasts were generated by determining the appropriate percentiles from the 5 years of historical data on a monthly basis. The performance of these climatological forecasts is summarized in Table 15.4b. An examination of the percentages of observations below, in, and above the intervals indicates that the climatological intervals were not quite as reliable as the intervals determined by the forecasters. Moreover, the average width of the climatological intervals is more than twice as great as the average width of the forecasters' intervals (cf. Table 15.4). Thus, the forecasters' variable-width intervals were more reliable and much more precise than intervals based on climatological data.

The results for the forecasters' fixed-width intervals and the corresponding climatological intervals are presented in Table 15.5. The average probabilities assigned by the forecasters to the 5°F and 9°F fixed-width intervals in the Denver

Table 15.4. Relative frequency of occurrence of observed temperature, and average interval width, for variable-width forecasts and climatological forecasts corresponding to variable-width forecasts

Experiment	Number of forecasts	Relative Frequencies[a]						Average Width (°F)	
		50% Intervals			75% Intervals			50% Intervals	75% Intervals
		BI	II	AI	BI	II	AI		
(a) Variable-width forecasts									
Denver	132	0.258	0.455	0.288	0.106	0.735	0.159	6.2	11.7
Milwaukee	432	0.181	0.539	0.280	0.081	0.794	0.125	5.9	10.1
(b) Climatological forecasts									
Denver	132	0.311	0.447	0.242	0.189	0.652	0.159	14.8	24.2
Milwaukee	432	0.178	0.569	0.252	0.081	0.817	0.102	14.5	23.7

[a] Below interval BI, in interval II, and above interval AI.

Table 15.5. Average probability assigned to intervals and observed relative frequency of observations in intervals for fixed-width forecasts and climatological forecasts corresponding to fixed-width forecasts

Experiment	Number of forecasts	Average Probabilities		Relative Frequencies	
		5°F Intervals	9°F Intervals	5°F Intervals	9°F Intervals
(a) Fixed-width forecasts					
Denver	122	0.60	0.80	0.46	0.66
Milwaukee	267	0.47	0.72	0.40	0.66
(b) Climatological forecasts					
Denver	122	0.23	0.37	0.19	0.43
Milwaukee	267	0.22	0.37	0.19	0.36

experiment differed considerably from the relative frequencies with which the OT fell in these intervals (see Table 15.5a). For both the 5°F and 9°F intervals, the average probability was 0.14 higher than the relative frequency. In the Milwaukee experiment, the differences between the average probabilities and the relative frequencies were considerably smaller; namely, 0.07 and 0.06 for the 5°F and 9°F intervals, respectively. Note, however, that the probabilities assigned to the intervals were, on the average, larger than the observations indicate that they should have been in all four "situations." In this regard, it is interesting to observe that in the Denver (Milwaukee) experiment the average width of the 50 percent variable-width intervals was 6.2°F (5.9°F), whereas the 5°F fixed-width intervals were assigned an average probability of 0.60 (0.47); the 75 percent variable-width intervals averaged 11.7°F (10.1°F) in width, whereas the 9°F fixed-width intervals averaged 0.80 (0.72) in probability. On the average, then, the fixed-width intervals have higher probabilities for narrower intervals when compared with the variable-width intervals in the Denver experiment (but not in the Milwaukee experiment). Narrower intervals are desirable provided that they are reliable, but the fixed-width intervals in the Denver experiment were not very reliable because they were too narrow. Moreover, the climatological fixed-width intervals were more reliable than the forecasters' intervals at both Denver and Milwaukee (cf. Tables 15.5a and 15.5b).

The results of the Denver and Milwaukee experiments provide strong support for the tentative conclusion reached by Peterson et al. (1972, p.969) that "weather forecasters can use credible intervals to describe the uncertainty inherent in their temperature forecasts." These results are particularly encouraging in view of the fact that the forecasters had no experience and very little training in making such forecasts. For a more detailed evaluation of the results of these experiments, including an examination of the forecasts formulated by individual forecasters, refer to Murphy and Winkler (1974a, 1977c) and Winkler and Murphy (1977).

TORNADO PROBABILITIES: EXPERIMENTAL FORECASTS

The Experiment

Forecasts of severe weather in the U.S. are formulated by NWS forecasters at the National Severe Storms Forecast Center (NSSFC) in Kansas City, Missouri. NSSFC issues two types of severe weather forecasts - severe weather outlooks and severe thunderstorm and tornado watches. In this context, severe weather is defined as the occurrence of one or more tornadoes and/or the occurrence of thunderstorms accompanied by hail greater than $\frac{3}{4}$ in. in diameter, by damaging winds, or by wind gusts exceeding 50 knots. An outlook is issued each day in the early morning hours (approximately 0900 GMT) and identifies those areas in the U.S. (if any) that are expected to experience severe weather during the subsequent 24 hours (1200-1200 GMT). A watch, on the other hand, can be issued at any time that one of the following events is expected to occur: one or more tornadoes - a tornado watch - and one or more severe thunderstorms (as defined above) - a severe thunderstorm watch. The watches are generally defined over a rectangular "box" approximately 25,000 square miles in area and are usually valid for a period of 4 to 6 hours. Most of the severe thunderstorm and tornado watches are issued during the period from February through June, although severe weather can occur in any month of the year.

Severe weather forecasts are generally expressed in categorical terms. However, an experiment was recently undertaken at NSSFC to determine whether NWS forecasters could subjectively quantify the uncertainty inherent in these forecasts in a reliable and skillful manner. Two groups of forecasters participated in the experiment - one group was concerned with the severe weather outlooks and the other group was concerned with severe thunderstorm and tornado watches. The forecasters who regularly formulate outlooks were asked to assign probabilities to the following events on an experimental basis: one or more tornadoes will occur in any of the severe weather areas identified in the outlook of the "day" in question (O1 forecast), and 10 or more tornadoes will occur anywhere in the U.S. on that day (O2 forecast). The forecasters who formulate watches were asked to assign probabilities to the following events: one or more tornadoes will occur within the watch (W1 forecast); three or more tornadoes will occur within the watch (W2 forecast); and at least one tornado occurring within the watch will attain an F rating of two or more on the FPP scale (W3 forecast). The FPP scale is a composite measure of tornado intensity, involving maximum wind speed, path length, and path width (see Fujita, 1973). An F rating of two refers to tornadoes with maximum wind speeds in the range from 113 to 157 miles/hr. Probabilities were assigned to these three events in connection with both severe thunderstorm and tornado watches. The period of the experiment was February through June 1976. In total, 92 O1 and O2 forecasts, 241 W1 and W2 forecasts, and 223 W3 forecasts were made during this 5 month period.

Some Results

We consider only the reliability of these experimental forecasts. The preliminary results of this experiment are described in greater detail by Murphy and Winkler (1977d).

The reliability of the O1 and O2 forecasts is indicated in Fig. 15.5, in which the observed relative frequency is plotted against the forecast probability. Examination of the curve for the O1 forecasts reveals that the forecasters tended to overforecast for this event. Nevertheless, these results are encouraging in that the relative frequency is a monotonically increasing function of the probability over the range of probability values from 40 percent to 100 percent (this range includes 77 of the 92 forecasts). Thus, the forecasters were able to "order" most of the forecasting occasions according to the relative likelihood of occurrence of

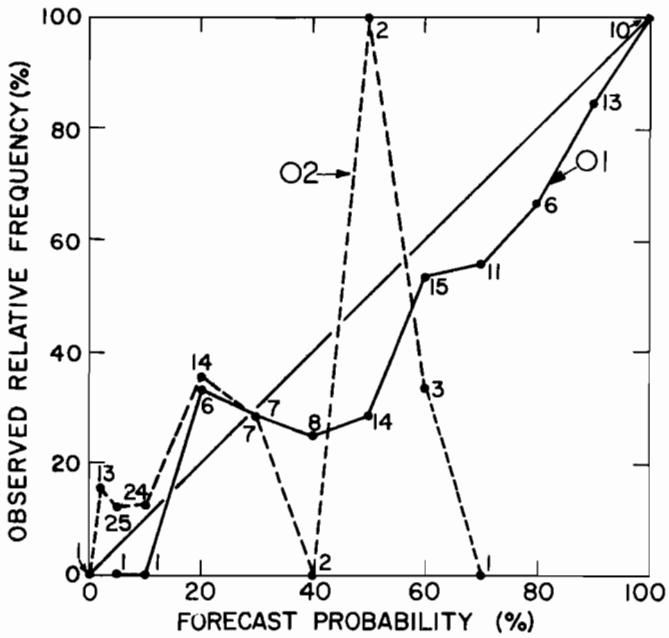


Fig. 15.5 The reliability diagram for the experimental outlook forecasts formulated by NWS forecasters during the NSSFC experiment.

the event of concern, even though their forecasts were not particularly reliable in terms of the difference between the probabilities and the relative frequencies.

Figure 15.5 reveals that the O2 forecasts were not very reliable. In this case, the forecasters exhibited a tendency to underforecast - the forecast probabilities are less than the relative frequencies for most of the probability values involving a significant number of forecasts. It should be noted that the large departures from the 45^{o} line, which are primarily associated with the higher probability values, involve very few forecasts.

The reliability diagram for the experimental watch forecasts is presented in Fig. 15.6. Examination of the curve for the W1 forecasts (one or more tornadoes in the watch area) indicates that the forecasters tended to overforecast for this event, and this tendency was most pronounced for probability values greater than or equal to 40 percent. However, the reliability of these forecasts is actually quite good, particularly when the forecasters' lack of experience in probabilistic forecasting and the relatively small sample size are taken into account. Moreover, with the exception of the 10 forecasts associated with a probability value of 90 percent, relative frequency is a monotonically increasing function of probability over the entire range of probability values used for these forecasts.

The reliability of the W2 forecasts (three or more tornadoes in the watch area) and the W3 forecasts (one or more F2 or stronger tornadoes in the watch area) is not as good as that of the W1 forecasts. The W2 forecasts exhibit only a slight tendency toward overforecasting, but no monotonically increasing relationship exists between observed relative frequencies and forecast probabilities for these forecasts. The curve for the W3 forecasts reveals a consistent and relatively strong tendency to underforecast for all probability values. However, it is encouraging to note

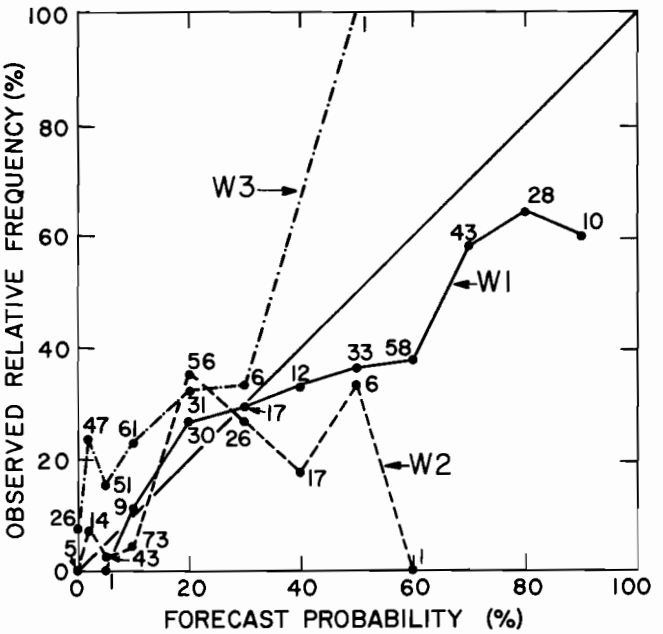


Fig. 15.6 The reliability diagram for the experimental watch forecasts formulated by NWS forecasters during the NSSFC experiment.

that the relative frequency is a monotonically increasing function of the probability for probability values greater than or equal to 5 percent for these forecasts.

SUMMARY AND CONCLUSION

The results of the area probability and point probability precipitation experiments have important implications for operational forecasting practices, particularly in locations in which a considerable amount of variation exists among points within the forecast area in terms of precipitation occurrence. The positive results concerning area probability forecasts at Rapid City provide encouraging evidence that forecasters can successfully make this different type of forecast.

The results of the temperature forecast experiments indicate that weather forecasters can formulate both reliable and skillful credible interval temperature forecasts, in the sense that the forecasters' intervals were more reliable and much more precise (i.e., narrower) than the corresponding climatological intervals at both Denver and Milwaukee. These results, which are particularly encouraging in view of the fact that the forecasters had no experience and little training in making such forecasts, also have important implications for the practice of temperature forecasting.

The preliminary results of the NSSFC tornado and thunderstorm forecasting experiments are also encouraging, particularly when the forecasters' lack of probability forecasting experience and the small sample sizes are taken into account.

The results of these operational and experimental programs provide convincing evidence that the output of numerical models, together with the knowledge and

experience of NWS forecasters, represents a sound basis for the formulation of subjective probability forecasts. In particular, these results indicate that NWS forecasters can subjectively quantify the uncertainty in real-time forecasts of a variety of weather variables in a reliable and skillful manner. In view of the fact that the value of probabilistic forecasts, in general, exceeds the value of either climatological or categorical forecasts (see Murphy 1977), the results of these programs appear to have important implications for operational forecasting practices.

In addition to the nationwide PoP program of the NWS and the experiments described in this paper, several other experimental and operational programs in subjective probability forecasting in meteorology have been undertaken in recent years. In this regard, an experiment involving probabilistic forecasts of minimum temperatures has been conducted during April and May in Albuquerque, New Mexico, since 1971 (see Gregg 1977), and these forecasts are now routinely disseminated to orchardists in New Mexico during the spring season. Moreover, subjective probability forecasts of various precipitation and temperature events are routinely formulated on an experimental basis in conjunction with the wide-ranging laboratory activities at several universities (e.g., Sanders 1973, Bosart 1975). For recent references concerned with various aspects of probability forecasting in meteorology, see Murphy (1976a). A comprehensive list of references related to subjective probability forecasting programs will be found in Murphy (1976b).

Finally, it should be mentioned that subjective probability forecasts, or assessments, are now being made in many different fields, including medical diagnosis, military intelligence, educational testing, and business, as well as in meteorology. Moreover, psychologists and others have conducted numerous experiments in laboratory settings involving quantification of judgements and decision making under uncertainty. For recent review papers concerned with subjective probability assessments in these areas, see Hogarth (1975), Lichtenstein et al. (1977), and Slovic et al. (1977).

ACKNOWLEDGEMENTS

The author is grateful to Mr. A.D. Cummings of the Southern Region Headquarters of the National Weather Service for permission to reproduce the results of the PoP program. The other experiments were designed and conducted in collaboration with Dr. R.L. Winkler of Indiana University.

REFERENCES

Bosart, L.F. (1975). SUNYA experimental results in forecasting daily temperature and precipitation. Mon. Weather Rev. 103: 1013-1020.

Brier, G.W. (1950). Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78: 1-3.

Cummings, A.D. (1974). Some aspects of PoP forecast performance by Southern Region WSFOs - revisited. Fort Worth, Texas, NOAA, National Weather Service, Tech. Memo. SR-77.

Curtiss, J.H. (1968). An elementary mathematical model for the interpretation of precipitation probability forecasts. J. Appl. Meteorol. 7: 3-17.

Epstein, E.S. (1966). Point and area precipitation probabilities. Mon. Weather Rev. 94: 595-598.

Fujita, T.T. (1973). Tornadoes around the world. Weatherwise 26: 56-83.

Glahn, H.R., and W.H. Klein (1975). Present status of statistical weather forecasting. Preprints of Fourth Conference on Probability and Statistics in Atmospheric Sciences (Tallahassee). American Meteorological Society, Boston, Massachusetts.

Gregg, G.T. (1977). Probability forecasts of a temperature event. National Weather Digest 2(2): 33-34.

Hogarth, R.M. (1975). Cognitive processes and the assessment of subjective probability distributions. J. Am. Stat. Assoc. 70: 271-294.

Hughes, L.A. (1976). On the verification of forecasts. Preprints of Sixth Conference on Weather Forecasting and Analysis (Albany). American Meteorological Society, Boston, Massachusetts.

Klein, W.H., and H.R. Glahn (1974). Forecasting local weather by means of model output statistics. Bull. Am. Meteorol. Soc. 55: 1217-1227.

Lichtenstein, S., B. Fischhoff, and L.D. Phillips (1977). Calibration of probabilities: the state of the art, in H. Jungermann and G. de Zeeuw, Eds., Decision Making and Change in Human Affairs. D. Reidel, Dordrecht, Holland.

Murphy, A.H. (1976a). Probability forecasts in meteorology: status, problems, and prospects. Preprints of Sixth Conference on Weather Forecasting and Analysis (Albany). American Meteorological Society, Boston, Massachusetts.

Murphy, A.H. (1976b). Subjective probability forecasting in meteorology: a review of operational and experimental results, unpublished outline. National Center for Atmospheric Research, Boulder, Colorado.

Murphy. A.H. (1977). The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. Mon. Weather Rev. 105: 803-816.

Murphy, A.H. (1978). The National Weather Service probability of precipitation forecasting program: the first ten years, in preparation. National Center for Atmospheric Research, Boulder, Colorado.

Murphy, A.H., and R.L. Winkler (1973). National Weather Service forecasters and probability forecasts: preliminary results of a nationwide survey, unpublished report. National Center for Atmospheric Research, Boulder, Colorado, and Indiana University, Bloomington, Indiana.

Murphy, A.H., and R.L. Winkler (1974a). Credible interval temperature forecasting: some experimental results. Mon. Weather Rev. 102: 784-794.

Murphy, A.H., and R.L. Winkler (1974b). Probability forecasts: a survey of National Weather Service forecasters. Bull. Am. Meteorol. Soc. 55: 1449-1453.

Murphy, A.H., and R.L. Winkler (1977a). Reliability of subjective probability forecasts of precipitation and temperature. Applied Statistics 26: 41-47.

Murphy, A.H., and R.L. Winkler (1977b). Point and area precipitation probability forecasts for a forecast area with significant local effects. Atmosphere 15: 61-78.

Murphy, A.H., and R.L. Winkler (1977c). The use of credible intervals in temperature forecasting: some experimental results, in H. Jungermann and G. de Zeeuw, Eds., Decision Making and Change in Human Affairs. D. Reidel, Dordrecht, Holland.

Murphy, A.H., and R.L. Winkler (1977d). Probabilistic tornado forecasts: some experimental results. Preprints of Tenth Conference on Severe Local Storms (Omaha). American Meteorological Society, Boston, Massachusetts.

Peterson, C.R., K.J. Snapper, and A.H. Murphy (1972). Credible interval temperature forecasts. Bull. Am. Meteorol. Soc. 53: 966-970.

Sadowski, A.F., and G.F. Cobb (1974). National Weather Service April 1972 to March 1973 public forecast verification summary. Silver Spring, Maryland, NOAA, National Weather Service, Tech. Memo. FCST-21.

Sanders, F. (1973). Skill in forecasting daily temperature and precipitation: some experimental results. Bull. Am. Meteorol. Soc. 54: 1171-1179.

Slovic, P., B. Fischhoff, and S. Lichtenstein (1977). Behavioral decision theory. Ann. Rev. Psychol. 28: 1-39.

Winkler, R.L., and A.H. Murphy (1976). Point and area precipitation probability forecasts: some experimental results. Mon. Weather Rev. 104: 86-95.

Winkler, R.L., and A.H. Murphy (1977). The use of probabilities in forecasts of maximum and minimum temperatures, unpublished paper. Indiana University, Bloomington, Indiana, and National Center for Atmospheric Research, Boulder, Colorado.

16 On-Line Forecasting for a Regulated River using Weather Data from Radar

J.A. Cole

This paper describes the design and operation of an on-line forecasting system on the River Dee, which rises in North Wales and enters the Irish Sea through a broad estuary near Chester (Fig. 16.1). The construction and use of reservoirs in the upper Dee to augment river flows downstream were originally described by Crann (1968) and Blezard et al. (1970).

The Dee reservoir system has been the subject of a general study from the control rule aspect; the research was supported by the former Water Resources Board. Much of this work was pursued at the University of Lancaster and has been reported by Wilkinson (1972), Jamieson and Wilkinson (1972) and Jamieson et al. (1976).

The water authority concerned (now the Welsh National Water Development Authority) has had to work to fulfill complex objectives, which are quoted here from Lambert and Cameron (1975):

Llyn Tegid (Bala Lake): releases controlled by sluice gate.

- Short-term retention of flood runoff, for subsequent controlled releases.
- Intensive amenity use, with higher than minimum lake levels in the period April 1 to September 30.
- Reservoir balancing from which closely controlled adjustments of day-to-day releases are made using telemetry data sent to the control center at Bala Area Office.

Llyn Celyn: up to 12 m^3/sec controlled release, spillage over a bell-mouth overflow weir.

- Principal reservoir utilization is for low-flow regulation of the Dee, with storage reserved to be released for fishery, water quality, or emergency purposes. Good natural refill characteristics because of reservoir size in relation to catchment area and rainfall. The water released passes into Llyn Tegid.
- Short- to medium-term retention of flood runoff, with a variable upper retention level depending on time of year; the flood runoff is released to the Dee via Bala Lake as soon as practicable after heavy rainfall ceases.

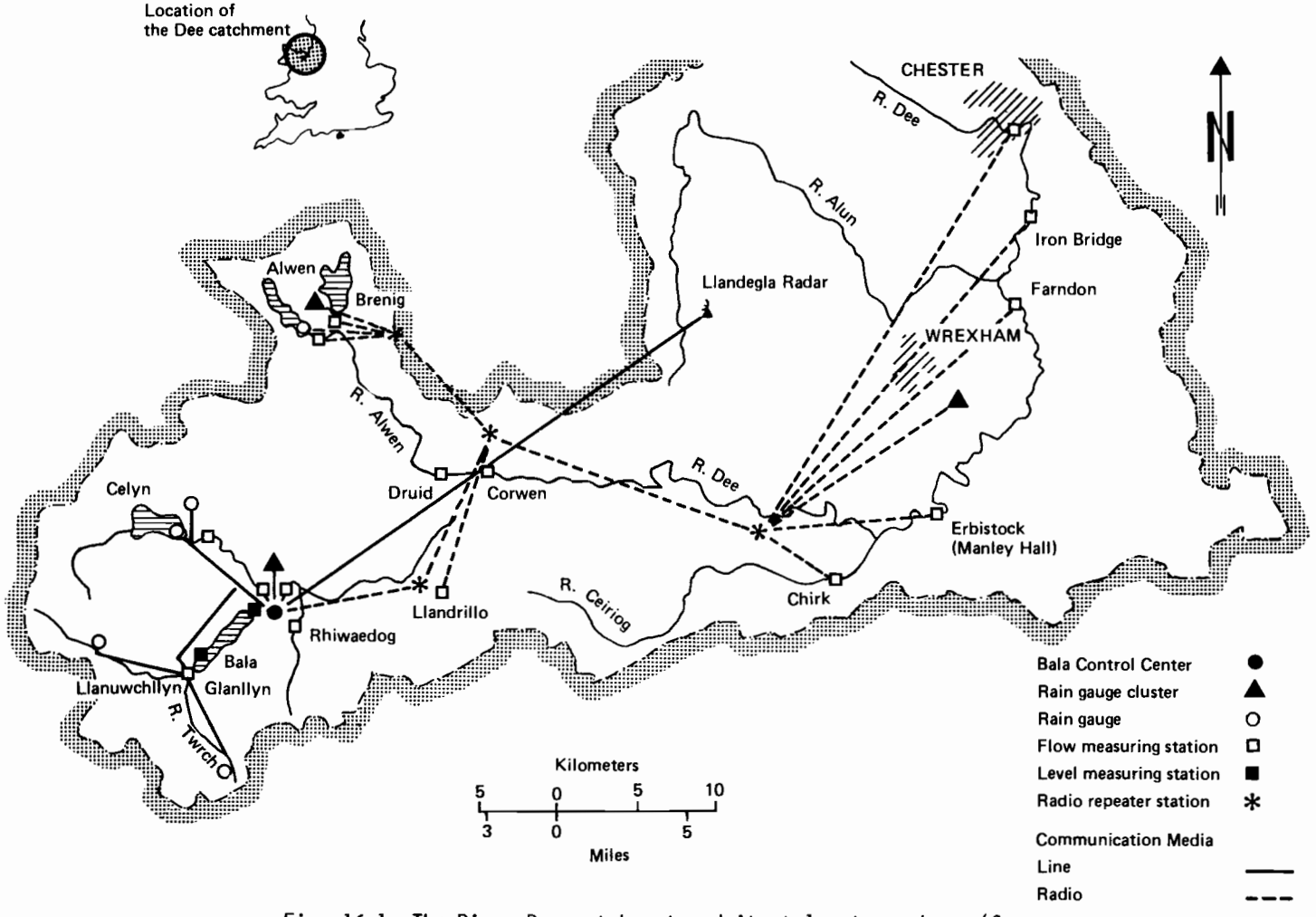


Fig. 16.1 The River Dee catchment and its telemetry scheme (from Water Research Centre, Medmenham, England, and Rowse and Roberts 1975).

- Generation of hydroelectric power for the national grid. This use is not given priority over either flood control or low-flow regulation; it is to be regarded as a useful byproduct when available.

Llyn Brenig: controlled release; spillage by overflow weir.

- Utilization of the reserve storage for assisting in low-flow regulation of the Dee, but only in very dry years; some storage of flushwater reserved for fishery, water quality, or emergency purposes. Poor natural refill characteristics are due to immense reservoir size in relation to catchment area and rainfall.

- Recreational use (boating, fishing, etc.).

- No significant flood regulation function.

The manner in which these objectives have been met has, of course, varied with the development of water storage in the Dee basin, as tabulated below, after Cole et al. (1975).

Table 16.1. Reservoirs on the Dee catchment

Reservoir	Date commencing operation	Volume ($10^6 \times m^3$)	Catchment area (km^2)	Principal uses	Secondary uses
Alwen	c.1920	14.5	25.5	Water supply (direct)	
Llyn Tegid (Bala Lake)	1955	6.4	266.8	Flood control; river regulation	Recreation
Llyn Celyn	1964	80.8	60.7	River regulation	Flood control; hydroelectric power
Llyn Brenig (1st stage)	c.1980	60.0	20.2	River regulation	

TELEMETRY SCHEME INCORPORATING WEATHER DATA FROM RADAR

For details of the design of the telemetry scheme in the Dee, see Rowse and Roberts (1975). As seen in Fig. 16.1, the hydrometric stations connect, by landline or UHF radio link, to a control center at Bala where there is a PDP 11/40 computer to handle data. The system is an extension of the one operated by the Water Authority since 1970 using a PDP 8 computer.

Interrogation of the rain-gauge and level measuring stations is performed automatically every half hour and the data are displayed on a mimic panel, a chart recorder, or are printed out by teletype.

The inclusion of weather radar data is one of the most significant recent developments in the telemetry scheme: the Llandegla radar has been in use for research

purposes since 1973 as part of a study on the accuracy of S and C band radars for the quantitative measurement of rainfall intensity; see Harrold and Austin (1974), Water Resources Board (1973), and Collier (1975).

The radar's own processor converts the areal pattern of microwave reflectivity into a rainfall intensity map expressed as a half-hourly rainfall on a 2 km x 2 km grid. Corrections are made for permanent echo of the signal because of the terrain, as well as for attenuation effects in the rainfall field itself. The measurements are best made with on-line calibration using telemetering rain gauges. The "rain gauge clusters" in Fig. 16.1 fulfill this purpose; their measurements are relayed on-line to Llandegla.

The rainfall data on the 2 km x 2 km grid are transmitted directly to Bala for display on a color television and are also delivered as subcatchment rainfalls for use in the hydrological model. The interactive software for transmitting the radar's rainfall data has been rather troublesome, partly because transmissions from Llandegla had to be given priority and this caused interruptions to the normal station calling sequence.

The televised rainfall maps are general guides to current weather, which are converted into a digital version from which subcatchment rainfalls are derived every quarter hour and relayed separately to the Bala computer.

As Taylor and Browning (1974) have emphasized, the radar measurements have wider implications, namely:

- Identification and tracking of weather hazards.
- Measurements of precipitation in river catchments which provide timely input data into objective forecast models of river flow.
- Assistance in the general analysis of the current weather situation by identifying mesoscale patterns of organization which otherwise would escape detection.
- Enabling, as a direct consequence of the third implication, more accurate short-term forecasts of precipitation intensity and other related phenomena such as lightning risk and wind shifts.

THE FLOW FORECASTING MODEL FOR THE DEE

The flow forecasting model has been described by Lowing et al. (1975) and will be presented here in summary form from Cole et al. (1975). There are three aspects of the model:

- Rainfall forecasting; it has not been implemented quantitatively, but much reliance is placed on the radar television picture of rainfall.
- Transformation of measured and forecast rainfall to a runoff hydrograph on a subcatchment basis.
- Tributary and main channel routing of subcatchment flows down to the regulating point.

The World Meteorological Organization (1975) has compared a variety of models suitable for on-line flow forecasting. The study has shown that in many river control systems the channel routing aspect is of paramount importance and the Dee appears to be no exception in this respect. Thus, it is highly desirable to have

good quantitative rainfall forecasts and a relatively simple tributary catchment model.

Rainfall Forecasting

Rainfall forecasts are needed to achieve efficient use of tributary flows far downstream of the regulating reservoirs. Currently, the only detailed rainfall forecasts available come from an assessment of the isohyetal maps from radar, as telemetered to Bala from Llandegla. These maps, in conjunction with synoptic weather situations obtained from regional branches of the British Meteorological Office, provide some deterministic rainfall forecasting. Research on storm cell growth, decay, and movement is currently underway within the British Meteorological Office, and improved estimates are expected within the next year or two. (See Harrold 1975.)

Rainfall-Runoff Model

There is an embarrassingly large choice of conceptual models of catchment response, even given the need to work at the half-hour time increments used in the Dee basin.

For real-time control purposes, a simple model has been adopted which has been shown to work well on the Dee catchment (Lambert and Cameron 1975). In this model there is a single storage state whose contents S_T at time T determine the outflow rate q_o at time 0 (i.e., now), T being a measure of catchment lag. Thus,

$$S_T = k \ln q_o + c ;$$

k and c are parameters of the catchment. This gives rise to the following sequential expressions for runoff with or without antecedent rainfall, P_L:

$$q_{t+T} = q_t \Big/ \left\{ \exp\left(\frac{-P_L}{k}\right) + \left(1 - \exp\frac{-P_L}{k}\right) \cdot \frac{Tq_t}{P_L} \right\} \text{ for } P_L \neq 0$$

or

$$q_t + T = q_t \ (1 + q_t T/k) \text{ for } P_L = 0$$

Future storage states (and hence outflow) are predicted by a simple water balance; rainfall adds to storage and outflow depletes it. Apart from its simplicity, this model has the major attribute that the current value of outflow (telemetered) may be used to update the catchment storage so that forecasts are continually being brought into line with the most recent data.

The calibration of the subcatchment model has been described by McKerchar (1975) who employs the criterion of least squares fit of observed and calculated hourly flows over a whole month of winter data. The values arrived at for lag time T and storage time constant k are tabulated in Table 16.2.

McKerchar recognized that k was not a constant parameter for a given subcatchment, but varied according to season, presumably reflecting a soil moisture state. Since summer recessions of storm runoff are particularly rapid, it is considered useful to retain the flexibility of the Lambert storage/lag model and only adjust k, which was made a function of antecedent rainfall and seasonal evaporation rate.

Table 16.2. Subcatchment model parameters

Catchment	Winter (Nov-Apr) T (hrs)	Winter (Nov-Apr) k	Summer (May-Oct) T (hrs)	Summer (May-Oct) k or k_1	Summer (May-Oct) Threshold q	Summer (May-Oct) k_2
Ceiriog	1.5	20.4	1.0	60.4	0.08	17.9
Alwen	1.0	8.6	1.0	23.6	0.10	9.1
Celyn	0.5	6.9	0.5	24.2	0.10	6.8
Hirnant	0.5	11.7	0.0	15.9	-	-
Upper Dee	1.0	4.9	1.0	7.9	-	-

River Flow Routing

To forecast flows at a downstream point requires a model that will account for translation and attenuation effects as water moves down the river channel. For the River Dee it was proposed to use Cunge's (1969) modification of the Muskingum methods. This assumes channel storage S is a linear function of inflow, Q_i, and outflow, Q_o, such that:

$$S = K\left\{xQ_i + (1 - x)\right\}Q_o$$

where K and x are fitted parameters of the reach. However, this model implies a wave celerity independent of discharge and in the upper Dee such an assumption proves untenable. The variable parameter diffusion (VPD) method developed by Price (1973) was used instead.

Price (1973) calculated the nonlinear variation of wave celerity $\bar{c}$ and attenuation $L\bar{a}$ as a function of discharge Q for the two main segments of the upper Dee, where $\bar{c}$ and $\bar{a}$ are discharge-dependent terms in the VPD equation for the transient flow propagation:

$$\frac{\partial q}{\partial t} + \bar{c}(Q)\frac{\partial Q}{\partial x} = Q\frac{\partial}{\partial x}\left[L\bar{a}(Q)\frac{\partial Q}{\partial x}\right]q + \text{negligible higher terms}$$

where

q = the lateral inflow at head of river reach,

t = time,

x = the downstream distance,

L = the length of the river reach, and

Q = discharge.

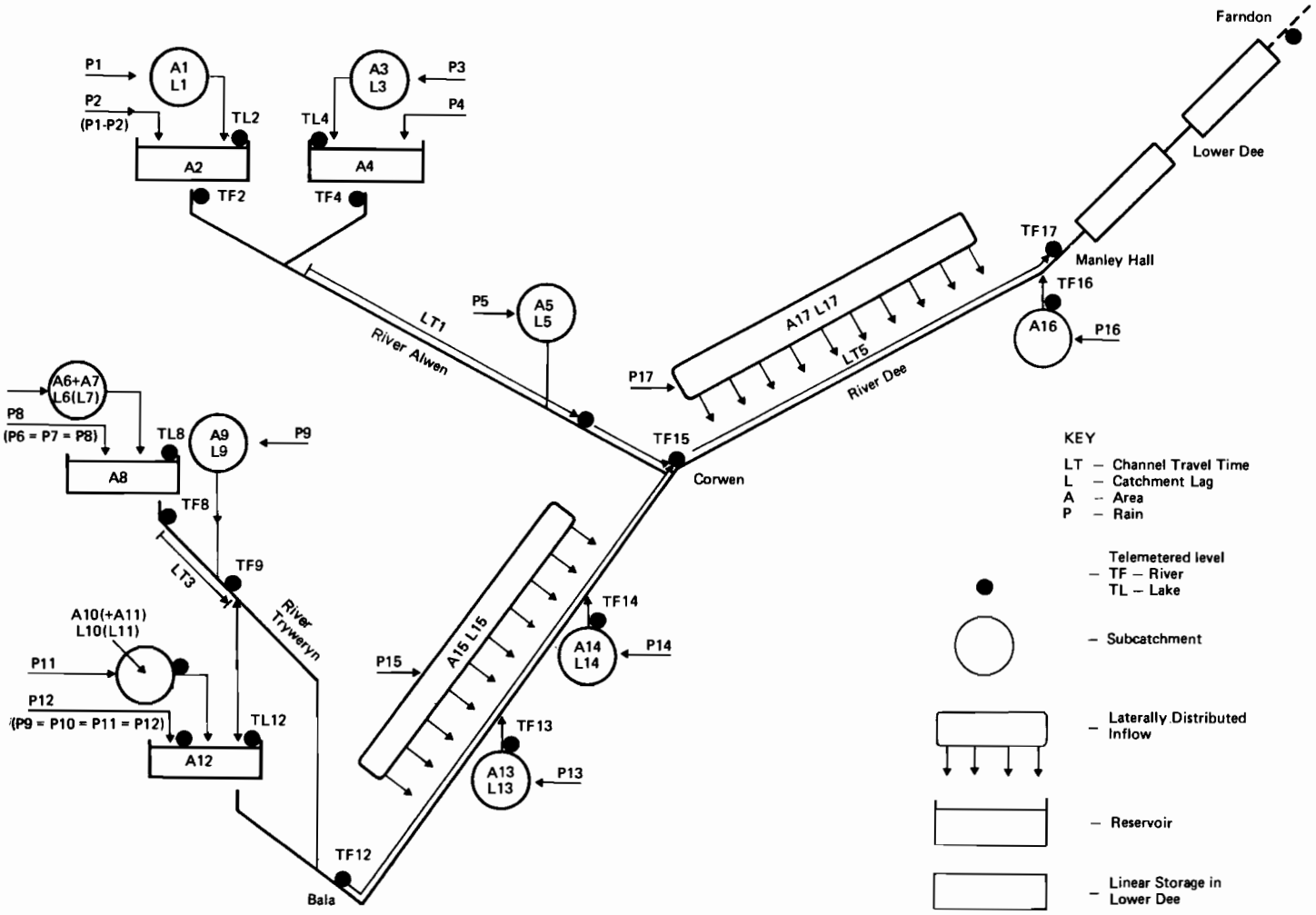


Fig. 16.2 Schematic drawing of the Dee system (from Water Research Centre, Medmenham, England, and Lowing et al. 1975).

Combination of Lateral and Subcatchment Flows

Figure 16.2 shows the nine subcatchments, shown as circles, feeding the upper River Dee and its major tributary the River Alwen. The two large areas of lateral inflow are each regarded as lumped subcatchments with regard to their rainfall runoff behavior, but their runoff is evenly distributed down the channel reaches. Most tributary flows are ungauged (or if gauged, not telemetered) which means that their flow and catchment storage forecasts are not amenable to continual updating. Thus, the updating procedure is applied for the data as a whole, either in proportion to the difference in forecast and gauged flows on the Hirnant catchment (for upper tributaries) or on the Ceiriog (for lower ones). The catchments are shown in Fig. 16.3.

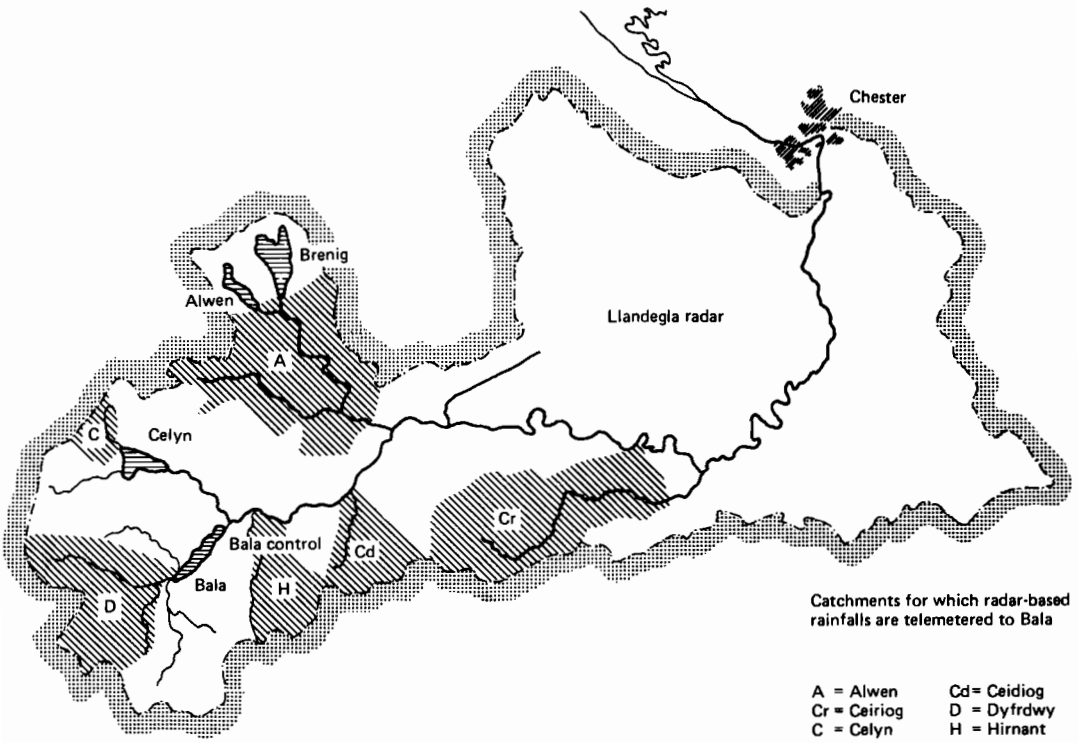


Fig. 16.3 Dee subcatchments (from Water Research Centre, Medmenham, England).

It will be seen from Fig. 16.3 that there are considerable areas for which no rainfall-runoff model has been established: in such cases there is a choice between adopting model parameters on some topographical basis or simply accepting interpolation of flows estimated in adjacent subcatchments.

Routing the Manley Hall Forecast Hydrograph to Farndon

Farndon is the effective control point for withdrawals from the river at Eccleston, near Chester. The Manley Hall - Farndon reach is modeled as two quasi-linear channels and two linear reservoirs in series. The main reason for using two reservoirs is to allow forecast spillage onto the flood plain between the two reservoirs at Worthenbury Brook confluence (WBC). Thus, between Manley Hall and

WBC there is assumed to be a linear channel with delay T hours and a linear reservoir with storage constant k hours; between WBC and Farndon an identical combination is assumed.

Values of T and k have evolved by trial and error. This has been more acceptable than optimization using "sum of squares" error minimization. The aim in finding suitable T and k values was to match observed attenuation and travel times while making allowance for the varying and unknown inflow below Manley Hall. The storage constant k was set for each reservoir at 3 hours. T is allowed to vary for each channel by the expression

$$T = \frac{140}{21 + FLOW}$$

where FLOW is the average of all Manley Hall flows from the flow measurement 5 hours before the event up to those flows forecast 24 hours ahead. This explains what is meant by quasi-linear channel. The lag and route procedure is linear. The lag does not change within the routing regardless of the variation in flow, but the lag may change before the next routing in an attempt to match the observed variation in pure translation time with discharge.

An estimate is needed of the future inflow from the main tributary catchments entering between Manley Hall and Eccleston (totaling 655 km^2). The assumption is made that this tributary inflow will recede at 3 percent/day from its current value; the current value is assumed to be the difference between the Manley Hall flow routed to Farndon and the gauged Eccleston flow.

This sequence of differences, which would seem to represent a lower Dee inflow hydrograph, usually requires smoothing and is accomplished by running a 12-hour moving average which is currently updated and is considered to apply to a time 6 hours before. The exponential decay is projected from this earlier time to produce the forecast hydrograph of lower Dee inflow up to 48 hours ahead. When added to the routed Manley Hall hydrograph, the 2-day forecast for Eccleston is thus obtained.

The lower Dee model is not very helpful for floods since the inflow predictions ignore any rainfall that may be observed or forecast in the lower Dee catchment. It is thought, however, that the assumptions, although crude, are satisfactory in the light of operational requirements. Despite this lack of emphasis on accurate high flow modeling, there is a requirement for flood warning in the lower Dee and this means the ability to forecast when and for how long overtopping of banks is likely to take place.

HYDROLOGICAL FORECASTS AND RIVER REGULATION

It is useful to depart momentarily from the Dee system and to consider, in principle, how a flow forecast is used for river regulation. For the system shown in Fig. 16.4, the river is regulated by the release of water at point Z to meet a prescribed flow at point X. The river is divided by broken lines representing isochrons of travel time; lag differences within each segment are ignored.

Our objective is to achieve the prescribed flow at X without wasteful surplus or harmful deficit. This calls for good forecasting of flows down the river, taking all tributary flows into account, T hours ahead. The cumulative amounts of absolute surplus or deficit will be an error which one should seek to minimize.

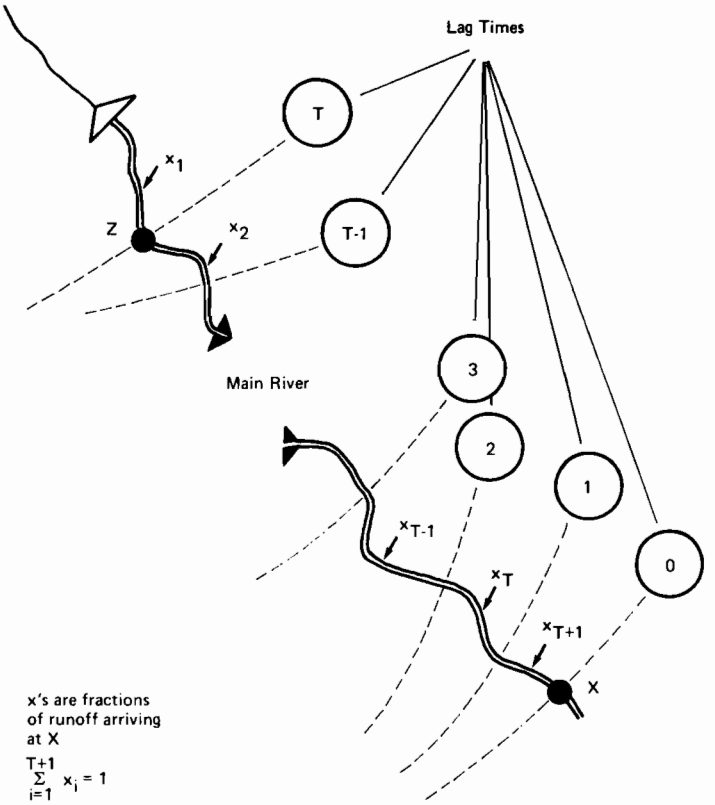


Fig. 16.4 Sketch of river receiving regulated releases at Z with time segments of travel to control point X (from Water Research Centre, Medmenham, England).

How the Flow at X is Predicted

Controlled releases at Z at time t	= R(t)
Unregulated releases at Z (e.g., spillage) at time t	= U(t)
Local inflow to segment j at time t	= $Q_j(t)$

In the Dee project, the $Q_j(t)$ are primarily derived from a rainfall storage-runoff model, but may be derived from tributary flow gaugings.

If for the moment we neglect hydrograph diffusion and attenuation, so that segment contributions simply add as plug flows (separate tributary contributions) with appropriate lag times, we compute the contributions to flow at X from the tributaries as shown in Table 16.3, where the circumflex (^) denotes an estimate.

The table below is purely qualitative in that it gives each Q "based on" a Q derived earlier. For a well-studied river system, such as the Dee, recession behavior of tributary catchments will be known and a realistic Q will be obtained under conditions of low rainfall, such that catchment storages are not replenished.

Table 16.3. Contributions to flow at X from the Dee tributaries

Segment	Contribution to Flow at X at Time $(t_o + T)$	Contribution to Variance of Total
1	$R(t_o) + U(t_o)$	Negligible
2	$\hat{Q}_2(t_o + 1)$ based on $Q_2(t_o)$	x_2 times a small error
j	$\hat{Q}_j(t_o + j)$ based on $Q_j(t_o)$	x_j times an intermediate error
T	$\hat{Q}_T(t_o + T)$ based on $Q_T(t_o)$	x_T times a large error

It is recognized that the contributions from each segment do not go by pure translation as plug flow to X, but diffuse into one another. This effect is of minor importance for the lowest segments (or reaches) which are most prone to the errors of rainfall forecasting. The effect becomes of prime importance when examining controlled releases from Z, which will appear as sketched in Fig. 16.5

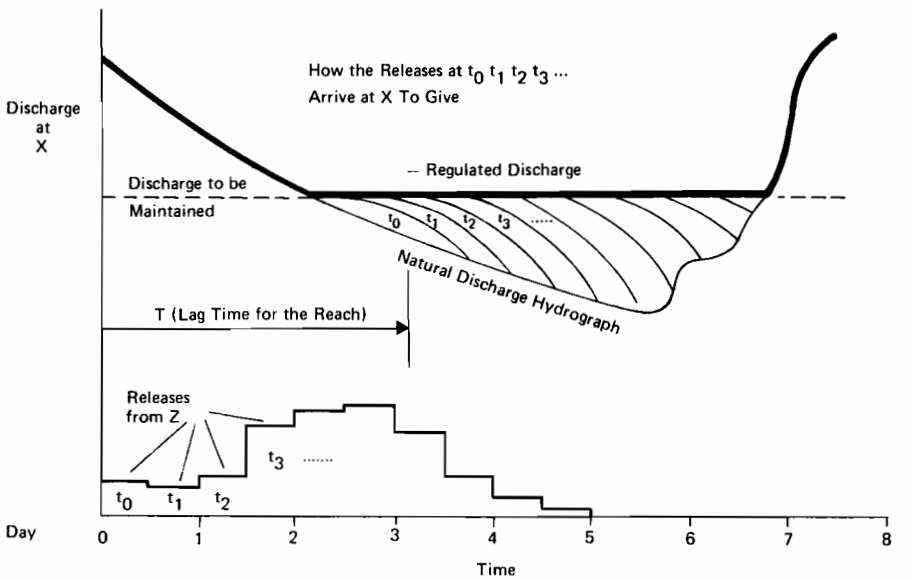


Fig. 16.5 How successive releases of water at Z arrive at X (from Water Research Centre, Medmenham, England).

on arrival at X. A further aspect is the flow-dependence of T itself; this cannot be ignored (as it varies at least twofold in the Dee between high and low flows) and is implicitly considered.

Error Terms

It is to be expected that errors in predicted flow will increase with the time required for the transfer of water from Z to X. It is quite likely that the error spread will increase exponentially with T.

With an unbiased estimator, the spread will be symmetrical about the time axis, but bias is likely to appear from a great many causes. The pure recession assumption may show a positive bias (observed flows greater than predicted flows) because of continuation of rainfall beyond time t_o when none was assumed within the range of the forecast.

In order to arrive at a satisfactory predictive model, we should test various assumptions concerning the rainfall prediction adopted with the forecast range t_o up to $(t_o + T)$. Criteria of success will be the absence of bias and any reduction in variance achievable. The total flow at X is the prime basis of such assessments, but there will also be opportunities to examine separate segments whose flows are gauged.

Controlling the River Dee

Referring again to Fig. 16.1, the Bala sluices at the outfall of Llyn Tegid may be considered as the regulation release point equivalent to Z in Fig. 16.4. Llyn Celyn works as the major backup storage, according to a seasonal control rule supplying enough for the Llyn Tegid releases in dry weather. Llyn Celyn releases are withheld in wet weather conditions when Llyn Tegid discharge can meet the regulation flows out of its own catchment and storage. Flood alleviation measures are possible by advance or delayed release from Llyn Tegid; see Wilkinson (1972). For a description of flood alleviation strategies in more complex configurations see Schultz and Plate (1976).

FORECAST ACCURACY

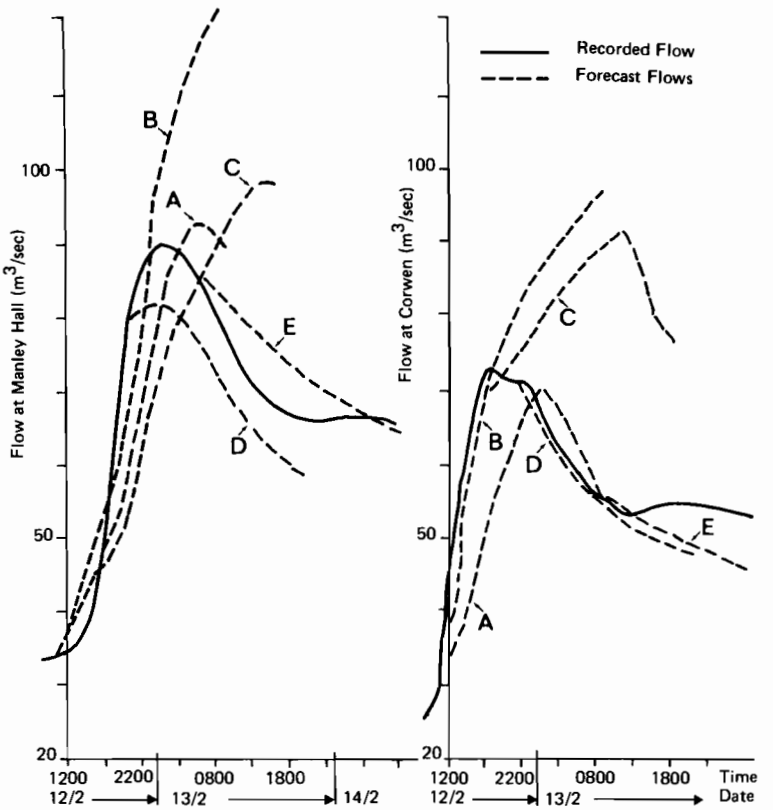
We have noted that the forecast accuracy at the control point is very much dependent on tributary flows whose influence has to be judged to a large extent from rainfall forecasts. An illustration of this fact is given in Fig. 16.6 which contrasts hydrographs at Corwen and Manley Halls on February 12, 1976, with a sequence of forecasts based on different rainfall predictions provided at the time by the regional Meteorological Office Forecasting Center; though expressed quantitatively, they were of a "broad brush" nature.

The sensitivity of flow forecasts to recent rain data is greatest in the headwaters. This will be seen from Figs. 16.8, 16.9, and 16.10, which relate to three points down the River Dee, for rainfall on September 11 and 12, 1976 (Fig. 16.7). The River Alwen is gauged close to its confluence with the Dee, just above Corwen. The three forecasting points are:

Corwen	22 km	all downstream of the Bala sluices on the main river.
Manley Hall	65 km	
Farndon	102 km	

The river response at Corwen is predicted:

- From the releases anticipated at Bala and routed down the Bala-Corwen reach.
- By adding in the Alwen flows at the confluence immediately upstream of Corwen.
- By addition and routing of lateral inflows.



Identifier	Time of Forecast	Assumption of Forecast
A	1000	1st Rain Forecast—Release from Bala at 20 m^3/sec
B	1100	2nd Rain Forecast—Release from Bala Rising to 30 m^3/sec after 3 hr
C	1700	2nd Rain Forecast—Release from Bala to Continue at 31 m^3/sec
D	2100	No Further Rain—Bala Release to Continue at 31 m^3/sec for 24 hr
E	0900	No Further Rain—Bala Release to Continue at 31 m^3/sec

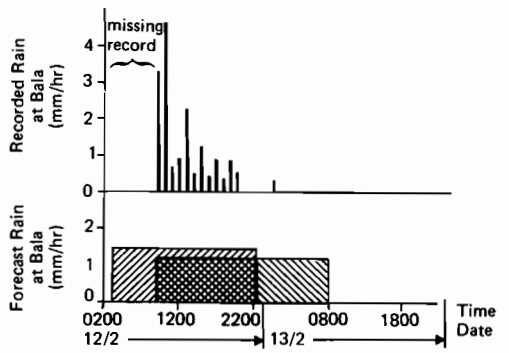


Fig. 16.6 Comparison of forecast hydrographs on basis of different rain forecasts (from Water Research Centre, Medmenham, England).

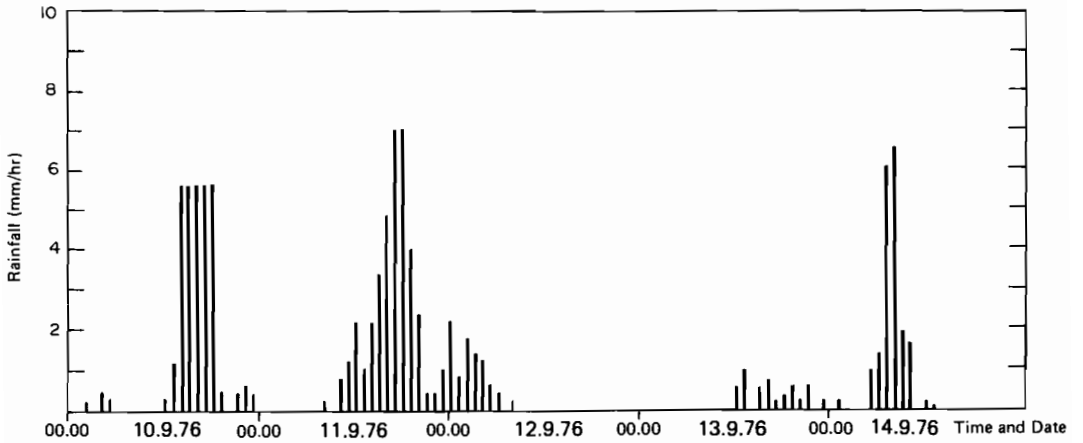


Fig. 16.7 Specimen rainfall at Celyn for September 10 to 14, 1976 (from the Dee Weather Radar and Real-Time Hydrological Forecasting Project Report 1977).

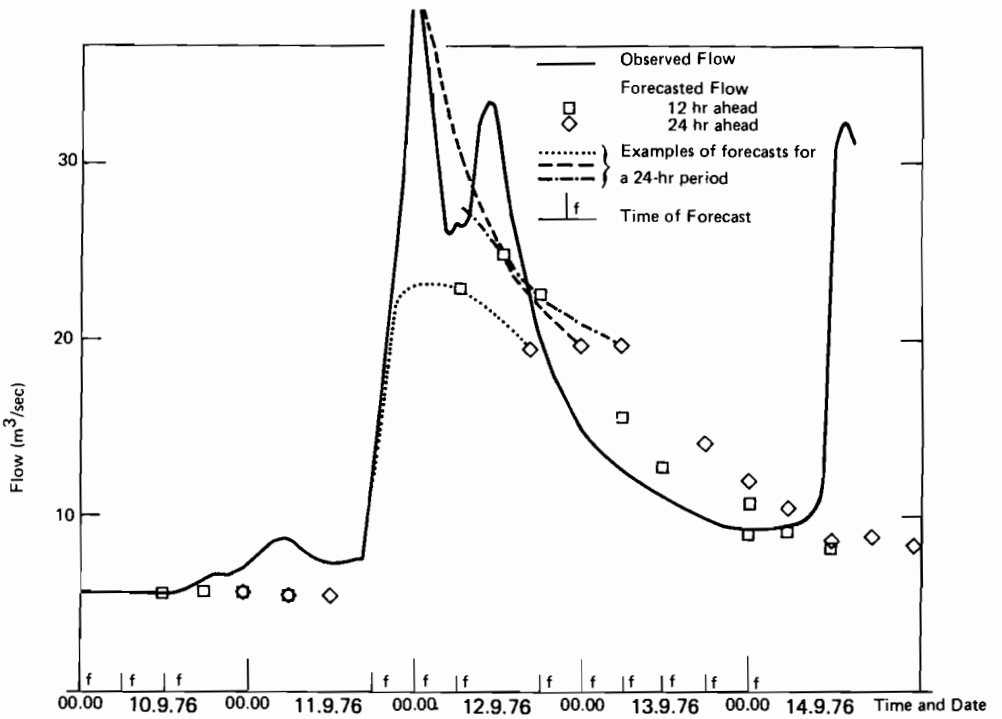


Fig. 16.8 Specimen of hydrograph and forecasts of the Dee at Corwen for September 10 to 13, 1976 (from the Dee Weather Radar and Real-Time Hydrological Forecasting Project Report 1977).

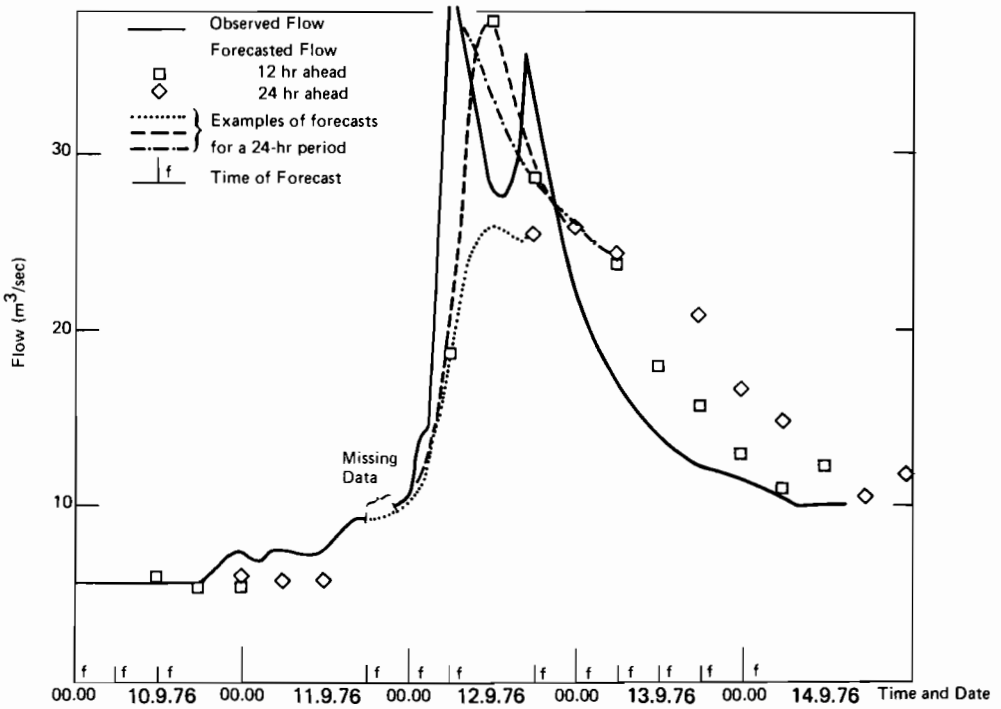


Fig. 16.9 Specimen hydrograph and forecasts of the Dee at Manley Hall for September 10 to 13, 1976 (from the Dee Weather Radar and Real-Time Hydrological Forecasting Project Report 1977).

Of these components, the first and the third contribute only slight errors to the forecast, but the second is heavily dependent on rainfall expectations, as the contribution from Alwen has a lag time of approximately 1 hour. The 24-hour flow forecast for 1800 hours on September 11, 1976, in Fig. 16.8 exhibits this fact; the runoff peak was underestimated by some 40 percent and the forecast assumed no future rainfall.

Figure 16.9 continues this story. The 12-hour forecast made at 2400 hours on September 11, 1976, for Manley Hall, gives a markedly better peak prediction than did the comparable 12-hour ahead forecast made at Corwen at 1800 hours on September 11.

With progressive movement of the flow peak downstream, the forecasting ability on the Dee improves, as Fig. 16.10 shows. Even though all the flow forecasts assumed no future rainfall, much of the forecast at Farndon is based on flow routing of water already on the ground or in the river's main channel. Thus, the assumption does not significantly add to the forecasting error.

From the experience with the Dee, the conflicting factors governing forecast error may be summarized thus: with progression downstream, factors improving the flow forecast are the gathering-in of more flow to the channel, the attenuation properties of the channel, which reduce hydrograph peaks, and long lag times of tributaries;

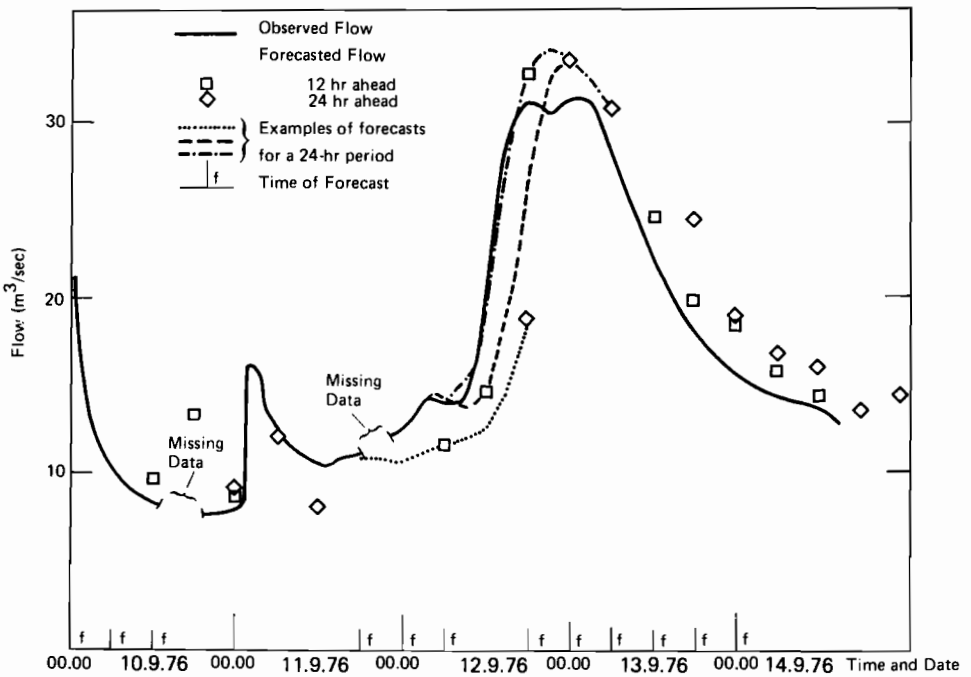


Fig. 16.10 Specimen hydrograph and forecasts of the Dee at Farndon for September 10 to 13, 1976 (from the Dee Weather Radar and Real-Time Hydrological Forecasting Project Report 1977).

factors worsening the flow forecast are the confluence of a major tributary (ungauged ones cause a more severe error than those which are gauged), the lack of rain forecasts for such tributaries and short lag times of tributaries.

The forecast accuracy in a particular case depends strongly on the network of rivers concerned: the confluence of a major tributary will probably cause a sawtooth function to be superimposed on any diminution of errors caused by the gathering-in of flow and attenuation properties of the channel as flows move downstream.

ACKNOWLEDGEMENTS

This paper is based on recent experience on the Dee Research Program, responsibility for which was transferred in 1974 to the Water Research Center (WRC) and the Central Water Planning Unit (CWPU) which jointly fund the research investigations. A large portion of the telemetry hardware costs except for radar, was borne by the Welsh National Water Development Authority. The Meteorological Office, the Plessey Company and the Royal Signals and Radar Establishment have been primarily responsible for the radar developments. The Institute of Hydrology and the Hydraulics Research Station, Wallingford, jointly developed the real-time flow forecasting procedures.

The views in this paper were formulated prior to completion of the Dee Research Program, and are those of the author alone - they are not to be construed as those of the WRC, CWPU or their collaborating organizations.

REFERENCES

Blezard, N., H.H. Crann, D.J. Iremonger, and E. Jackson (1970). Conservation of the environment by river regulation. Association of River Authorities, Chester Conference.

Cole, J.A., A.I. McKerchar, and R.J. Moore (1975). An on-line flow forecasting system, incorporating radar measurements of rainfall, as used to assist the short-term regulation of the River Dee in North Wales. IAHS-AISH Publ. No. 115, pp. 57-66.

Collier, C.G. (1975). Rainfall measurement by radar. Water Research Center and Royal Radar Establishment, Symposium paper.

Crann, H.H. (1968). The design and construction of Llyn Celyn. J. Inst. Water Eng. 22: 13-71.

Cunge, J.A. (1969). On the subject of a flood propagation method. J. Hydraul. Res. IAHR: 7(2): 205-230.

Harrold, T.W. (1975). Rainfall forecast in the United Kingdom using radar data. Water Research Center and Royal Radar Establishment, Symposium paper.

Harrold, T.W., and P.M. Austin (1974). The structure of precipitation systems - A review. J. Rech. Atmos. pp. 41-47.

Jamieson, D.G. (1972). Operating a multi-purpose reservoir system for water supply and flood alleviation. Water Resour. Res. 8(4): 899-903.

Jamieson, D.G., and J.C. Wilkinson (1972). A short-term control strategy for a multi-purpose river system. Water Resour. Res. 8(4): 911-920.

Jamieson, D.G., D.K. Smith, and J.C. Wilkinson (1976). Evaluation of short-term operational policies for a multi-purpose reservoir system. J. Hydrol. 28: 191-213.

Lambert, A.O., and R.J. Cameron (1975). Control rules for long and short-term objectives. Water Research Center and Royal Radar Establishment, Symposium paper.

Lowing, M.J., R.K. Price, and R.A. Harvey (1975). Real-time conversion of rainfall to runoff for flow forecasting in the River Dee. Water Research Center and Royal Radar Establishment, Symposium paper.

McKerchar, A.I. (1975). Subcatchment modeling for Dee River forecasting. International Report No. 30, Institute of Hydrology, Wallingford, Oxon.

Micholass, C.A., and T.W. Harrold (1975). The distribution of rainfall over subcatchments of the River Dee as a function of synoptic type. Meteoral. Mag. 104: 208-217.

Price, R.K. (1973). Flood routing method for British rivers. Proc. Inst. Civ. Eng. Part 2 55: 913-930.

Rowse, A.A., and B.H. Roberts (1975). Design of the Dee telemetry system, with computer acquisition of data. Water Research Center and Royal Radar Establishment, Symposium paper.

Schultz, G.A., and E.J. Plate (1976). Developing optimal operating rules for flood protection reservoirs. J. Hydrol. 28(2/4): 245-254.

Taylor, B.C., and K.A. Browning (1974). Towards an automated weather radar network. Weather 29(6): 202-216.

Water Resources Board, Operations Systems Group (1973). Dee weather radar project: The use of a radar network for the measurement and quantitative forecasting of precipitation.

Wilkinson, J.C. (1972). A long-term control strategy for a multipurpose reservoir. Water Resour. Res. 8(4): 904-910.

World Meteorological Organization (1975). Intercomparison of conceptual models used in operational hydrological forecasting. Geneva, WMO Operational Hydrology Report No. 7.

17 Real-Time Hydrological Forecasting Systems in Poland

Eryk Bobinski
Teofil Piwecki
Janusz Zelanzinski

River hydrograph forecasts for flood protection and reservoir operation for several rivers in the mountainous region of Poland are urgently needed. The catchment areas, upstream of reservoirs, vary from several hundred to a few thousand square kilometers; the surface runoff travel time is from 6 to 24 hours. River forecast information based on conventional methods like API, regional runoff ratios, and so on was considered inadequate. For this reason, a new system using a computer simulation model was implemented in 1973 for the Sola River and in 1974 for the Dunajec River.

A related problem is the forecasting of the river stage hydrograph for specific reaches of the two main rivers in Poland, the Vistula and the Oder. The forecast is for inland river navigation; a system providing river stage hydrograph forecasts for two cross sections of the lower Vistula that have unsteady flow conditions is being implemented.

FLOW FORECAST SYSTEM FOR THE SOLA RIVER SYSTEM

The Sola River is a mountain tributary of the Vistula with a catchment area at the gauging station Zywiec of 780 km^2. The altitude of the catchment is between 340 and 1,500 m above sea level. About one half of the catchment area is forested. The lag time after a heavy rain is about 6 hours and the maximum rainfall is about 100 mm in 24 hours.

The observational network incorporated into the system consists of 11 recording rain gauges, 1 climatological post, 1 actinometric post, and the river stage recording gauge at the Zywiec cross section.

Data from the posts are reported to the collecting station Zywiec by radio and from Zywiec the input data message (named HYDRA) is transmitted by teletype to the computer at the Institute for Meteorology and Water Management in Warsaw.

The HYDRA message, after reception in Warsaw, is checked by hand - the checking includes a formal control of the message headline, the group and station numbers, and the total message layout. Then, to the real-time data, are added the values of forecasted meteorological variables: precipitation, wind speed, air humidity deficit, and total radiation for the coming 48 hours, divided into two intervals of 24 hours each. This augmented input message, after completion, is punched on cards

and fed into an ODRA 1305 computer (a machine compatible with the ICL 1900 computer). The flow forecast is computed using a model (MONIKA) based on the concepts of the SSARR catchment model with several modifications introduced. A description of the MONIKA model was presented at the WMO-UNESCO Symposium in Bratislava (Bobinski et al. 1975). The system is operated for a summer period from May 1 until October 31 when the precipitation is mostly rain and the rain recording gauges can work without interruption caused by snow.

Computation

A flow chart of the forecasting model is shown in Fig. 17.1. The time step for the input data is 3 hours and for the computation and output 1 hour. However, the output message (called HYFOR) is transmitted to the collecting station by teletype and gives the forecasted flow values in 3-hour intervals.

Usually when the model is operated under real-time conditions, part of the input data (usually precipitation) is missing or contains errors different in type and magnitude. For this reason, the pre-processing of the input data, i.e., error checking, estimation of missing data, estimation of the actual evapotranspiration, computation of mean areal precipitation etc., is required and constitutes the large portion of the program. Quite often a few rain recording gauges are out of order. In such cases estimation of missing data uses the 12-hour and 24-hour total precipitation from the standard raingauges and the rainfall time distribution pattern available from the working recording gauges.

Evapotranspiration estimation. Evapotranspiration estimation is done for both the last 24-hour period and the forecasted 48-hours using observed values and forecasted data from the HYDRA input message. Potential evapotranspiration is estimated by use of the St. Bac (Poland) formula:

$$E_p = 0.001\ (100 \cdot d \cdot v^{0.5} + 4T) \qquad (17.1)$$

where

E_p = potential evapotranspiration (mm/day);

d = air humidity deficit (mbar);

v = wind speed (m/s);

T = total radiation (cal/cm^2).

The 24-hour value of E_p is divided into 3-hour intervals after a statistical analysis of data for the summer season. These computed 3-hour values are then reduced, taking into account actual soil moisture and precipitation conditions, to obtain actual evapotranspiration data, which are deducted from the interception or soil moisture storage.

Interception storage. It is assumed that rainfall excess begins only after filling the total interception storage, which is set equal to 3.5 mm for this catchment. When the precipitation is less than interception storage, then intercepted water can only evaporate.

Updating of the parameters. Two model parameters, the soil moisture index WN and the base flow infiltration index B, represent the current conditions of the model. Updating is done in relation to these two parameters. Beginning with the initial values of WN and B given in the program, the values are then changed to obtain a

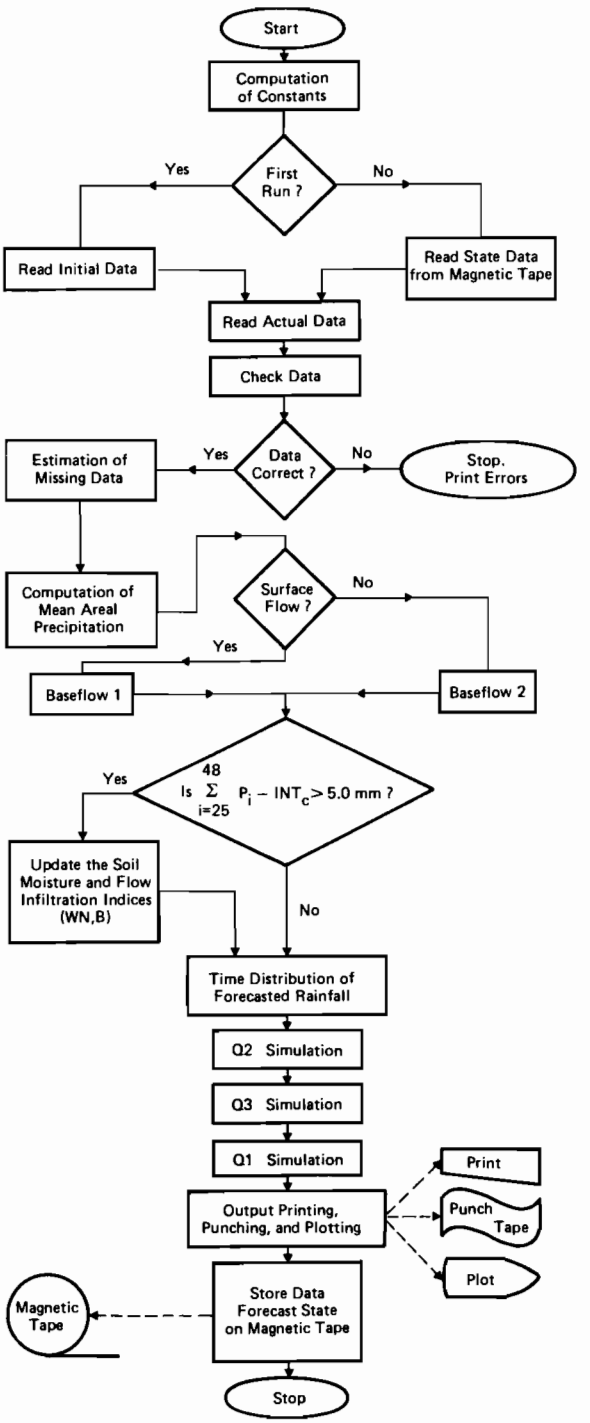


Fig. 17.1 A flow chart of the forecasting model for the Sola and Dunajec Rivers. See the text for explanations of $\underline{P_i}$ and INT_c.

minimum of the criterion function

$$F = \sum_{i=1}^{n} \left(\frac{Q_o(i) - Q_s(i)}{Q_o(i)} \right)^2 \qquad (17.2)$$

where

$Q_o(i)$ = the flow at the gauging station recorded at hour i;

$Q_s(i)$ = the flow at the same hour simulated in the model for the recorded precipitation (see Fig. 17.2);

n = the number of hours for which the updating is done.

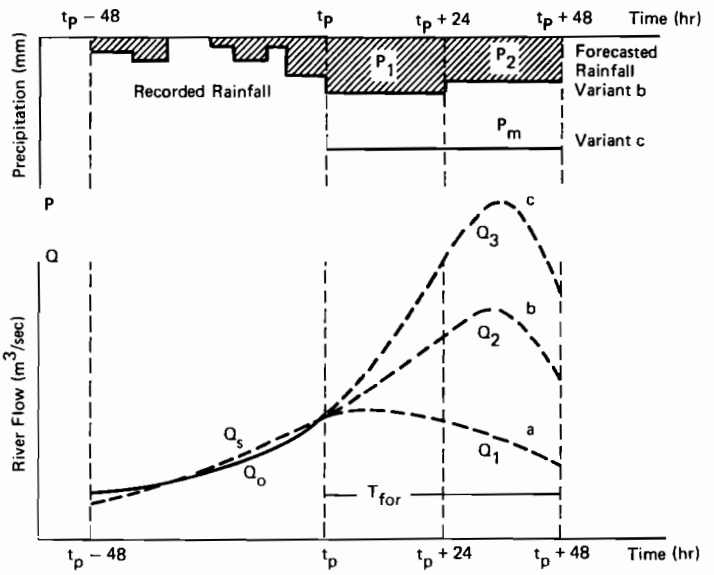


Fig. 17.2 The diagram of the forecasting model principles. t_p = the initial moment of the forecast period; t_p+48 = the end of the hydrograph forecast period; T_{for} = the forecast period; Q_s = simulated hydrograph; Q_o = recorded hydrograph; a, b, c, = three variations of the hydrograph forecast corresponding to three precipitation forecasts.

At the beginning of the forecasts, when the model is started for the first time, n = 3 or a multiple of 3. After two days of operation n is 48. The updated values of the parameters WN and B are stored in the memory and serve as initial values for the computation in the following cycle.

The updating procedure is executed only when the following condition is fulfilled:

$$\sum_{i=25}^{48} P_i - INT_c > 5.0 \text{ mm} \qquad (17.3)$$

where

$\sum_{i=25}^{48} P_i$ = the precipitation amount observed for the last 24 hours preceding the forecast time t_p (see Fig. 17.2);

INT_c = total interception storage.

If this condition is not fulfilled, then the WN and B parameters are not updated and the values obtained at the end of the last computation run serve as the initial values for the actual run.

Such an updating procedure does not exist in the original SSARR model. Instead, following Rockwood et al. (1972), the watershed initial conditions are adjusted in some other way to the observed discharges. If the computed discharge at a point in time does not match the observed discharge within a tolerance of error, an adjustment factor is applied to change the rain quantities, and the simulation is run again. Both approaches, either applied in the original SSARR or in the MONIKA model, may be questioned and the problem seems open to discussion.

Baseflow switching. When both surface and subsurface flow components are present, the baseflow component is computed using "Baseflow 1" parameters. When the baseflow dominates, then the parameters of "Baseflow 2" are applied. The difference is that the time of storage parameter T_{SB}, is equal to 120 hrs for Baseflow 1, and 1,200 hrs for Baseflow 2. The following conditions should be satisfied for Baseflow 2:

$$\sum_{i=25}^{48} P_i \leq 5 \text{ mm} \tag{17.4}$$

$$\Delta Q \leq 5 \text{ m}^3/\text{sec} \tag{17.5}$$

and

$$\frac{\Delta Q}{24} \leq 0.5 \text{ m}^3/\text{sec per hr} \tag{17.6}$$

where

$$\Delta Q = Q_{max} - Q_{48} \text{ or } Q_{48} - Q_{25}, \tag{17.7}$$

Q_{25} = the flow rate at the beginning of the last 24-hr period;

Q_{48} = the flow rate at the end of the period;

Q_{max} = the maximum flow during the last 24 hrs.

All the above conditions must be satisfied. If not, Baseflow 1 is used. Equations (17.4), (17.5), and (17.6) are, of course, a rough estimation of the surface and subsurface flow components. For this model, these criteria, together with the baseflow switching operation, give fairly accurate results.

The time distribution of the forecasted rainfall. The distribution of rainfall is done in 3-hr intervals using a distribution function depending on the precipitation pattern: (a) convective, (b) frontal, (c) uniform, or (d) random. The distribution patterns of types a and b were identified with a statistical analysis of the precipitation data of the river catchment in question. The distribution is chosen by the meteorologist-forecaster on duty, and appropriate numbers are input into the

HYDRA message.

The flow hydrograph forecast. Hydrograph forecasts are presented for variations depending on the forecasted precipitation (see Fig. 17.2). These variations are:

- The total precipitation over the period T_{for} is equal to zero; it gives forecast Q_1.
- The total precipitation within the T_{for} will amount to the forecasted values P_1 and P_2. The values P_1 and P_2 are given separately for the first and second 24-hr forecast period by a synoptic meteorologist on duty and result in the forecasted outflow Q_2.
- The maximum precipitation P_m of the 24-hr total of 100 mm will occur in the time T_{for} resulting in the forecasted outflow Q_3.

These assumptions are made because quantitative precipitation forecasting methods are not reliable enough, especially for small mountain catchments. With these assumptions, three forecasted hydrographs are obtained instead of one. Variation (a) determines the lower limit, (c) the upper limit and (b) gives an intermediate hydrograph related to the precipitation forecast. In flood conditions, when the simulation is repeated every 3 hrs, such an approach makes it possible to correct the forecast according to the actual development, and a choice can be made between Q_2 and Q_3. In normal conditions the choice is made between Q_1 and Q_2 by the forecast hydrologist on duty who is responsible for the forecast transmitted to the user.

The output HYFOR message is printed by the line printer, punched on a paper tape that is fed to the teletype, and transmitted directly to the data station Zywiec where the forecast hydrologist on duty chooses the proper flow forecast and transmits it to users.

All output and input data, together with the parameter values representing the actual state of the model, are stored on magnetic tape and used as data files. About once a month, the recorded and forecasted hydrographs and hyetographs are plotted to help estimate model performance. An example of the plot for the Dunajec river at Kowaniec, for July 1976, is presented in Figs. 17.3 and 17.4.

System Operation

The forecast system is operated under two conditions: normal flows and flood conditions. Under normal flows a forecast is computed once a day in the morning hours and under flood conditions the computing runs are repeated every 3 hrs, that is, after each observation cycle.

During normal operations, the full cycle of the system starting from the observation time until the release of the output forecast (HYFOR) takes about 3 hrs, of which the machine time for the analysis and the forecast computation (the operations shown in Fig. 17.1) takes only 3 minutes. Under flood conditions the cycle duration is about 2 hrs.

Figure 17.5 presents a diagram of the length of specific operations. It shows that after the reception of the HYDRA input message in the computer center, about 1 hr is needed to complete the meteorological forecast data. This delay results from the present arrangement of the meteorological forecast system. After some rearrangement, the time for the system operation will be about 2 hrs. In any case, there is a remarkable inconsistency between the machine time and time spent on all other operations, most of which are done by hand. One may wonder whether this is a "real-time" system; it is the best to be done with the present technological options available in the Polish hydro-meteorological service. In the

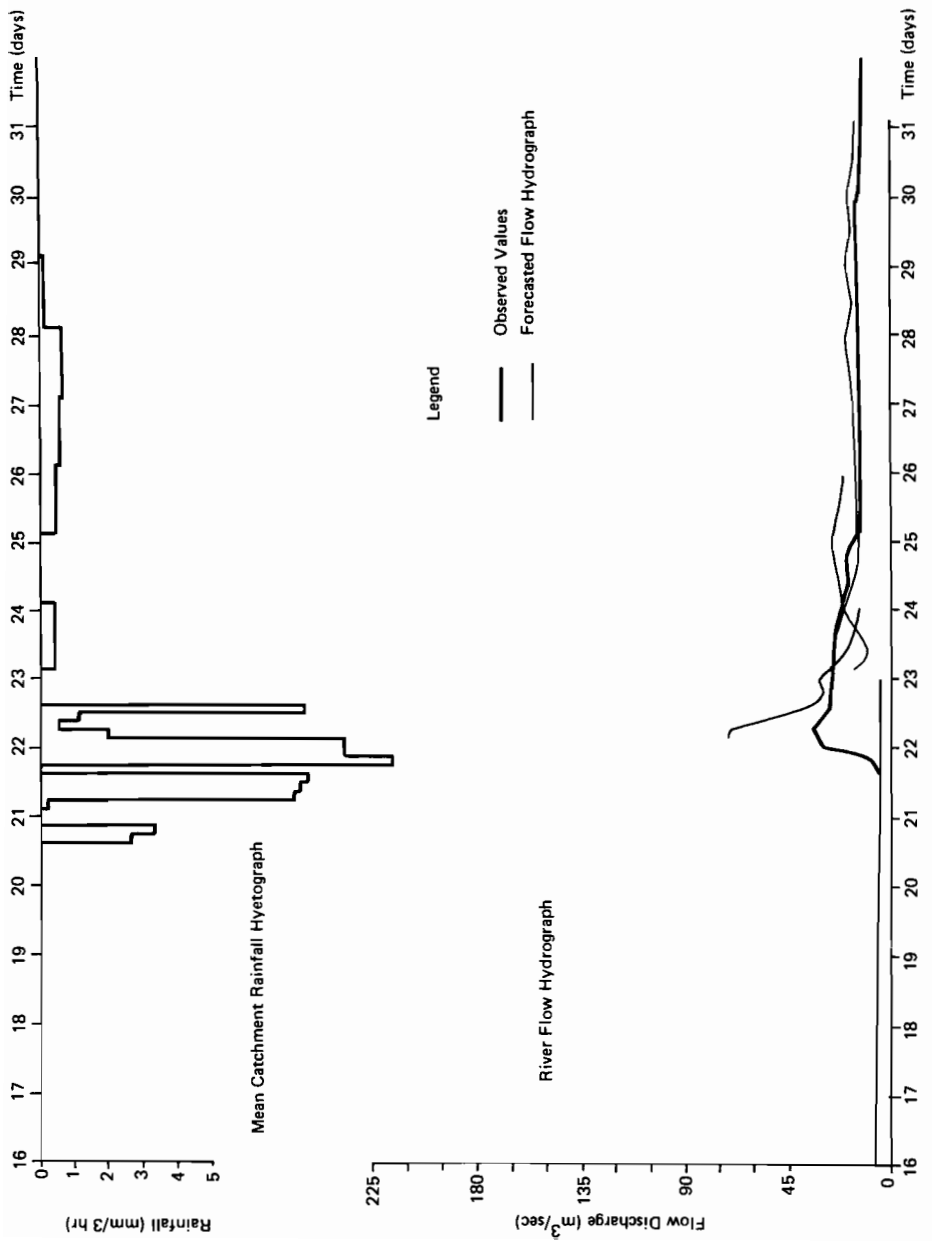


Fig. 17.3 Plot of variant Q1 of model Anna for the Dunajec River, cross-section Kowaniec, for the period 16-31 July 1976. Forecasted precipitation amount is always zero.

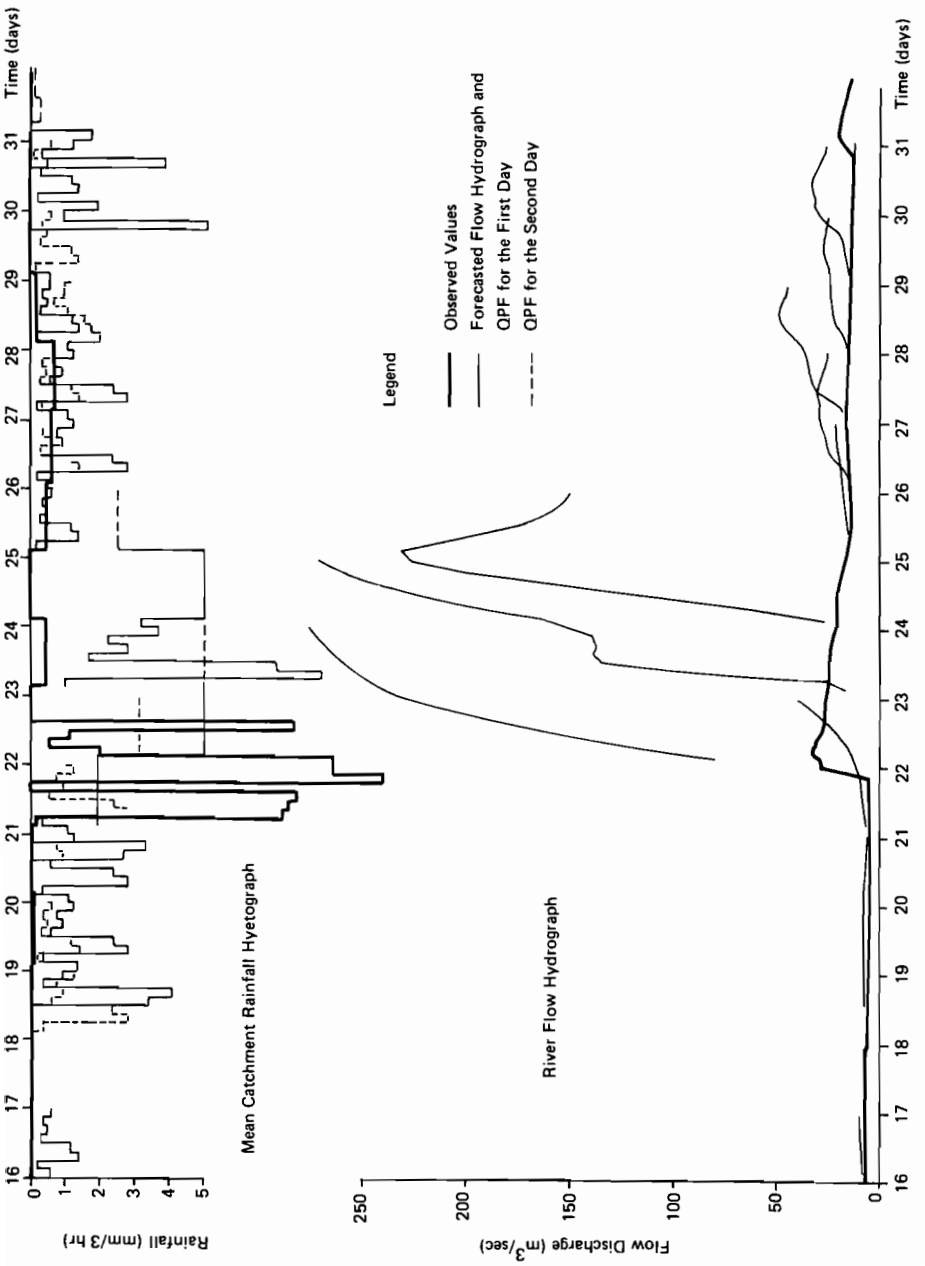


Fig. 17.4 Plot of variant Q_2 of model Anna for the Dunajec River, cross-section Kowaniec for the period 16-31 July 1976. QPF stands for quantitative precipitation forecasts.

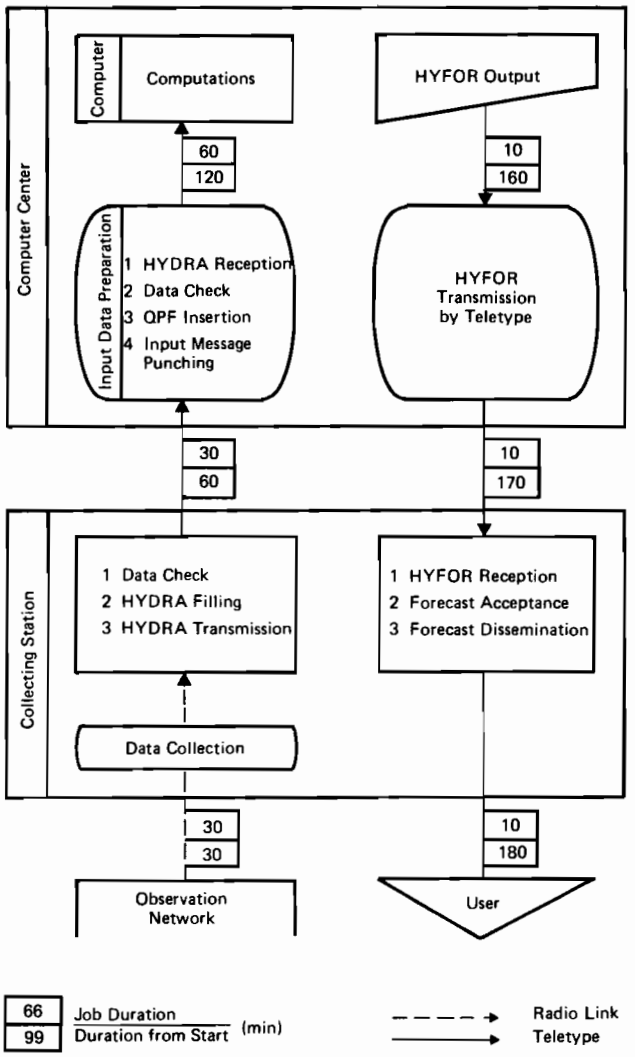


Fig. 17.5 A diagram of data collection, transmission, processing, and dissemination for the Sola and Dunajec River forecasting systems. HYDRA is the input data message; HYFOR is the output data message. QPF stands for quantitative precipitation forecasts.

future, a step-by-step automation of all operations which are now done by hand is expected; but even after automation of all the operations a meteorologist will still be in command of the system operation as an integral "on-line" part of the total, a man-machine system.

THE DUNAJEC RIVER SYSTEM

The Dunajec, a mountain tributary of the Vistula, rises from the Tatra mountains.

The flow forecast is computed for two cross sections: Kowaniec with a catchment area of 680 km^2 and an altitude between 250 and 580 m above sea level, and Kroscienko with a catchment area of 1,579 km^2 and an altitude between 250 and 420 m above sea level. About 30 percent of the catchment is forested.

The travel time of the water after a heavy rain is a few hours for the Kowaniec sub-basin and about 12 hours for the Kroscienko. An observation network incorporated into the system consists of 17 recording rain gauges, of which 11 belong to the Kowaniec catchment, two river stage recording gauges, at Kowaniec and Kroscienko, and one synoptic meteorological station equipped with actinometric instruments. This station at Zakopane collects information from the network for the system operation, and transmits the input message by teletype to the computer center.

The flow forecast for the two stations is computed using a catchment model with a structure similar to that of the MONIKA model applied for the Sola River. The difference between the two models lies in the values of the constant parameters. The output forecast messages for the Dunajec stations have a format identical to that used for the Sola River. The processing time for the forecast, under normal conditions, is the same as for the Sola River, about 3 hrs (see Fig. 17.5), and in flood conditions it is about 2 hrs.

EVOLUTION OF THE SYSTEM

The system was put into service in 1973 on the Sola River, in 1974 on the Dunajec River for the Kowaniec station, and in 1975 for the Kroscienko. Every year, after an evaluation of the system performance in the preceding period, some improvements are introduced. This is a cyclic process of continuous system modification and improvement. The main modifications up to date are as follows:

Reporting network. The number of reporting posts incorporated in the system has been increased. In the Sola catchment area there were, in the beginning, 9 rain gauges and 1 river stage gauge. Now the network consists of 11 rain gauges, a climatological post, an actinometric post and the river stage gauge. In the Dunajec catchment area in the beginning there were 11 rain gauges and 1 river stage gauge (Kowaniec). It too has been substantially enlarged. In 1973 and 1974 only rain and river stage gauges were working in the system. Since 1975 the climatological posts, a meteorological synoptic station, and an actinometric post were incorporated into the system reporting network for evapotranspiration estimation.

Telecommunication network. Since 1973 the reliability of the telecommunication system has been improved. Most of the posts are equipped with two types of communication: telephone and radio operating in the 34-MHz frequency band. In addition, the collecting stations are equipped with teletypes. All telecommunication facilities are operated by people rather than machines.

Data processing facilities. In 1973 and 1974 the forecast computation was performed on an ODRA-1204 computer and since 1975 by the larger ODRA-1305. This hardware improvement created the possibility to improve the forecasting model and to speed up computation time.

Precipitation data processing. The rain recording gauges are quite often out of order. In these cases, the 24-hr total precipitation from the standard rain gauge is the only available information. In order to provide additional information for such circumstances, 12-hr precipitation totals from the standard rain gauges were introduced into the HYDRA input message in 1975.

Initially, the time distribution of the forecasted rainfall was simulated by a

random number generator giving a random time distribution of the forecasted 24-hr total. In 1974, this was replaced by the 4 distribution functions related to storm patterns that were described above. This produced better model performance.

Evapotranspiration estimation. Initially, evapotranspiration was estimated by using average daily climatological values for the river catchments under study. The daily values have been uniformly divided into 3-hr intervals. In 1975, estimation of the actual evapotranspiration as described above was introduced and resulted in the improvement of hydrograph forecasting.

Meteorological forecast data. Initially, only quantitative precipitation forecasts (QPF) 24 hrs ahead were given. In 1974, the time distribution of the forecasted rainfall and the initial hour of the forecasted rainfall were added. In the 1973 and 1974 model versions, the QPF were given for the first 24 hrs of the 48-hr forecast period and the forecasted rainfall amount for the second day was assumed to be equal to zero. Since 1975, QPF are given separately for the first and second day of the forecast period (see Fig. 17.2). In 1975, a quantitative forecast of the meteorological and radiation variables needed for evapotranspiration estimation was also introduced. These were estimated separately for each day of the forecast period.

Interception storage. This variable was absent in the 1973 and 1974 model versions. At that time, the simulated hydrograph was rising much faster than the observed hydrograph when light rain occurred after a dry period. This meant that the model did not take into account the initial losses due to the interception and the surface detention storage. In 1975, interception storage, as described above, was introduced.

Baseflow procedure. The model was calibrated on a set of flood events which resulted in a fairly good fit between simulated and observed hydrographs. However, when the model was operated on a day-to-day basis during low flow periods, the computed hydrographs dropped faster than the observed ones.

As a result, the two separate baseflow procedures, as described earlier, were introduced in 1976. Those two procedures represent two distinct flow regimes; Baseflow 1 reflects the surface and the subsurface flow components and Baseflow 2 reflects the prevailing groundwater components of the river channel flow. This model modification has improved flow forecasts.

SYSTEM PERFORMANCE

The quality of the forecasts by the system in 1973 and 1974 was not completely satisfactory. While the system improvements described above have eliminated several deficiencies, the goodness-of-fit of the computed hydrograph still depends upon the QPF. However, the relation between QPF and the flow forecast error depends on the initial conditions.

The last model version was implemented on July 7 1976. This version gives fairly good results as can be seen from Figs. 17.3 and 17.4 and Table 17.1.

Note that the model sensitivity to the QPF errors depends on the soil moisture conditions. For instance, on July 22 a considerable amount of precipitation was observed (see Fig. 17.3) which was underestimated by the QPF as seen in Table 17.1. This big QPF error however did not produce a flow volume forecast error of the same magnitude because the rainfall occurred after 10 (essentially) rainless days when the soil moisture deficit was high. On the other hand, after two rainy days, the forecasts for July 23, 24, and 25 produced large overestimates of QPF resulting in large flow volumes (see Table 17.1).

Table 17.1. Forecast errors for the Dunajec River, cross section Kowaniec for the period July 21-26, 1976

	Precipitation, P (mm/24 hr)			Flow Volume, V($10^6 m^3$/24 hr)				
					Q_1 Variation		Q_2 Variation	
Date	Observed	Forecast	ΔP	Observed	Forecast	ΔV_1	Forecast	ΔV_2
July 21	5.9	10	+4.1	0.63	0.63	0	0.72	+0.09
July 22	50.1	15	-34.1	1.07	0.68	-0.39	0.98	-0.09
July 23	11.2	40	+28.8	3.53	5.70	+2.17	19.60	+16.07
July 24	3.7	30	+26.3	2.98	1.90	-1.08	12.20	+9.22
July 25	0	40	+40	2.39	2.86	+0.47	15.80	+13.41
July 26	3.81	5.0	+1.19	1.85	1.80	-0.05	2.04	+0.19

NOTE: All forecast values, FOR, are for the first day of the forecast period.

During periods with little or no rain, when the 24-hr precipitation totals were below 10 mm and the QPF or the QPF errors were of the same order of magnitude, then the flow forecasts were fairly good (see Figs. 17.6 to 17.9), especially the Q_1 forecasts. This means that when the Carpathians had little or no rain, Q_1 (QPF equal to zero) provides the most probable forecast. For the days with considerable precipitation, a good QPF is very important, especially for heavy, flood-producing rainfall. However, obtaining accurate QPF with a sufficient lead time (i.e., 24 hrs) seems at present to be unlikely, especially for small mountain basins.

Figures 17.6 to 17.9 present model performance as related to the forecasts for the first day of the 48-hr forecast period. The corresponding figures for the second day were worse, though the scatter diagram of precipitation was very similar to Fig. 17.8. Similar results were obtained for August 1976, but they are not presented here.

In comparing the scatter diagrams of Figs. 17.8 and 17.9, one can note that the model is filtering the QPF errors, and that the scatter diagram of flow volumes is better than that of precipitation. This is further evidence of fairly good model performance.

Forecast errors estimation. The hydrograph forecast errors are random variables resulting from a nonstationary stochastic process. Its nonstationarity depends on the nonstationarity of the runoff process and its related factors, including the length of the forecast period. That is, with increasing length of the forecast period the standard deviation of the forecast variable increases. For this reason, all statistical estimates of the forecast errors may be applied for the given forecast time only (i.e., 24 hrs) and for a relatively short period (i.e., 1 month) during which the runoff process is supposed to be stationary.

Many statistical estimates of the forecast errors have been proposed in the

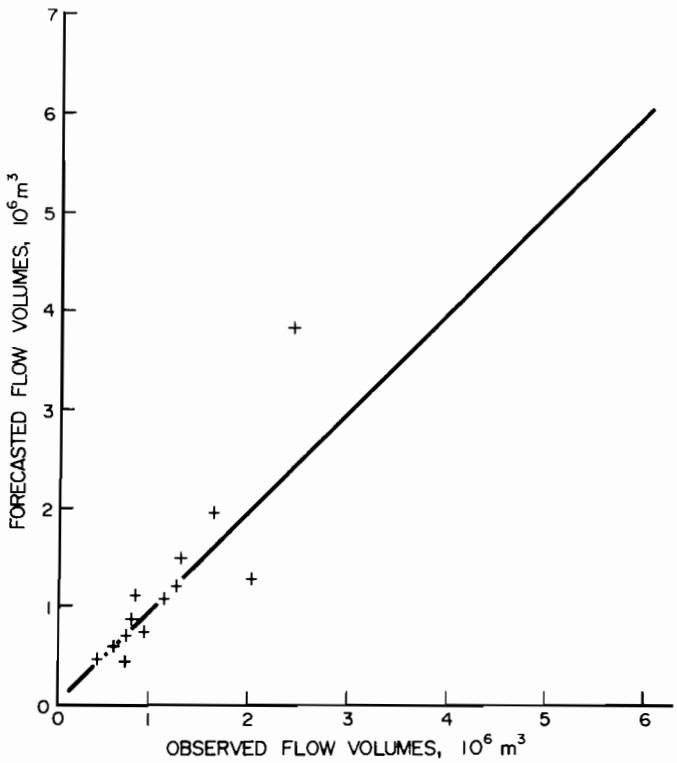


Fig. 17.6 Scatter diagram of forecasted versus observed 24-hr flow volumes using variant Q_1. (Dunajec River, cross section Kowaniec, 1-31 July 1976).

literature. The choice of the proper one is not an easy task. There are some estimates accepted by international organizations that allow intercomparison of a specific class of models. In order to obtain comparable results, we have adopted as a standard the following criteria used during the WMO Intercomparison of Conceptual Models Project (WMO 1975).

Coefficient of variation of residual of errors:

$$Y = \frac{\left[\frac{\Sigma(y_c - y_o)^2}{n}\right]^{1/2}}{y_o} \tag{17.8}$$

Ratio of relative error to the mean:

$$R = \frac{\Sigma(y_c - y_o)}{n\bar{y}_o} \tag{17.9}$$

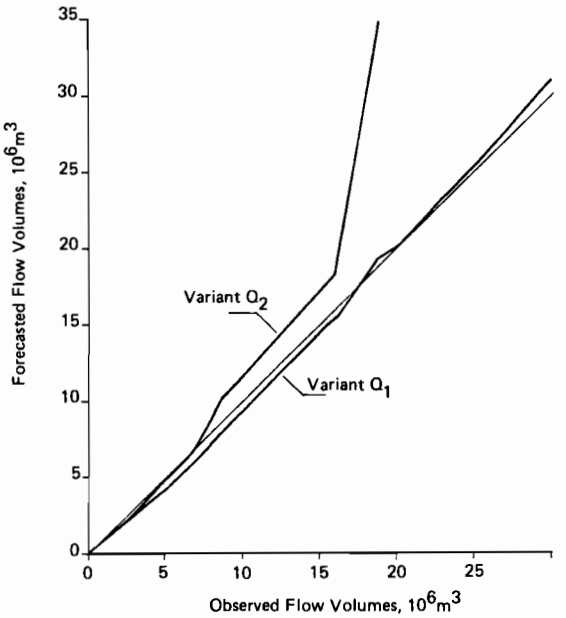


Fig. 17.7 A double mass plot of forecasted versus observed monthly flow volumes. (Dunajec River, cross section Kowaniec, 1-31 July 1976).

Ratio of absolute error to the mean:

$$A = \frac{y_c - y_o}{n\bar{y}_o} \tag{17.10}$$

where

y_o = observed discharge, y_c = computed discharge, n = total number of observations and

$$\bar{y}_o = \frac{\Sigma y_o}{n}$$

The above criteria are useful for a long series of observations and are sensitive to large errors like those presented in Fig. 17.4 for July 23, 24, and 25. When such errors occur in one month, all the statistics for this month are sharply increased by one order of magnitude. So, for short periods these criteria are not representative. The latest model version will be checked using these criteria on the historical data series produced by the previous model versions and, of course, on the new data series in the coming months.

THE RIVER STAGE HYDROGRAPH FORECAST SYSTEM FOR THE LOWER VISTULA RIVER

Downstream of the Włocławek dam and hydropower station, the daily river flows vary from 350 m^3/sec up to about 2,000 m^3/sec; the latter figure corresponds to the peak

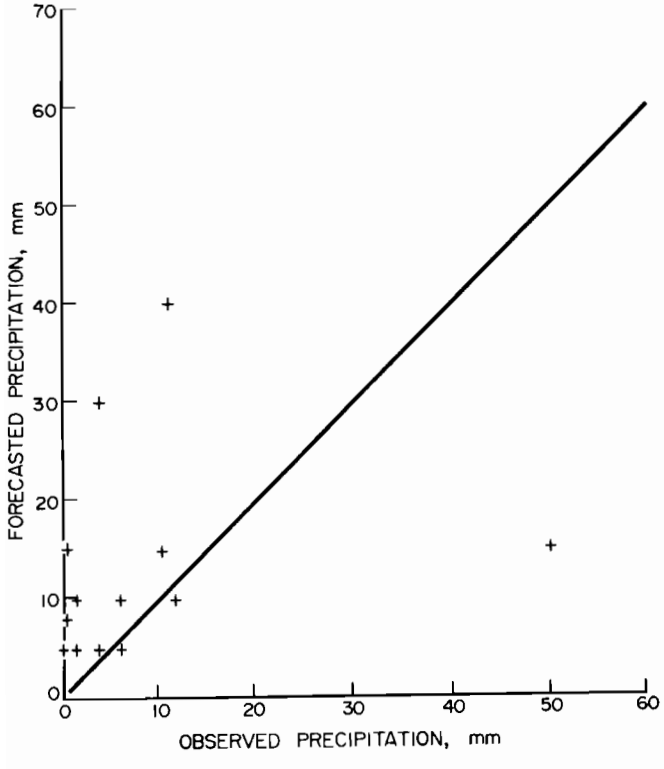


Fig. 17.8 Scatter diagram of forecasted versus observed 24-hr mean catchment precipitation. (Dunajec River, cross section Kowaniec, 1-31 July 1976).

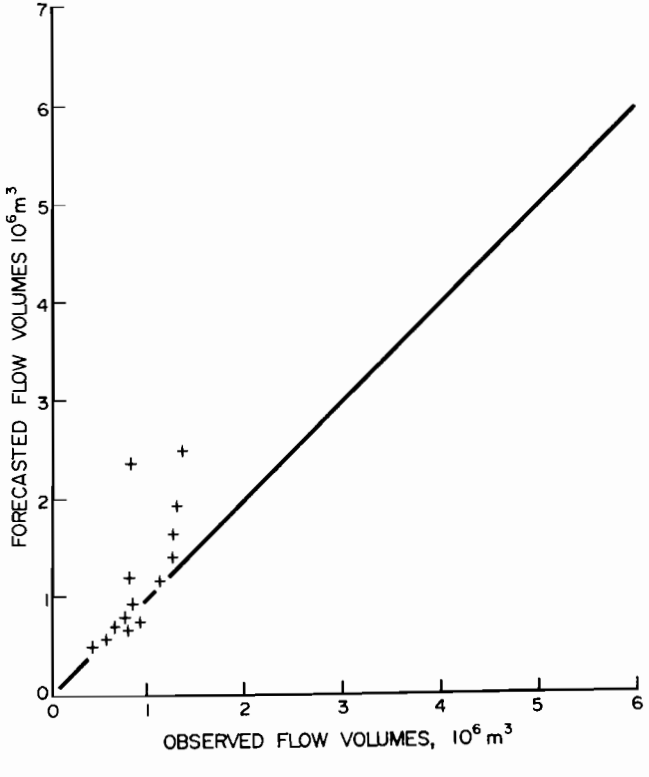


Fig. 17.9 Scatter diagram of forecasted versus observed 24-hr flow volumes using variant Q_2. (Dunajec River, cross section Kowaniec, 1-31 July 1976).

power generation of the station. Usually there are at least two peaks daily. The range of river stages directly downstream of the dam is about 1.5 m.

Such unsteady flow conditions create certain difficulties for navigation along the 100-km reach downstream of Włocławek. In order to facilitate navigation, a river stage hydrograph forecast is needed to provide a critical depth hydrograph forecast at the specific river cross sections where the shallows exist. The shallows are moving downstream, so the model has to provide a hydrograph forecast for any chosen cross section. This problem cannot be solved using conventional methods such as river stage relationships.

The travel time along the river reach under discussion is about 24 hrs, the same as the hydrograph period. Input data for the system is as follows: the hydrograph of discharge from the power station during the last 24 hrs, the forecasted discharge hydrograph for the next day and the river stage hydrographs from the three recording gauges located at the top, middle, and lower end of the river reach for the last 24 hrs. The upper gauge records stages with 15-minute intervals, the middle and the lower gauges with 1-hr intervals. The output from the system is the river stage hydrograph for the middle and the lower gauges and a critical depth hydrograph for specific cross sections.

The river reach is modeled in two segments. For the upper segment, the model is structured as a cascade of linear storages and linear channels, whereas for the lower reach the model is based on kinematic wave equations. The latter model was calibrated using results of continuous discharge measurements at the lower two gauges over a week. The linear cascade model was calibrated using optimization methods. The system is now being implemented and further system developments will be performed after an evaluation.

CONCLUSIONS

After several years of development and modification, a daily flow forecasting system is providing valuable information for users and performs fairly well. The amount of information is greater and the quality higher in comparison with conventional methods. The time for data collection, transmission, and processing has been shortened, although further reductions are needed.

Under flood conditions, the system produces forecasts every 3 hrs, which provide valuable information for the forecaster-hydrologist working under considerable stress.

The system collects input and output data files on magnetic tapes; the input data are checked, so the data files are of higher quality than those stored by conventional methods, and the files are readily accessible for any kind of processing. The system development has lead to improvements in both the reporting network and telecommunication facilities and also in the standard of the staff. One of the main achievements has been the establishment of a system development team which has gathered considerable experience and is a basic requirement in obtaining satisfactory results.

Other key conditions for successful system development are close cooperation with an existing network, the good telecommunication and data processing facilities of the meteorological service, and easy access to meteorological output such as weather analyses, quantitative precipitation forecasts and other variables. From this point of view, a combined meteorological and hydrological service has a distinct advantage over separate services.

The first forecasting models were calibrated on historical data that were incomplete

and of poor quality. As a result, the outputs from these models were not very good. After this initial operation, the input and output data files were improved and the model was recalibrated (with some modifications). This process of model development presented here represents the way the forecasting system in Poland is developing. It seems that such an approach can be applied in many situations.

It appears that the lack of adequate data for model calibration and testing is a common situation. For this reason, it is important to start operation as soon as possible in order to gather data, even if the model forecasts in the beginning are not very good. In operational real-time forecasting systems, greater effort should be devoted to data collection, transmission, and processing subsystems, including data checking and estimation of missing data, in order to obtain the best possible hydrological analysis. The sophistication of the forecasting model, which transforms the runoff volume into a flow hydrograph, is not so important in the initial stage of model development in comparison with the data processing procedures. Improving the model seems to be the next step after gathering the proper data.

For mountainous catchments with a lag time of a few hours, the model performance is highly dependent on the accuracy of the quantitative meteorological forecasts, especially those of precipitation during floods. Since the accuracy of QPF is not reliable, it seems reasonable to issue flow forecasts for a few precipitation conditions. When QPF are not available, an alternative approach would be to issue the flow forecasts, assuming no future precipitation, frequently and to shorten the whole operational cycle. Thus, the forecast after each cycle will be adapted to the observation. Such an approach would enable the user to make frequent adjustments to his operation and shorten the time that the model runs on inaccurate data. This could be helpful during the flood period.

The development of on-line automated systems incorporating radar for the aerial precipitation estimation is the way to improve the system. These systems however should have a person as an integral "on-line" part of the operation.

REFERENCES

Bobinski, E., T. Piwecki, and J. Zelazinski (1975). The mathematical simulation model for the Soła and Dunajec rivers real-time flow forecasting. International Symposium and Workshops on the Application of Mathematical Models in Hydrology and Water Resources Systems, Bratislava.

Rockwood, D.M., E.M. Davis, and J.A. Anderson (1972). User manual for COSSARR model. U.S. Army Engineer Division, North Pacific, Portland, Oregon.

World Meteorological Organization (WMO) (1975). Operational hydrology report No. 7. Intercomparison of conceptual models used in operational hydrological forecasting. WMO Report No. 429, Geneva.

18 The United States Weather Service River Forecast System

David C. Curtis and George F. Smith

The primary purpose of the United States National Weather Service (NWS) hydrology program is to provide accurate and timely hydrologic information to the general public. While flood forecasts and warnings are the most widely known hydrology products, NWS river forecasts are also used for water supply, navigation, irrigation, power, reservoir operation, recreation, and water quality interests. These forecasts are an effective tool in the development and management of water resources.

The River Forecast Center (RFC) is the focal point of the river forecast network (Fig. 18.1). It is staffed by professional hydrologists who receive hydrometeorologic data, prepare forecasts, and transmit forecasts to other Weather Service Forecast Offices (WSFOs) for dissemination. Accurate and timely forecasts can minimize loss of life and property due to extreme river flows or related events.

Twelve RFCs prepare river forecasts and warnings for approximately 2,500 communities. Approximately 97 percent of the United States (including Alaska) is covered by this service. The area of responsibility of each RFC includes at least one major river system (Fig. 18.2).

Forecasts of seasonal snowmelt or yearly water runoff are prepared by five RFCs in the Western United States. Two additional RFCs in the Northeast prepare seasonal snowmelt and monthly runoff forecasts. Water supply forecasts for 600 points where snow is the principal source of stream flow are distributed to water users monthly by local WSFOs.

RIVER FORECASTING TECHNIQUES

In the late 1960s a commitment was made by the NWS to move from an index-type catchment response function to continuous conceptual hydrologic models for use in river forecasting. Conceptual models with a strong physical basis have several distinct advantages over index-type relationships, as follows:

- Accurate mathematical representation of a catchment enhances the probability of adequately predicting future events, especially events of a magnitude unexperienced in the past.

- Parameters based on conceptual considerations can sometimes be subjectively altered to reflect changes made or to be made to the physical characteristics of the catchment.

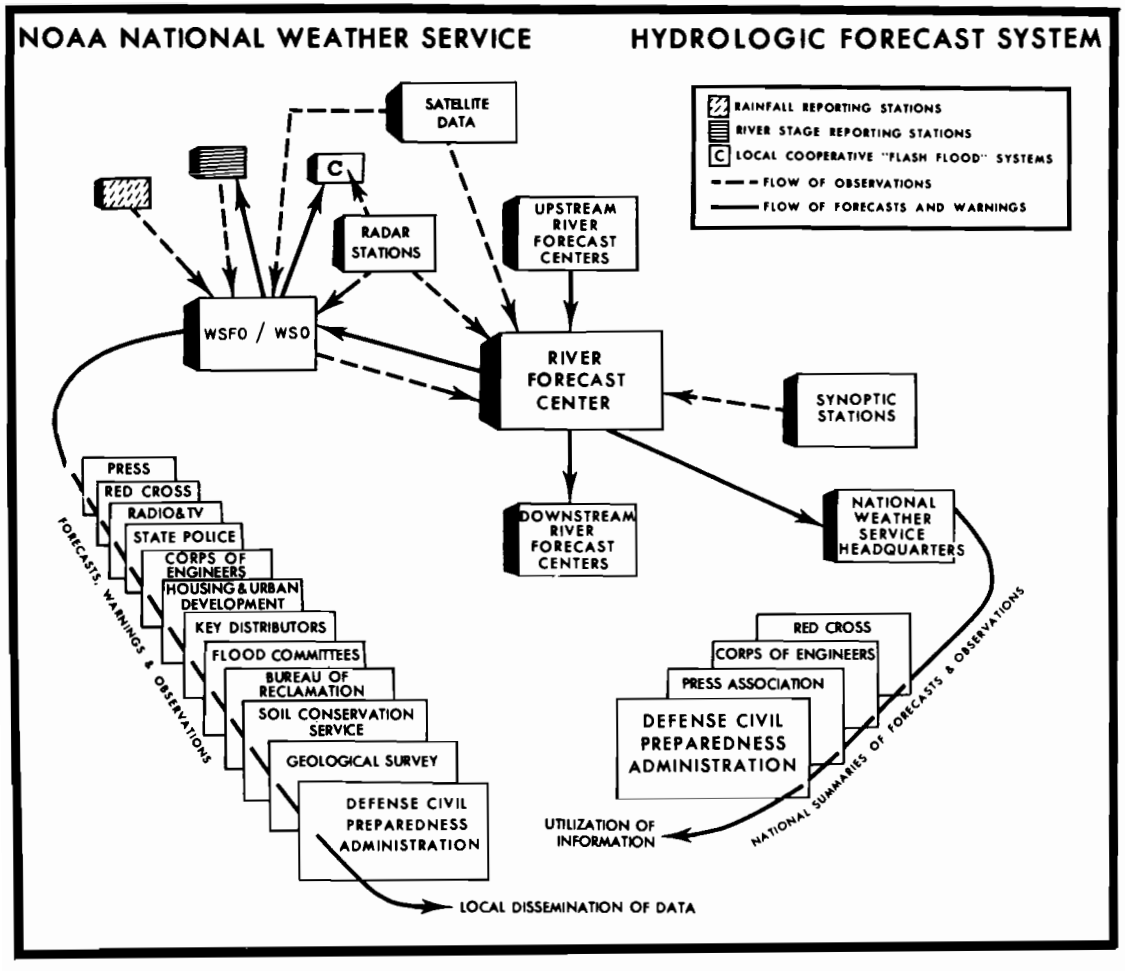


Fig. 18.1 National Weather Service River Forecast System. WSFO stands for Weather Service Forecast Office, WSO for Weather Service Office, and NOAA for National Oceanic and Atmospheric Administration.

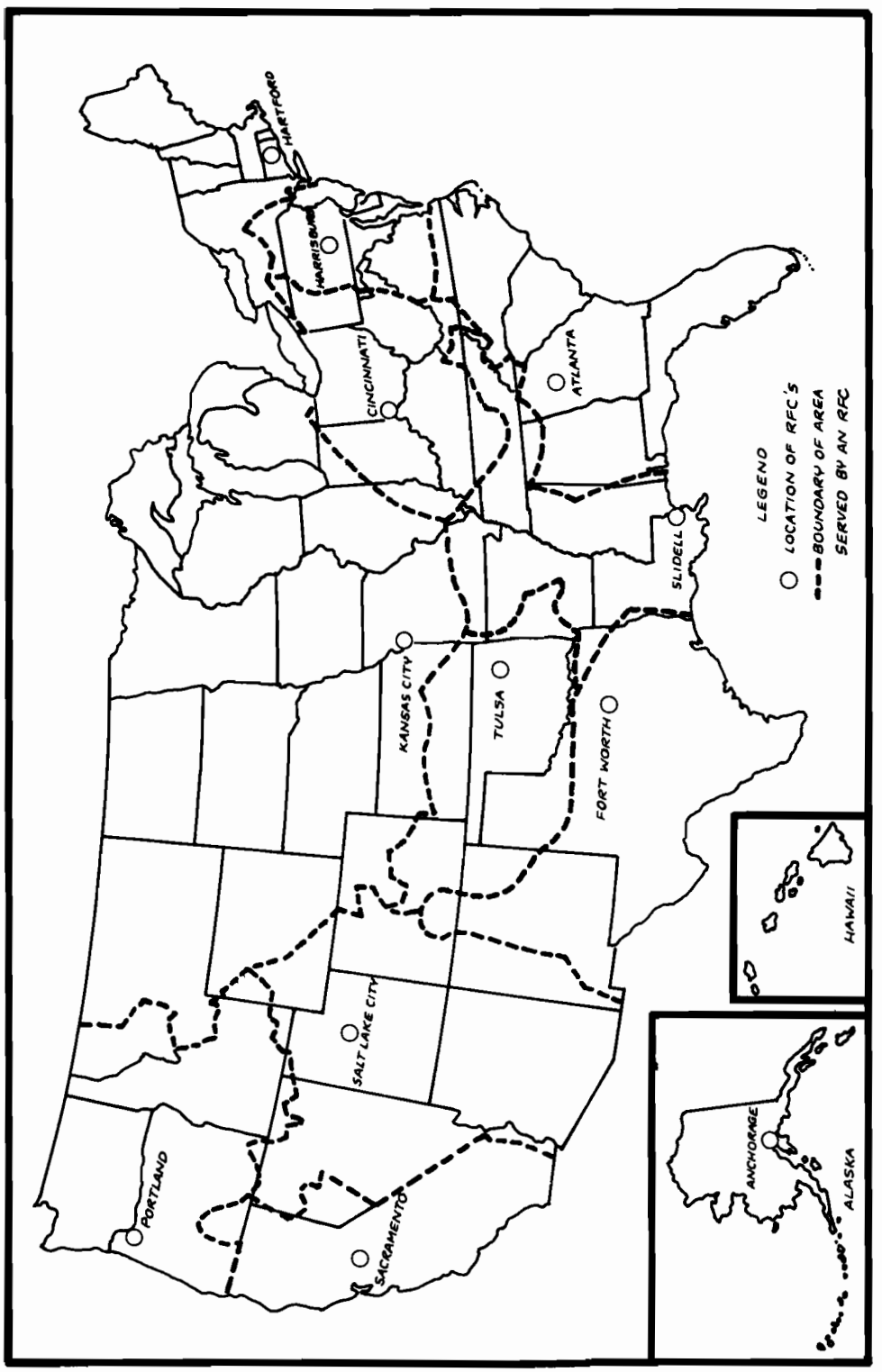


Fig. 18.2 River forecast areas of responsibility.

- A conceptual model can be extended to problems other than simulating catchment streamflows. For example, algorithms could be added to simulate the movement of pollutants through the soil matrix.

- A model that is physically based is an effective tool for future research and modification.

Studies were undertaken by the NWS Hydrologic Research Laboratory (HRL) to determine which of the hydrologic models available at the time were best suited for river forecasting. Based on these studies, a modified version of the Stanford watershed model (3) was included in the initial version of the National Weather Service River Forecast System (NWSRFS) (4). Since 1971, several important modifications to the NWSRFS have been made.

- The Stanford soil moisture accounting routine has been replaced by a model developed by the NWS RFC at Sacramento, California (5).

- A snow accumulation and ablation model has been added (6).

- A dynamic river routing model has been added (7).

- The data management capabilities have been greately expanded and changed from a data storage system oriented to magnetic tape to a direct access disk storage system.

This paper summarizes the various elements and experiences of the NWSRFS to date. The NWSRFS is the set of techniques and computer programs used to produce river forecasts. Included are programs to manage the large volumes of data associated with a national forecasting system and programs to perform the hydrologic and hydraulic computations necessary to forecast river system response.

The basic elements of the NWSRFS are:

- Data management. Routines that store, retrieve, and manipulate data from the appropriate direct access disk files.

- Soil moisture accounting. Routines that simulate the movement of water through the soil profile.

- Snow accumulation and ablation. Routines describing the growth and subsequent melting of a snow cover.

- Channel routing. Hydrologic and hydraulic techniques to route flows through natural channels.

- Mean areal precipitation. Routines converting point precipitation values to areal means.

- Mean areal evapotranspiration. Routines to compute mean areal evapotranspiration.

- Mean areal temperature. Routines to convert point temperature values to areal means.

CALIBRATION DATA MANAGEMENT

Vast amounts of data are required to implement the NWSRFS throughout the United States. One component of the NWSRFS is designed to manage and prepare data for use

in hydrologic calibration models. Except for units conversion, the data management portion of the NWSRFS does not transform the data. It is a system that selects appropriate time series from a magnetic tape or card source and transcribes them into a format more efficiently managed and used by the hydrologist.

The data management system consists of three parts (Fig. 18.3). Each of these

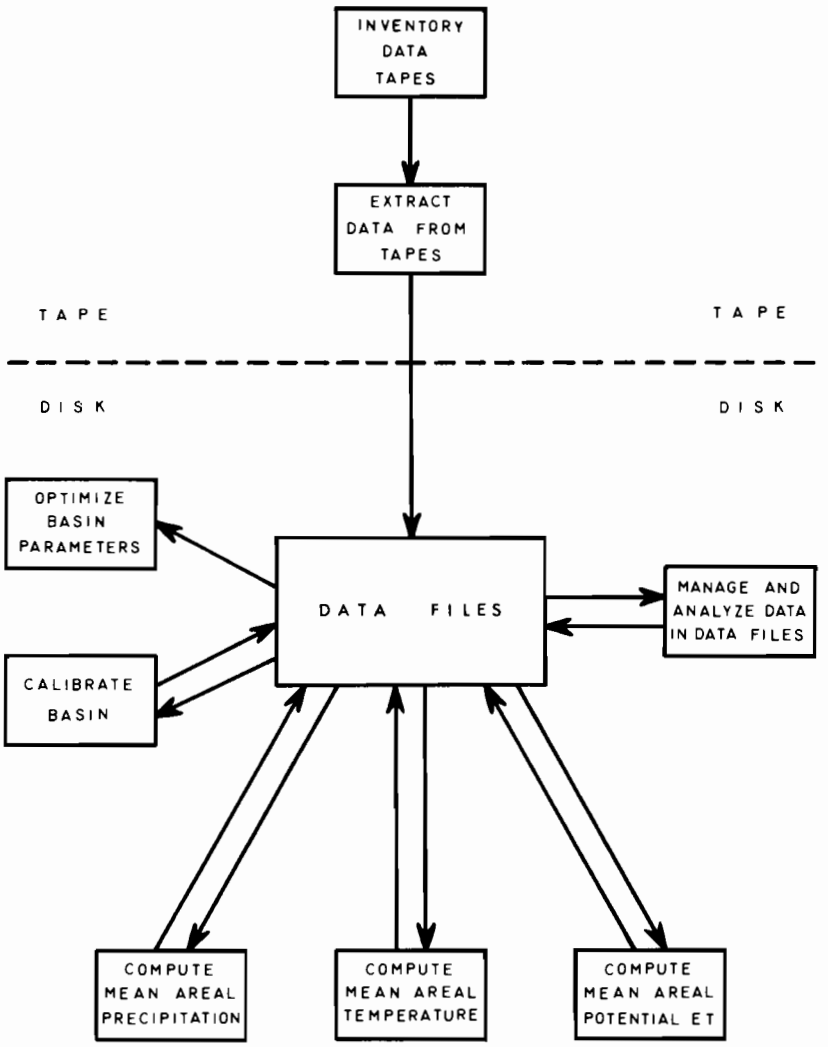


Fig. 18.3 Calibration data management system. ET stands for evapotranspiration.

parts performs a function necessary in reducing the available data into an efficient format for use by the hydrologic programs. The three parts are:

- Inventory of data on National Climatic Center (NCC) or U.S. Geological Survey (USGS) magnetic tapes.

- Copying of selected data from the tape of direct access data files.

- Management of data in data files.

The four types of data stored on magnetic tape are hourly precipitation, daily climatological observations, synoptic meteorological observations, and daily stream-flow data. For each data type, programs are available to inventory a tape, extract and label selected data (time series), and place them into data files, on a direct access disk. Once the various data are in data files, a package of subroutines can be used to manage and analyze these data. This package of subroutines is called the data file utility system and can perform such tasks as listing which files are available for data storage or the contents of a particular file, copying all or part of a file from one location to another, editing data within a file, removing unwanted data from a file, and plotting one or more time series as a function of either time or another time series.

A key component of the file utility system is the labeling of each time series to allow rapid data access and retrieval. This label, or time series header, contains unique identifiers, as well as information about the locations, length, and type of data in the time series. Identifiers contained in the time series header enable the direct access device to move to the exact location of the requested data residing in a disk file. This shortens computer access time by eliminating sequential searches of entire data files to find the desired time series.

SOIL MOISTURE ACCOUNTING

The version of the Sacramento soil moisture accounting model (Fig. 18.4) included in the NWSRFS is a deterministic model with a limited distributed parameter capability. Soil moisture accounting is accomplished for a particular soil moisture accounting area. Within a soil moisture accounting area, precipitation inputs are evenly distributed and soil conditions are assumed uniform. Thus, with respect to a soil moisture accounting area, the model is a lumped input and lumped parameter type. However, it is not necessary for the boundaries of a soil moisture accounting area and the natural catchment to coincide. This feature allows the catchment to be described by more than one soil moisture accounting area. If the inputs and parameters differ between soil moisture accounting areas, the model is "distributed" with respect to the inputs and parameters for the catchment.

On a vertical plane, the model defines two soil moisture accounting zones. An upper zone represents the upper soil layer and interception storage, and a lower zone generally accounts for most of the soil moisture and the groundwater storage.

Moisture storage

Both the upper and the lower zones store "tension" and "free" water. Tension water storage represents water closely held by soil particles. Free water storage represents water that is available for drainage, either horizontally or vertically. In the upper zone, tension water requirements must be met before water is transferred to upper zone free water storage. The stipulation that tension water requirements be met before substantial drainage begins represents the movement of a wetting front through the soil mantle. In the lower zone a fraction of the incoming water can be directly transmitted to free water storage even if the lower zone tension storage is not full. The capacity to "short circuit" tension water requirements in the lower zone aids the simulation of catchments where significant lower zone drainage is evident, even though area-wide lower zone tension water requirements have not been fulfilled.

Free water can move vertically through percolation or horizontally as interflow; it can be depleted by evapotranspiration, or it can replenish tension water

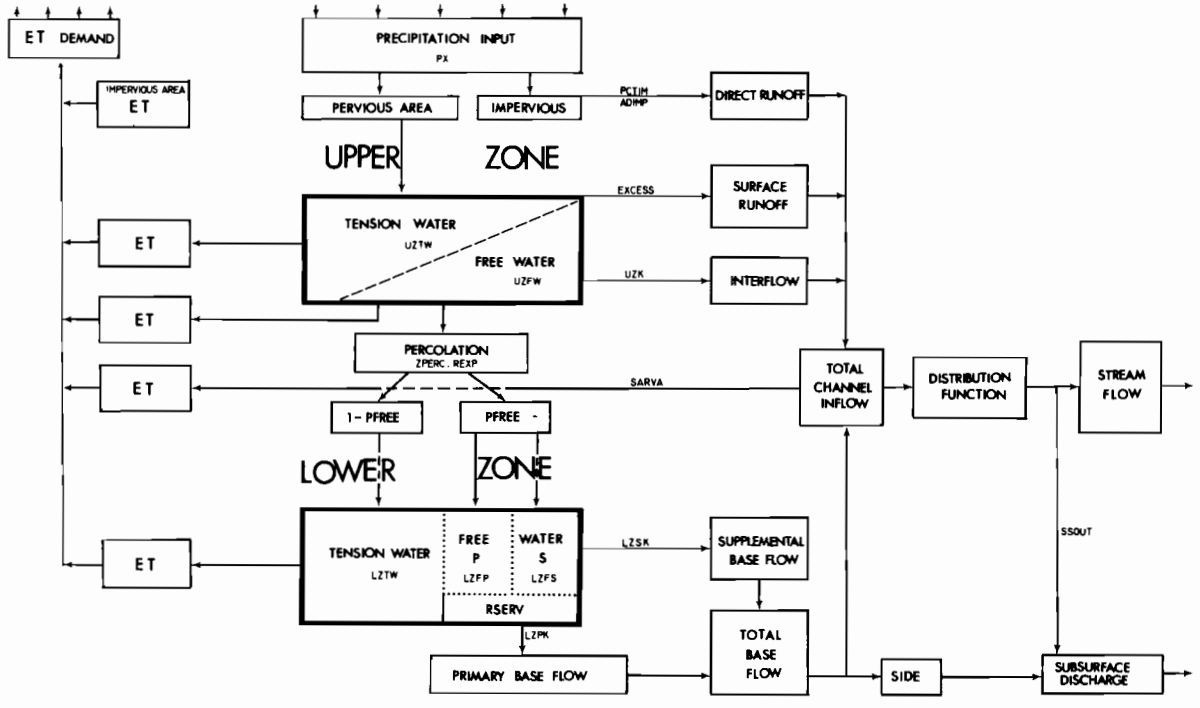


Fig. 18.4 Sacramento soil moisture accounting model. ET-evapotranspiration, PX- precipitation inflow, UZTW-upper zone tension water, UZFW- upper zone free water, ZPERC and REXP-coefficients regulating the percolation from the upper zone to the lower zone (ZPERC refers to tension water and REXP to free water), PFREE-coefficient for calculating percolation of the free water, LZTW-lower zone tension water, P- percolated water, LZFP- lower zone free percolation, S-water storage, LZFS- lower zone free water storage, RSERV-reserve storage for free percolated water, LZPK-coefficient for response of free percolated water to primary base flows, PCTIM and ADIMP-coefficients that transform the precipitation on the impervious ground into discharge- versus-time outflows, UZK-coefficient of the response of the upper zone free water for calculating the amount of flow in the unsaturated zone, SARVA-coefficient used to calculate evapotranspiration from the channel, LZSK-coefficient for calculating the lower zone supplemental base flow from lower zone storage, SSOUT-coefficient for calculating the infiltration from the stream to the groundwater that leaves the catchment as groundwater. SIDE refers to groundwater that does not show up in the stream because it leaves the catchment through the sides.

requirements. Tension water storages can only be depleted by the evapotranspiration process.

Percolation

Movement of water from the upper zone to the lower zone is controlled by a percolation algorithm that relates the contents and capacities of upper zone storages as well as drainage parameters for the respective free water storage. The formula controls the movement of water in all portions of the soil profile, both above and below the percolation interface, and is itself controlled by the current state of the soil moisture storage system.

Evapotranspiration

In most rural catchments, evapotranspiration is a dominant hydrologic process; thus, accurate continuous hydrograph simulation is heavily dependent upon successful description of evapotranspiration. Two types of evapotranspiration information can be input to the Sacramento soil moisture accounting model: a seasonal evapotranspiration demand curve or potential evaporation data with an adjustment curve to account for the effect of the current state of the vegetation cover on the actual evapotranspiration.

Variable Impervious Area

A fraction of the precipitation falling on a particular catchment is assumed to be deposited on impervious area directly connected to or adjacent to the channel system. This fraction contributes directly to channel flow and does not enter the soil matrix. In the Sacramento soil moisture accounting model a minimum and maximum percentage impervious area is specified on the theory that as a given soil moisture storage becomes satisfied, an increasing amount of pervious area begins to behave as impervious area. An algorithm evaluates the current state of the soil moisture storage system and adjusts the total percentage of impervious area accordingly.

Flow components

The model recognizes and generates five components of channel flow:

- Direct runoff, resulting from moisture applied to the variable impervious area.
- Surface runoff, resulting from moisture applied at a rate faster than upper zone intake.
- Interflow, i.e., lateral drainage from upper zone free water storage.
- Supplementary base flow, i.e., drainage from lower zone supplementary storage.
- Primary base flow, i.e., drainage from lower zone primary storage.

Computational Technique

Movement of water through a soil matrix is a continuous process. The rate of movement at a particular point is a function of the current state of the moisture supply and soil moisture storage system. A quasi-linear, open form computation is used to model soil moisture movement. Use of this technique assumes that movement of soil moisture during a time step is defined by conditions existing at the beginning of each time step. To use this assumption effectively, a short time step

must be selected. The basic computational interval of the NWSRFS is 6 hrs. However, in the soil moisture accounting model, time steps are set such that no more than 5 mm of water is involved in a single execution of the computational loop. This rather arbitrary limit has been set large enough to logically fulfill its function but not so small as to cause unwarranted execution times. The state of the soil moisture storage system can be output at the end of each 6-hr period.

SNOW ACCUMULATION AND ABLATION MODEL

The snow accumulation and ablation model is a conceptual model that describes the important physical processes taking place during the accumulation and ablation of a snow cover (Fig. 18.5). Although written for NWSRFS, the model can be used in conjunction with almost any soil moisture accounting and channel routing routine. Output from the snow model serves as input to the soil moisture accounting procedure. The output from the snow model is snow cover outflow (snowmelt water and rain water leaving the snow cover) plus rain that fell on bare ground.

The snow model uses air temperature as the only index of energy exchange across the snow-air interface. Two basic reasons exist for using air temperature as the sole index of energy exchange.

- Air temperature data are readily available throughout the United States on a real-time operational basis.

- Comparison tests conducted by the NWS HRL, though limited to two experimental watersheds, have indicated that at least in these two cases the hydrograph simulations produced by using air temperature as the sole index of snow cover energy exchange are comparable to those produced using an energy balance snow cover model.

Obviously, under certain meteorological and physiographic conditions, an energy balance model will provide more accurate estimates of snow cover energy exchange if the necessary data are available. Extensive research on snow cover energy exchange is being conducted by the HRL (8,9). It is planned to include an areal energy balance snow cover model in the NWSRFS within the next few years for use in areas where the necessary data are available and a meaningful increase in accuracy can be attained.

Model Components

Accumulation of the snow cover. The model first determines the form of precipitation input by using a reference air temperature. Precipitation falling when air temperature is greater than the reference value is assumed to be rain, and snowfall is assumed when air temperature is below the reference value. Generally, the reference temperature is set at about 1°C.

Accurate precipitation data are important if snow cover accumulation is to be simulated satisfactorily. Considerable variation can exist between actual and measured snowfall during a particular event resulting from inaccurate catch caused by the aerodynamic inadequacies of the precipitation gauge (10). Therefore a constant multiplier, called the snow correction factor, can be used to adjust recorded precipitation amounts to more adequately describe the actual snowfall.

Heat exchange at the snow-air interface. Heat exchange at the snow-air interface is the most critical factor in controlling the ablation of the snow cover. When air temperature is used as the only index to heat exchange, two basic situations arise for which heat exchange must be estimated: the ambient air is warm enough to cause melting at the snow surface, and the ambient air is too cold for melting to occur.

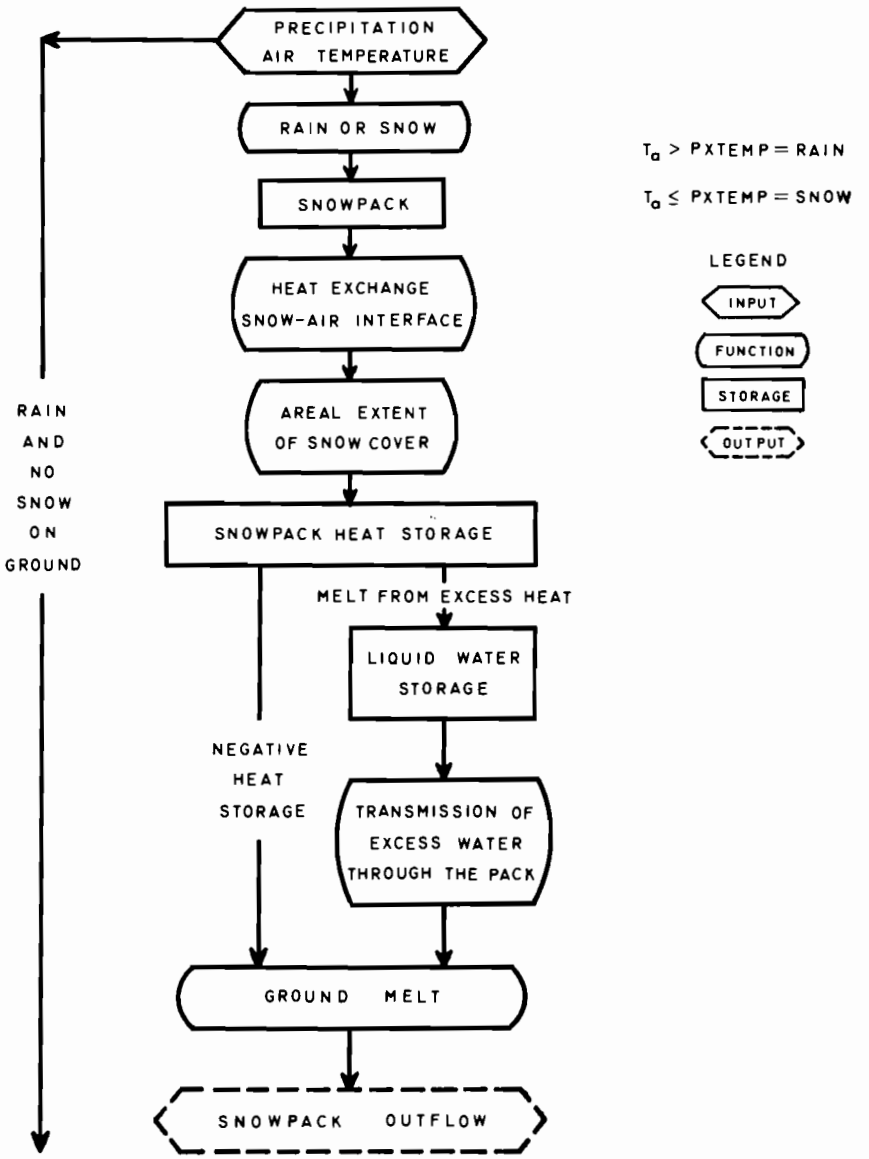


Fig. 18.5 Snow accumulation and ablation model.

The model assumes that melting can occur at the snow surface when air temperature is above a base temperature (usually 0°C). To calculate the melting rate, the model distinguishes between rain and nonrain periods.

Development of the energy balance equation to compute the snowmelt rate during rain is based on several assumptions: solar radiation is zero, incoming long-wave radiation equals the blackbody radiation at the ambient air temperature, snow surface temperature is 0°C, the relative humidity is 90 percent, and temperature of the rainwater is equal to the ambient air temperature. The energy balance of a melting snow cover is then expressed as the sum of the net radiation heat transfer,

latent heat transfer, sensible heat transfer, and the heat transfer by rainwater.

During nonrain periods, melt at the snow surface is assumed to be proportional to the difference between air temperature and the base temperature. A constant of proportionality referred to as the melt factor is used to linearly relate the temperature difference with snowmelt. To account for seasonal variation in various meteorological factors that affect melt, the melt factor is allowed to vary from a minimum on December 21 to a maximum on June 21. A sine curve is used to interpolate melt factors for other dates. The use of a sine curve to describe the seasonal variation in the melt factor has proved to be adequate throughout the conterminous United States. However, a different curve, which produces a more delayed increase in the melt factor, is provided for use in Alaska.

If the ambient air temperature is below 0°C, the model assumes that snowmelt does not occur. In this situation, the snow cover is either gaining or losing heat. The direction of the heat flow is dependent upon the relative temperatures of the snow surface (assumed equal to air temperature) and the temperature at some depth below the snow surface. Heat conduction to or from a snow cover is not only a function of the temperature gradient but also of the snow density. The model indirectly accounts for the effect of snow density on the rate of melt by seasonally varying the potential rate of heat transfer into or out of the snow cover.

Areal extent of snow cover. To estimate the total amount of melt generated over a given area, the portion of the total area covered by snow must be known. If rain is falling, the areal extent of snow cover must also be known to determine how much rain is falling on bare ground and how much rain is falling on the snow.

Since the snow accumulation pattern for a given area is reasonably similar from year to year, a unique curve can be drawn that relates the extent of snow cover to the current state of the snow cover in terms of water equivalent. Thus, once the water equivalent of the snow has been computed, the areal extent of the snow cover can be determined from the areal depletion curve.

Snow cover heat storage. The snow model keeps a continuous accounting of heat storage in the snow cover. Maximum heat storage occurs when the snow cover is isothermal at 0°C. When air temperatures are lower than the snow cover temperature, heat is transferred from the snow to the air, creating a heat deficit or negative heat storage in the snow cover. The model accounts for negative heat storage; and as air temperatures rise, sufficient heat must be transferred to the snow cover to eliminate existing heat deficits before melt can occur.

Liquid-water retention and transmission. Snow crystals form a porous medium similar to soil particles that retains and transmits water. The amount of liquid water that can be retained by the snow cover is assumed to be a constant percentage of the ice content.

Equations for the transmission of excess liquid-water through the snow cover are used to lag and attenuate the flow of liquid water to account for the time delay and storage characteristics of the snow cover.

Ground melt. Heat exchange at the soil-snow interface is usually negligible when compared with the heat exchange at the air-snow interface. However, in some catchments, sufficient melting takes place continuously at the bottom of the snow cover to maintain a significant base flow. Therefore, the model allows a constant amount of melting to take place at the soil-snow interface.

CHANNEL ROUTING

The soil moisture accounting model generates a volume of runoff available for channel inflow per 6-hr period. Initially, this volume of available water is assumed to be distributed uniformly over a soil moisture accounting area. To account for the temporal distribution of runoff volume reaching the catchment outfall, a time delay histogram and linear reservoir are used.

When runoff reaches the channel, the water is transmitted downstream in the form of a flood wave. As the flood wave moves downstream storage characteristics of the channel and fluid flow dynamics cause the shape of the flood wave to change. Normally, the flow exhibits a time lag between points in the channel due to finite wave speeds. Also, the flood wave is attenuated owing to the storage characteristics of the channel and, to a lesser extent, the inertial properties of the wave.

In rivers where the flow dynamics (i.e., backwater and/or flood wave inertial effects) are not important in describing the movement of a flood wave, a simple hydrologic (storage) routing procedure can be used. If the flow dynamics are important, a more complex hydraulic routing technique may be required.

Currently, the NWSRFS uses a simple storage routing technique known as lag and K in situations where flow dynamics are relatively unimportant (11). Experience has shown that the lag and K technique (both variable and constant with respect to flow) is very useful in describing flood flows in uncontrolled rivers where the effects of backwater and in-bank storage are negligible.

In rivers, reservoirs, and tidal reaches, where backwater conditions occur or where hydrologic techniques cannot achieve the desired degree of accuracy, a dynamic wave model (7) based on the following one-dimensional equations of unsteady flow is used:

$$\frac{\partial Q}{\partial x} + \frac{\partial A}{\partial t} - q = 0 \qquad (18.1)$$

$$\frac{\partial Q}{\partial t} + \frac{\partial [\alpha Q^2/A]}{\partial x} + gA\left[\frac{\partial h}{\partial x} + S_f\right] - qv_x + W_f B = 0 \qquad (18.2)$$

where

x = distance along the channel axis, positive in downstream direction;

t = time;

A = cross-sectional area of flow;

Q = average discharge across a section;

B = width of cross section at the water surface;

h = water surface elevation;

q = lateral flow per unit length along the channel; positive in inflow and negative in outflow;

v_x = velocity of lateral flow in the direction of the channel flow;

α = velocity distribution coefficient;

S_f = resistance slope given by the Manning equation for uniform turbulent flow;

W_f = energy loss due to wind per unit width of channel; and

g = acceleration due to gravity.

These equations can be used to predict stages and discharges. The first equation conserves the volume of water in the system and the second conserves the momentum.

Applications of these equations to a river system requires some idealization of channel geometry. Channel cross sections are specified at points along the river where significant changes occur. Typically, for large rivers with slowly varying transients, 15 to 25 km between cross sections will be sufficient. Observed stages and discharges can be used to determine channel roughness coefficients. With this information and the upstream discharge or stage hydrograph(s), flow can be routed through the river system.

The unsteady flow equations are solved by the "four-point implicit method," a finite difference technique. This solution technique allows unequal distance intervals between cross sections. Also, the numerical stability properties of the implicit method do not restrict the size of the computational time step (12). The desired accuracy can be the sole criterion in choosing the time step; and for non-tidal situations, time steps on the order of several hours can be used.

MEAN AREAL PRECIPITATION

A component of the NWSRFS, called the mean areal precipitation (MAP) program (4), objectively transforms hourly and daily point precipitation measurements into an areal mean. In addition, MAP estimates missing data based on nearby precipitation records and time distributes daily data based on hourly patterns over time.

Three weighting techniques are available for distributing precipitation gauge information throughout an area. The relative importance of each precipitation gauge in an area can be determined as the sum of $1/d^2$, where d is the distance from each point in the area to the gauge. Normalizing the sums of all precipitation gauges yields the proportion (or weight) of the MAP contributed by each gauge. In certain locations (especially in mountainous regions), the distribution of precipitation gauges throughout an area may not accurately reflect the actual pattern of precipitation. Under these circumstances, MAP can accept predetermined weighting factors to distribute the point measurements. Use of predetermined weights allows the modeler to analyze any circumstances peculiar to an area and adjust the precipitation gauge weights accordingly. A third technique of determining the distribution of precipitation is by Thiessen weight factors. In this method, the area is divided such that each precipitation gauge is centered in a region containing all points closer to it than to any other gauge. The proportion of the total area contained in each region specifies the weight assigned to the associated precipitation records.

The point precipitation weightings described above all rely on complete precipitation gauge records to successfully compute the MAP. To utilize information available in partial records, a technique for estimating missing data has been developed. The crux of this method relies upon determining the importance of nearby precipitation gauge measurements for the estimation of data for the incomplete station record. The importance of nearby stations is computed as $1/D^2$, where D is the distance from the station with incomplete records. With this information, available precipitation measurements can be distributed to complete partial records.

The precipitation patterns exhibited at hourly recording stations are used to

distribute the depths of precipitation at daily stations. Again, a $1/D^2$ weighting factor is used to determine the relative effects of nearby hourly stations on the daily record.

However, this technique will not result in estimations greater than the largest known value or less than the smallest. In mountainous regions, it may be desired to alter the estimated value at a station because of known orographic effects. The MAP program accounts for such modifications when completing a precipitation gauge record by applying "characteristic station adjustments." A ratio of the characteristics of the estimating and estimated stations relates the known depths of precipitation to the depth at the incomplete station. For example, setting the characteristic of the station being estimated equal to two doubles the depth of precipitation recorded at an estimating station that will be used by the MAP program to estimate the unknown value. If no specification is made, the characteristics of all stations are assumed equal.

The MAP program produces an estimate of the depth of water falling on an area, which is used as input to the snow accumulation and ablation model and the soil moisture accounting program.

MEAN AREAL TEMPERATURE

The snow accumulation and ablation model uses air temperature as an index to heat exchange processes. The air temperature used in the model is the mean temperature over the area being simulated. Since air temperatures are measured at discrete points, it is desirable to transform point temperature data to mean areal values before the data are used by the snow model. Also, the mean areal temperatures (MAT) must conform to the computational interval used by NWSRFS. The NWSRFS MAT (6) transforms observed minimum and maximum daily temperatures into 6-hr mean areal temperatures.

The computation of MAT involves inferences regarding the temperature at all points within the area. Available observed temperature records at points within and surrounding the area are used to compute the MAT. When portions of the observed records are missing (because of equipment malfunction, missed observations and so on), the MAT program estimates missing data using records available at surrounding points. This avoids discarding observed records that are not continuous and losing the information contained in the partial record.

Two separate estimation algorithms are used. One procedure is used in non-mountainous areas, where temperature can be assumed to vary linearly with distance. The estimated temperature is a weighted average (1/D) of surrounding temperatures. The second procedure is used in mountainous areas, where temperature variation between points generally does not vary linearly with distance, but where temperature variation is primarily influenced by elevation differences. A quasi-objective technique is used that employs a weighting procedure of distance and elevation to estimate the unknown temperatures.

Once all maximum and minimum daily temperature time series have been completed, the 6-hr MAT can be computed. The first step is to convert the point minimum and maximum temperatures to point 6-hr temperatures. Four equations, one for each 6-hr period of the day, are used to convert the minimum and maximum temperatures to 6-hr temperatures. The 6-hr MAT is then calculated as the weighted average of the point 6-hr temperatures.

EVAPOTRANSPIRATION

The determination of the volume of water removed from a basin through evaporation and transpiration is important in accurately predicting the amount of water available for runoff. The mean areal potential evapotranspiration (MAPE) program computes areal values of potential evapotranspiration. As with the MAT program, the weighted average of a number of point measurements are computed over an area. The technique for distributing point potential evaporation values throughout an area is analogous to that for distributing point temperature values in nonmountainous areas. This technique gives a MAPE value for the total area.

However, evapotranspiration does not occur at the potential (maximum) level at all times. An adjustment must be made to reduce the potential evaporation value to potential basin evapotranspiration. This reduction accounts for such factors as watershed albedo and vegetative cover. The MAPE program uses a set of 12 values corresponding to the area-wide potential evapotranspiration demand for day 16 of each month. A linear interpolation between these 12 values yields an adjustment for each day of the year.

MODEL CALIBRATION

To use the soil moisture accounting, snow accumulation and ablation, and channel routing models for river forecasting, model parameters for each river basin must be estimated. Both trial-and-error methods and automatic methods are in use. Trial-and-error methods involve subjective adjustments to parameters, based on specific characteristics of previous model output. Automatic techniques involve the use of direct-search optimization algorithms for the catchment model and an interactive gradient adjustment procedure for the dynamic routing model.

Catchment Calibration

To achieve efficiently a satisfactory set of model parameters by trial and error calibration, two elements are required from the hydrologist: he must understand the physical processes taking place in the catchment and the mathematical modeling of the catchment. The hydrologist compares simulated and observed hydrographs and manually adjusts model parameters based on knowledge of the model mathematics and the physical processes of the natural catchment. However, even an experienced modeler may find the trial-and-error method time consuming. To shorten calibration time, an automatic calibration technique is available.

Experience has shown that calibration efficiency is enhanced considerably by accurate initial estimates of model parameters. Much effort has been directed toward identifying ways to determine initial parameter estimates from an observed hydrologic and geologic data base (5,13). Physically realistic parameter values will improve the representation of catchment response.

The automatic catchment model calibration technique is a direct-search optimization technique known as "pattern search" (14). The concept of this strategy is to increase the size of parameter adjustments at each stage of optimization if a persistent direction (pattern) of adjustments has been established. The success of improving model performance by parameter adjustment is measured by the sum of the squared differences (errors) between the simulated and observed daily flow.

The recommended calibration procedure includes three stages. The initial stage incorporates the experience of the hydrologist with trial-and-error calibration to test initial parameter estimates and to reveal any gross errors in the data. After reasonable parameter estimates have been obtained, intermediate calibration can proceed using pattern-search optimization to further refine parameter values.

In the final stage of calibration, the hydrologist reviews observed and simulated hydrographs for the entire period of record. If bias is absent at low, medium, and high flows, the calibration is considered complete. If, however, bias is present, additional parameter adjustment is necessary.

Dynamic Routing Calibration

The calibration of the dynamic routing model is accomplished primarily through adjustment of the channel roughness coefficients. Channel roughness is assumed constant throughout specified river reaches; however, it is allowed to vary with discharge.

Manual Calibration

The manual calibration technique uses observed stages and discharge throughout the river system as a measure of the model's accuracy. Boundary conditions (upstream and downstream discharge and stage hydrographs) are input to the model, and computed stages and discharges at internal (test) points are compared with observed values. Roughness coefficients are adjusted and the simulation is repeated.

Altering a roughness coefficient affects stages and discharges throughout the river system, but the greatest effect is immediately upstream of the altered reach (7). The manual calibration technique begins with the upstream reach and adjusts the roughness to match computed and observed stages at the upstream test station. The calibration proceeds downstream, matching computed and observed values at each test station in a sequential manner. In general, the reaches with constant roughness are established so that one test point falls within each reach. The manual calibration method requires numerous submissions of the routing system, and only a few adjustments are made each time.

Automatic Calibration

The automatic calibration procedure computes a set of roughness coefficients by calibrating the river system one reach at a time. With this technique, roughness reaches are established so that test stations are at both ends of the reach. Discharge is input at the upstream boundary, while stage is specified downstream. Observed stages at the upstream boundary are tested against computed stages at that point. Statistics are computed for several ranges of discharge so that the roughness coefficient can be calibrated as a function of discharge. For each range of discharge, the adjustment procedure uses the root-mean-square (RMS) error to determine whether the required change should be positive or negative. Adjustments are automatically made to the roughness coefficients for the reach and the one-reach system is rerun. The cycle is repeated until a minimum RMS error for the reach is found. The discharge computed at the downstream boundary using the coefficients associated with the minimum RMS error are input as upstream boundary conditions for the next reach. A compatible set of roughness coefficients which minimize RMS errors throughout the river system is determined by rationally proceeding through the river system one reach at a time. The automatic calibration procedure makes efficient use of time of both man and machine.

OPERATIONAL FORECAST PROGRAMS

Catchment Model

Once a catchment is calibrated, the conceptual models mathematically represent the important hydrologic processes of the catchment. The hydrologic parameters derived in the calibration phase are transferred to the operational forecast programs. In terms of hydrologic computation, there are no differences between calibration and operational programs. The only differences lie in the timeframe of catchment

simulation and in the number of catchments simulated per computer run. In calibration, interest is in the reproduction of a long series (5-15 yrs) of historical data; however, operationally the timespan is reduced to forecasting catchment response a few days or weeks into the future. During calibration, no more than three adjacent catchments are simulated during one computer run. However, the operational programs are designed to interact with an operational direct access disk file system to retrieve and store information necessary to simulate entire river systems (which may include as many as 600 local catchments) on a continuing real-time basis.

The operational forecast programs in the NWSRFS have been organized into the three modules shown in Fig. 18.6: a data management module, a pre-processor module, and a

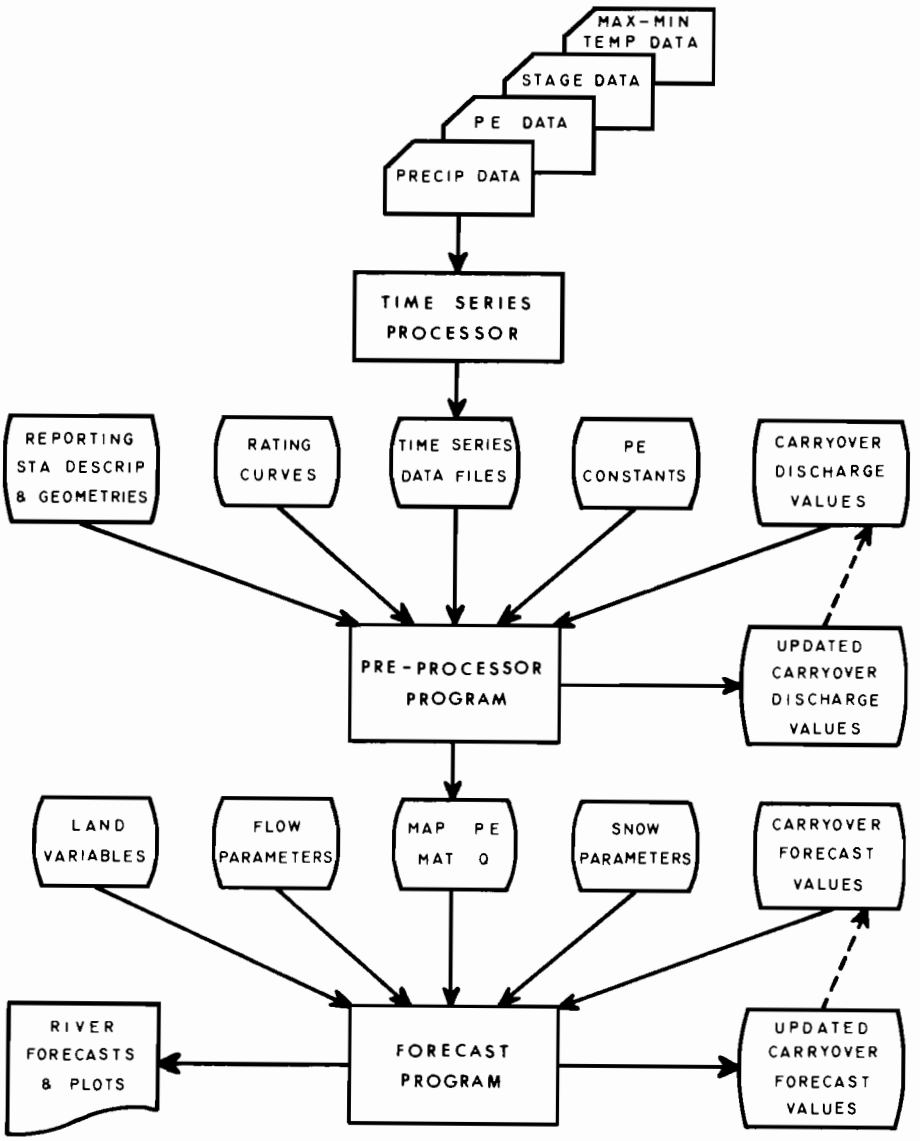


Fig. 18.6 Operational catchment response model. PE stands for potential evapotranspiration, STA for station, and Q for discharge.

forecast module. The data management module is the only interface between the user and the forecast system. The data management module allows the user to enter time series data (i.e., precipitation, temperature, stage/discharge, and potential evaporation data values); enter model parameter data; and display/print time series data, parameters, and forecast output. The pre-processor module uses the time series data and corresponding model parameters to compute mean areal precipitation, potential evapotranspiration, stream flows, and mean areal temperatures. The forecast module uses the values from the pre-processor module along the model state variables carried over from previous computational periods to compute conditions at specified forecast points.

Dynamic Wave Model

The core of the operational dynamic routing program is the dynamic routing basic element described above. The modifications to that basic element of the NWSRFS consist primarily of an expanded data management package (Fig. 18.7). A large portion of the information required to simulate a river system does not change daily. The cross-sectional data, roughness coefficients, and information

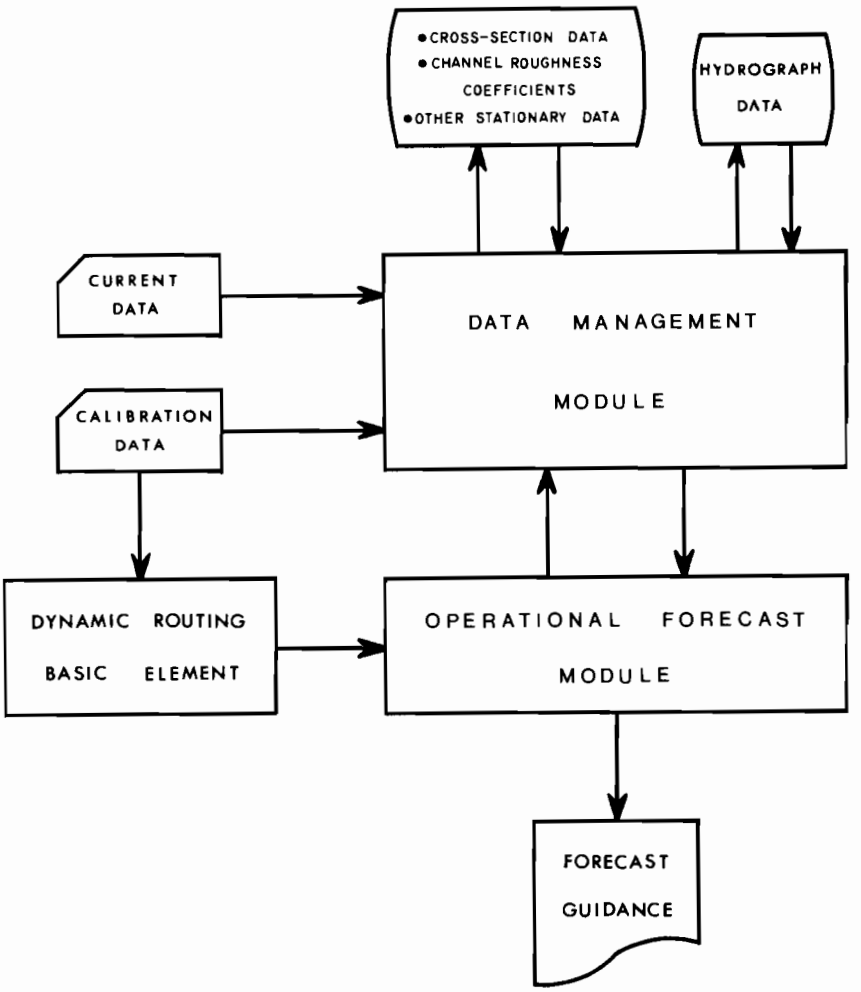


Fig. 18.7 Operational dynamic routing model.

specifying the routing configuration will remain constant for long periods of time. It is only the hydrographs that must be updated for daily operational use. To manage these data requirements efficiently, a package of subroutines to store and retrieve the unchanging portion of the input data from disks and to update the hydrographs has been developed.

By specifying which river system is to be simulated and the period of simulation, the data are automatically prepared for use by the dynamic routing basic element. A minimum of data handling is required, as only the latest values needed to update the hydrographs must be prepared.

A feature included in the operational dynamic routing program, which will improve forecasts, is the ability to bring the entire system (all computed stages and discharges) up to date with the most recent observed stages before proceeding into the future. Since all the information available about the conditions in the river is contained in the stages and discharges, by updating the entire system to present observed stages, the startup errors in a forecast of any given length can be minimized.

Extended Streamflow Prediction

Water supply forecasting is another function performed by the NWSRFS. An operational element capable of extended stream-flow prediction (ESP) provides the forecast.

The ESP program makes multiple simulations with the catchment model using the current hydrologic conditions and precipitation and temperature data representing periods from numerous years. The stream flows obtained from these simulations are analyzed to provide a frequency distribution (and thereby a probability distribution) of any flow level. The probability distribution relates any flow to a specified chance of occurrence. The time span to which a probability distribution applies, as well as the date at which the span begins, can be input into the ESP program.

With this information, the expected flow at any selected probability level can be obtained for the time period chosen. Forecast information for peak, low, and mean flow levels, as well as volume of flow, is available.

COMPUTER FACILITIES

The computational elements of the NWSRFS are being implemented on a computer facility at Suitland, Maryland. The operating system consists of three IBM 360/195 computers with a capacity of 64 disk drives. (Trade names are included for identification purposes only. No endorsement by the U.S. Department of Commerce, the National Oceanic and Atmospheric Administration, or the National Weather Service is implied.)

Each RFC will access the computing system through remote terminals with batch processing capabilities. Input/output can be accomplished by means of punched card, paper tape, printer or magnetic tape mass storage. Limited off-line processing capabilities will also be available. The remote terminal serving the HRL will perform batch processing, as well as allow interactive access with the operating system.

EXPERIENCE WITH THE NATIONAL WEATHER SERVICE RIVER FORECAST SYSTEM

Since 1971, the Lower Mississippi RFC at Slidell, Louisiana, has been using the initial version of NWSRFS for operational river forecasting. This RFC has forecast

responsibilities for the lower portion of the Mississippi River system (see Fig. 18.2) and some adjacent systems draining to the Gulf of Mexico. Nearly 200 catchments have been calibrated by the Lower Mississippi RFC and most are forecast on a daily basis.

The use of the NWSRFS at Slidell has proved to be especially effective in forecasting extreme hydrologic events. For example, in March 1973, a major flood on the Amite River at Denham Springs, Louisiana, crested at an elevation of 9.93 m (flood of record is 10.8 m). NWSRFS was used to forecast a flood crest elevation of 9.90 m. Prior to the storm, the Amite River was flowing at an elevation of 3.05 m. The time to peak of the flood hydrograph was 72 hrs and the NWSRFS crest forecast was released approximately 30 hrs before the crest occurred.

There was a similar situation on the Leaf River at Hattiesburg, Mississippi, in December 1973. A total 2-day rainfall in excess of 150 mm produced a major flood on the Leaf River. In 48 hrs, the river rose from an initial elevation of 2.13 m to the crest elevation of 8.25 m. Nearly 36 hrs before the crest was observed at Hattiesburg, a crest forecast of 8.07 m was generated by the NWSRFS. These two examples are a good indication of the potential of a conceptual model with a strong physical base to simulate catchment response.

The experiments with the NWSRFS by the Lower Mississippi RFS have been instrumental in making many of the operational modifications and improvements that are included in the current version of NWSRFS. As remote job processing terminals connected to the IBM 360/195 in Suitland, Maryland, are installed, the remaining RFCs will have the computing capacity to implement NWSRFS. It is expected that in the next 5 to 10 yrs the NWSRFS will be implemented nationwide.

In addition to the operational application of NWSRFS by the Lower Mississippi RFC, several other RFCs, as well as the HRL, have calibrated a wide variety of catchments. These catchments represent most of the hydrologic conditions found in the United States.

REFERENCES

1. Sittner, W.T. (1973). Modernization of National Weather Service river forecasting techniques. Water Res. Bull. 9(4): 655-659.

2. Monro, J.C., and E.A. Anderson (1974). National Weather Service River Forecasting Service. ASCE J. Hydraul. Div. 100 (HY5): 621-630.

3. Crawford, N.H., and R.K. Linsley (1966). Digital simulation in hydrology: Stanford Watershed Model IV. Dept. of Civil Eng. Tech. Report 39, Stanford University, Stanford, California.

4. Staff, Hydrologic Research Laboratory (1973). National Weather Service River Forecast System, forecast procedures. NOAA Tech. Memo. NWS HYDRO-14, U.S. Dept. of Commerce, Silver Spring, Maryland.

5. Burnach, R.J.C., R.L. Ferral, and R.A. McGuire (1973). A generalized streamflow simulation system: Conceptual modeling for digital computers. Joint Federal-State River Forecast Center, Sacramento, California.

6. Anderson, E.A. (1973). National Weather Service River Forecast System, snow accumulation and ablation model. NOAA Tech. Memo. HYDRO-17, U.S. Dept. of Commerce, Silver Spring, Maryland.

7. Fread, D.L. (1974). Implicit dynamic routing of floods and surges in the Lower

Mississippi. Presented at AGU National Meeting, Washington, D.C.

8. Anderson, E.A. (1968). Development and testing of snow pack energy balance equations. Water Resour. Res. 4(1): 19-37.

9. Anderson, E.A. (1976). A point energy and mass balance model of a snow cover. NOAA Tech. Report NWS-19, U.S. Dept. of Commerce, Silver Spring, Maryland.

10. Larson, L.W., and E.L. Peck (1974). Accuracy of precipitation measurements for hydrologic modeling. Water Resour. Res. 10(4): 857-863.

11. Linsley, R.K., M.A. Kohler, and J.L.H. Paulhus (1958). Hydrology for Engineers. McGraw-Hill, New York.

12. Fread, D.L. (1974). Numerical properties of implicit four-point finite difference equations of unsteady flow. NOAA Tech. Memo. NWS HYDRO-18, U.S. Dept. of Commerce, Silver Spring, Maryland.

13. Peck, E.L. Catchment modeling and initial parameter estimation for the National Weather Service River Forecast System. NOAA Tech. Memo. NWS HYDRO-31, U.S. Dept. of Commerce, Silver Spring, Maryland.

14. Monro, J.C. (1971). Direct search optimization in mathematical modeling and a watershed model application. NOAA Tech. Memo. NWS HYDRO-12, U.S. Dept. of Commerce, Silver Spring, Maryland.

15. (1975) Interim Report: Operational forecast programs and data components for the National Weather Service River Forecast System (NWSRFS): Data files design. Prepared for the National Weather Service under Contract No. 5-35422, by Data Sciences, Inc., Silver Spring, Maryland.

Index

List of Contributors

M.B. Beck
Control and Management Systems Division
University Engineering Department
Cambridge University
Cambridge, England

E. Bobinski
Institute for Meteorology and Water Management
Warsaw, Poland

R.L. Bras
Department of Civil Engineering
Massachusetts Institute of Technology
Cambridge, Massachusetts

J. Cole
Water Research Center
Medmenham, England

D.C. Curtis
Hydrologic Research Laboratory
National Weather Service
Silver Spring, Maryland

T. Ganendra
Department of Civil Engineering
Imperial College of Science and Technology
London, England

S. Ikeda
Department of Applied Mathematics and Physics
Kyoto University
Kyoto, Japan

T. Katayama
Department of Applied Mathematics and Physics
Kyoto University
Kyoto, Japan

T. Kawaguchi
Sumitomo Steel Industries, Ltd.
Osaka, Japan

D.B. McLaughlin
Resources Management Associates
Lafayette, California

D. Maidment
Water and Soil Division, MWD
Christchurch, New Zealand

R. Mehra
Scientific Systems, Inc.
Cambridge, Massachusetts

R.J. Moore
Institute of Hydrology
Wallingford, England

A. Murphy
Department of Atmospheric Sciences
Oregon State University
Corvallis, Oregon

G. Olsson
Department of Automatic Control
Lund Institute of Technology
Lund, Sweden

T. Piwecki
Institute of Hydraulic Engineering
Technical University
Gdansk, Poland

Y. Sawaragi
Department of Applied Mathematics and Physics
Kyoto University
Kyoto, Japan

G.F. Smith
Hydrologic Research Laboratory
National Weather Service
Silver Spring, Maryland

H. Stehfest
Kernforschungszentrum
Karlsruhe, FRG

A. Szöllösi-Nagy
Research Institute for Water Resources Development
Budapest, Hungary

H. Tamura
Department of Precision Engineering
Osaka University
Osaka, Japan

G. Weiss
Department of Statistics
Tel Aviv University
Tel Aviv, Israel

E.F. Wood
Department of Civil Engineering
Princeton University
Princeton, New Jersey

J. Zelanzinski
Institute of Hydraulic Engineering
Technical University
Gdansk, Poland

REAL-TIME FORECASTING/CONTROL OF WATER RESOURCE SYSTEMS
Selected Papers from an IIASA Workshop,
October 18–21, 1976
Eric F. Wood, Editor,
with the assistance of András Szöllösi-Nagy

When water resource systems are not under control, the consequences can be devastating. In the United States alone, flood damage costs approximately $1.5 billion annually. These losses can be avoided by building more reservoirs to hold the flood waters, but such construction is very expensive, especially because reservoirs have already been built on the best sites. A better and less expensive alternative is the development of more effective management methods for existing water resource systems, which commonly waste approximately 20 percent of their capacities through mismanagement.

This book contains selected papers from a workshop devoted to the consolidation of international research on statistically estimated models for real-time forecasting and control of water resource systems. The workshop was sponsored by the International Institute for Applied Systems Analysis and contributed to the research on the methodology of real-time forecasting and control of water resource systems being carried on in the Water Project of the Institute.

Statistical models first appeared in hydrology at the beginning of the 1970s. Hydrologists began to use the techniques of time series analysis and system identification in their models, which seemed to give better results than the earlier, deterministic simulation models. In addition, real-time control of water resources was being developed at the practical level and on-line measurements of rainfall and runoff from a catchment were becoming available. The conceptual models then in use could not take advantage of measurements from the catchment, but on-line measurements now allow an operator to anticipate flood waters upstream or a water shortage downstream.

The book is divided into three parts. The first part presents several methods of forecasting for water resource systems: distributed lag models, maximum likelihood identification, nonlinear catchment models, Kalman filtering, and self-tuning predictors. The papers in the second part present methods for controlling stream quality and stream flow, and the third part describes forecasting in the United States, the United Kingdom, and Poland.

0 08 024486 6