IMPERIAL COLLEGE OF SCIENCE AND TECHNOLOGY

DEPARTMENT OF CIVIL ENGINEERING

REAL-TIME FORECASTING AND CONTROL IN THE OPERATION OF WATER RESOURCE SYSTEMS

by

Torkil Ganendra, B.Sc., M.Sc., D.I.C.

Thesis submitted to the University of London in partial fulfilment of the requirements for the Degree of Doctor of Philosophy in the Faculty of Engineering

June 1979

"But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and the highest aim of all computations made concerning phenomena must approximate, as nearly as practicable, to the truth".

C.F. Gauss, 1809

REAL-TIME FORECASTING AND CONTROL IN THE OPERATION OF WATER RESOURCE SYSTEMS

by

Torkil Ganendra

ABSTRACT

Problems associated with the real-time management of a complex water resource system are discussed with particular reference to real-time river regulation. A real-time control strategy is suggested in which the water resource system is sub-divided into a number of subsystems and one regional centre. This could correspond to a computer network comprising subsystem mini/micro computers linked to a regional computer centre. Each subsystem can then be individually controlled in real-time while the global system operation is determined by medium-term policies which are developed within the regional centre.

Subsystem control is dependent on models for forecasting hydrological variables. The self-tuning regulator, designed to control processes with unknown parameters, forms the basis of the self-tuning predictor and the self-tuning controller, which are applied to forecasting and control respectively in this study. The self-tuning predictor and self-tuning controller assume that the system being modelled can be represented by a linear ARMAX model. The parameters of the predictor and controller are estimated using recursive least-squares.

The self-tuning predictor is applied to the Brosna catchment in Ireland and the Hirnant sub-catchment of the River Dee in North Wales. The results obtained are good for lead-times not greater than the pure time delay between the rainfall and runoff response. The predictor parameters are also shown to converge rapidly. The self-tuning predictor is also used with an auxialiary 'perlog' function input. The resulting non-linear model takes into account antecedent catchment conditions. The versatility of the ARMAX model is also shown by an example in which stage is forecasted using three self-tuning predictors with stage, flow rate and cross-sectional area of flow as the forecast variables respectively.

A self-tuning controller suitable for real-time river regulation is derived and applied to three simulated subsystems. The controls are applied to reservoir releases and groundwater abstraction discharge rates for low flow augmentation. It is shown that by controlling the reservoir releases or the groundwater abstraction rates, a substantial reduction in excess compensation flow may be achieved.

CONTENTS

Page

ACKNOWLEDGEMENTS

Acknowledgements are due to Dr Enda O'Connell at the Institute of Hydrology, Wallingford, and Elizabeth Shaw at Imperial College, for their constant and unselfish support and the many hours they have devoted towards the preparation of this thesis. Acknowledgements are also due to the late Professor J.R.D. Francis and to Dr R. Perry for the use of the facilities at Imperial College.

The author also wishes to thank the numerous individuals who participated in fruitful discussions, and in particular Dr D.W. Clark and Dr P.J. Gawthrop of the Engineering Science Department, University of Oxford, Dr M.H.A. Davis of the Computing and Control Department, Imperial College and many of the staff at the Institute of Hydrology.

Finally the author is grateful to his colleagues at Imperial College for their help and encouragement and in particular to Paul Johnston for his assistance and advice in the preparation of this thesis, and to Judith Barritt for typing the manuscript.

NOTATION

A list of the principle notation used in the text is given below. For the convenience of the reader, a bracketed figure indicates the equation in which a particular notation has first been used.

$A(z^{-1})$	polynomial as defined by eqn. 3.22
$\mathcal{A}(z^{-1})$	polynomial as defined by eqn. 5.13
$A'(z^{-1})$	polynomial as defined by eqn. 3.38
a_i	coefficient of $A(z^{-1})$ (3.15)
a_i'	coefficient of $A'(z^{-1})$ (3.38)
a	model parameter (4.8)
$\bar{a}$	attenuation parameter (6.55)
AML	approximate maximum likelihood
AR	autoregressive
ARMA	autoregressive moving average
ARMAX	autoregressive moving average with exogenous inputs
$B(z^{-1})$	polynomial as defined by eqn. 3.22
$\mathcal{B}(z^{-1})$	polynomial as defined by eqn. 5.13
$B'(z^{-1})$	polynomial as defined by eqn. 3.38
b	delay parameter
b_i	coefficient of $B(z^{-1})$ (3.15)
b_i'	coefficient of $B'(z^{-1})$ (3.38)
b	model parameter (4.8)
b(t)	baseflow (6.53)
$C(z^{-1})$	polynomial as defined by eqn. 3.22
$\mathcal{C}(z^{-1})$	polynomial as defined by eqn. 5.13
$C'(z^{-1})$	polynomial in h(t) (5.39)
c_i	coefficient of $C(z^{-1})$ (3.20)

$\overline{C}(Q)$	average wave speed along the reach (6.55)
$E(z^{-1})$	polynomial as defined by eqn. 6.25
$E\{.\}$	expectation operator (3.3)
$e(t)$	Gaussian white noise sequence distributed as $N(0,\sigma^2)$ (3.20)
e_i	parameter of $E(z^{-1})$ (6.27)
$e_u(t)$	prewhitened u(t) series (3.56)
$e_r(t)$	prewhitened r(t) series (5.30)
$F(z^{-1})$	polynomial as defined by eqns. 4.5
f	a constant (4.11)
f_i	coefficient of $F(z^{-1})$ (4.5)
$G(z^{-1})$	polynomial as defined by eqns. 4.5
G_t	cumulative net gain (6.53)
g_i	coefficient of $G(z^{-1})$ (4.5)
g_a	parameter corresponding to variable qa(t) (6.56)
g_c	parameter corresponding to variable qc(t) (6.56)
g_h	parameter corresponding to variable qh(t) (6.56)
H	Kalman filter measurement matrix (3.2)
H_{res}	reservoir level (6.46)
h	water table elevation (6.50)
h_i	rainfall parameter (6.45)
h(i)	ordinate of the unit hydrograph (3.16)
h(t)	stage (5.38)
I	identity matrix
IV	instrumental variables
J	quadratic loss function defined by eqn. 3.3
K	number of sample autocorrelation coefficients (3.70)
K(t)	Kalman gain matrix (3.7)
k	pure time delay (4.1)
$L(\varepsilon\|\theta)$	log likelihood function defined by eqn. 3.45

ℓ	order of polynomial in the input term (4.14)
MA	moving average
m	order of polynomial in the output term (4.14)
m	number of ordinates of the unit hydrograph (3.16)
N	sample size (3.24)
$N(0,\sigma^2)$	Normal distribution with zero mean and variance σ^2
n	order of polynomial
P_a	pumping rate - wells 1, 5 and 8 (6.45)
P_b	pumping rate - wells 3, 4, 6, 7, 9 and 10 (6.45)
P(t)	error covariance matrix (3.30)
$P(t+1\|t)$	predicted error covariance of the Kalman filter state estimates (3.6)
$p\ell(t)$	perlog (5.34)
Q	value used in the Q-test (3.70)
Q	flow rate from aquifer (6.50)
Q(t)	reservoir outflow (6.47)
Q_i	lateral inflow per unit length
q(t)	streamflow (output) (3.15)
q'(t)	storm runoff (3.16)
q_{min}	minimum constraint on prediction $\hat{q}(t+k\|t)$
q_{max}	maximum constraint on prediction $\hat{q}(t+k\|t)$
q_{in}	reservoir inflow (6.48)
qc(t)	River Ceiriog flow
qh(t)	River Hirnant flow
qa(t)	River Alwen flow
r(t)	rainfall input (3.15)
r'(t)	excess precipitation (3.16)

$r_{xx}(\tau)$	lag τ sample autocorrelation function for sequence x(t)
$r_{xy}(\tau)$	lag τ sample cross-correlation function between x(t) and y(t)
S	value used in the S-test (3.73)
s	specific yield (6.50)
T	transmissivity (6.50)
t	time-step
U(t)	vector of input variables (3.1)
u(t)	system input (3.20)
u(t)	control variable (4.1)
u_{max}	maximum constraint on u(t) (6.36)
u_{min}	minimum constraint on u(t) (6.36)
V	quadratic loss function defined by eqn. 3.25
V_1	quadratic loss function defined by eqn. 4.3
V_2	quadratic loss function defined by eqn. 6.10
V_3	quadratic loss function defined by eqn. 6.11
V(t)	vector of white noise variables distributed as N(0,R) (3.2)
V_{LS}	quadratic loss function given by eqn. 4.10
V_{res}	reservoir volume (6.46)
VPD	variable parameter diffusion
v_i	impulse response weights (3.54)
W(t)	vector of Gaussian white-noise variables distributed as N(0,Q) (3.1)
w(t)	set-point (6.4)
X(t)	Kalman filter state vector (3.1)
$\tilde{X}(t\|t)$	state estimation error (3.4)
x	horizontal space coordinate (6.50)
x(t)	cross-sectional area of river flow (5.40)
y	vertical space coordinate (6.50)
y(t)	system output (3.20)

$\hat{y}(t\|t-k)$	prediction of y(t) using information up to time t-k (3.26)
$y'(t)$	transformed output series (3.65)
$y_r(t)$	output error defined by eqn. 6.4
$\hat{y}_{IV}(t)$	predicted output given by the IV method (3.39)
$Z(t)$	vector of measurements (3.2)
z^{-1}	backward shift operator (3.21)
$Z(t)$	vector of instrumental variables (3.34)
$z(t)$	instrumental variable (3.36)
α	forgetting factor (5.22)
$\alpha(t)$	time varying forgetting factor (5.23)
α_o	coefficient for rate of change of $\alpha(t)$ (5.23)
$\alpha_{\tau\tau}$	sample partial autocorrelation function (3.59)
$\alpha_{i(i\neq 0)}$	coefficients of polynomial $A(z^{-1})$ (5.13)
β_i	coefficients of polynomial $B(z^{-1})$ (5.13)
Γ	Kalman filter control transition matrix (3.1)
γ_i	coefficients of polynomial $C(z^{-1})$ (5.13)
$\gamma_{yy}(\tau)$	lag τ autocovariance function for sequence y(t) (3.48)
$\gamma_{yu}(\tau)$	lag τ cross-covariance function between y(t) and u(t) (3.49)
$\rho_{yy}(\tau)$	lag τ autocorrelation function for sequence y(t) (3.50)
$\rho_{yu}(\tau)$	lag τ cross-covariance function between y(t) and u(t) (3.51)
∇	backward difference operator (3.52)
$\varepsilon(t)$	residual error between observed and predicted (3.25)
θ	parameter vector (3.24)
$\hat{\theta}(t)$	estimate of θ at time t
$\phi(t)$	data vector (3.27)
Φ	Kalman filter state transition matrix (3.1)
σ	standard deviation (3.45)
$\nu(t)$	innovation sequence of Kalman filter (3.15)

$\xi(t)$	a noise process (3.15)
$\psi(t)$	output function defined by eqn. 6.20
$\hat{\psi}(t+k\|t)$	prediction of output function (6.21)
λ^*	loss function weight for V_1 and V_2 (6.10)
λ	weight λ^*/b
ζ	a noise term (3.54)
μ	scalar multiplied by I-matrix to give P(0)
μ_x	mean of x (3.48)
ω	control law parameter (4.9)
χ^2_5	Chi-square at the 5 percent point
$\hat{\cdot}$	estimated

1. INTRODUCTION

1.1 TRENDS IN WATER RESOURCE SYSTEM MANAGEMENT

Water resources have always been a major factor in the economic development of a region. The extent of urbanization and industrial and agricultural development has been dependent upon reliable water supplies. Rapid population and economic growth, together with the more intensive use of water resources have made it necessary to develop water resource systems to high levels of complexity and sophistication. Perhaps the most important new trend in water resource management has been the gradual broadening of its scope. Several decades ago water resource management consisted of purely engineering solutions to water supply problems as they emerged in connection with particular engineering projects such as irrigation or hydropower schemes (this is still often the case). Only with the increase in the number and size of projects associated with large water resource systems has it been necessary to consider the interaction between the different projects, and the extent to which multiple and often conflicting objectives can be met.

Although water is normally 'used' in one small geographic area, each use has its interactive effects on other users in the area and on other areas. With the increasing intensive use of water resources, and with environmental concerns often in conflict with economic opportunities, a policy of un-coordinated, ad hoc development is not appropriate even in the areas of abundant water. Consequently there has been a gradual shift in water resource management from engineering technology alone to a broader technology incorporating planning.

Today water resource planning is a multi-disciplinary field involving the interaction between the social sciences (e.g. economics), applied hydrology and water quality engineering. Applied hydrology is concerned with the quantitative evaluation of the spatial and temporal behaviour of water within the hydrosphere. Hydrological

data provide a measure of the available water resources and are fundamental to the operation of any water resource system. Water quality engineering is connected with water supply treatment and distribution, wastewater treatment and disposal and environmental conservation activities. The need for careful planning and efficient operation or management of the broad constituents of a water industry has led to the demand for mathematical techniques which can be applied to aid decision-making. Systems analysis backed by the extensive use of digital computers has been the answer to this challenge. Systems analysis has introduced to water resource management modern mathematical concepts such as decision theory, optimization, mathematical modelling, control theory, statistics and probability theory, and simulation techniques.

The implementation of a water resource system involves three stages, namely the planning and design procedures, the construction and commissioning of civil works, and the operational management of the system. The planning and design stage is often characterised by the following tasks which must be carried out:

(i) estimation of water quality and quantity requirements;

(ii) the location of the project(s) in relation to the demand centre(s) and the selection of potential storage sites;

(iii) yield capacity assessment and sizing of project components based on demand and hydrological inputs;

(iv) minimum cost strategy for the timing of project implementation to satisfy increasing water quantity and quality demands throughout the planning horizon;

(v) an operational management framework which is based on demand for, and availability of the water resources;

(vi) detailed engineering design of civil works, wastewater treatment plants, instrumentation networks etc.

The result of the planning and design phase is a set of structural and non-structural (i.e. management objectives, legislation) measures

which establish the water resource system, and allow for the development of resources to meet new demands in the long term. The overall objective is to provide an effective solution to resource system expansion in order to achieve a maximum net benefit (difference between costs incurred and benefits generated).

When the construction and commissioning of the various components of a water resource system have been completed, the system manager is faced with the task of operating the system so that the planning and design objectives are fulfilled. For example, a frequent problem is the operation of a dual purpose reservoir where the objectives are both those of water supply and flood alleviation. To ensure an adequate water supply it is in a broad sense desirable that the volume of impounded water be maximised. This is in conflict with flood protection which requires a certain proportion of the reservoir to be empty as storage available for flood alleviation. The reservoir management problem is usually resolved by some form of simulation study at the planning and design stage which results in a reservoir rule curve or decision table. This specifies a policy of desired reservoir levels for different times of the year and it is the task of the system manager to ensure that the reservoir is operated within this policy.

The water resource system must then be operated within the constraints of the policies derived in (v) above. These policies are usually based on a monthly, seasonal or annual time scale and will be referred to as medium-term strategies. In order to operate the system within the constraints laid down by the medium-term strategy, a short-term management strategy is required which is based on a time scale ranging from less than one hour to several days. Short-term management measures could for example be the daily scheduling of reservoir releases. Thus there are two levels of time dependency, namely short and medium-term which can be associated with the operational management of a water resource system.

The subject of this thesis is the short-term or 'real-time' operation

of a water resource system which is dependent on the availability of forecasts of hydrological variables such as rainfall or streamflow. Hydrological models are required to provide these forecasts which largely govern the outcome of short-term management decisions. It is interesting to note that the limitations of hydrological, water quality and water supply demand forecasts separates the short-term and medium-term management of the water resource system. For while forecasting in the short-term is concerned with quantity and time of occurrence, medium-term forecasting is concerned with the forecasting of quantity and frequency of occurrence. In short-term forecasting, if at some time t a forecast is made of the forecast variable at time $t+\Delta t$, then the time increment Δt is referred to as the lead-time. It is more difficult to forecast as the lead-time becomes greater and the physical dynamics of the hydrological system, which are dependent on unknown meteorological inputs, become increasingly nebulous. When the lead-time becomes too large for short-term forecasting and it is not possible to forecast quantity at time $t+\Delta t$, forecasts must be based on some concept of frequency of occurrence. Short-term management is thus bounded by the constraints on the lead-times associated with short-term forecasts.

In many parts of the world, flooding not only regularly claims many lives, but the loss of property is also of great economic concern. Thus, while the discussion so far has been focussed on large water resource systems, there is also an urgent need to develop low cost and effective flood warning systems , which could be operated for a small river basin.

Scope of Study

In Chapter 2 some notion of the complexity and possible configurations of the components of a water resource system are introduced. This highlights some of the problems that are usually encountered with the real-time forecasting and control of such a system. A framework for the operational management of a water resource system at the short-term and medium-term levels is then proposed.
The methodology reduces the complexity of the global system real-time management problem down to the real time control of individual

subsystems. In Chapter 3, linear filter and prediction theory is introduced. This gives a brief introduction to some recursive estimation methods which have, in the field of research hydrology, been successfully applied to real-time flow forecasting. Important aspects of system identification and model performance evaluation are also treated in this chapter. Modern control theory leading up to stochastic adaptive controllers is outlined in Chapter 4. The self-tuning regulator, on which the self-tuning predictor and the self-tuning controller are based, is then described. In Chapter 5 rainfall-runoff models in general are classified, and in particular, real-time models based on estimation theory are reviewed. The self-tuning predictor, on which the writer's work on real-time forecasting is based, is then described and applied to streamflow forecasting for the River Brosna and the River Hirnant, and the results are discussed. A modificiation of this model incorporating an auxiliary input is then developed and tested. The auxiliary input is known as the perlog function, and the resulting model is non-linear and dependent on catchment antecedent conditions. Finally, in this chapter the choice of variable to be forecast (flow rate, stage or cross-sectional area of flow) is studied. In Chapter 6 a self-tuning controller is derived and applied to problems of reservoir release control and groundwater abstraction control for real-time river regulation. Both the predictor and the controller described in Chapters 5 and 6 can be adopted as standard modules to be used within the real-time subsystem control framework developed in Chapter 2.

1.2 FORECASTING AND CONTROL OF HYDROLOGICAL VARIABLES

The broad concept of the field of water resource management and its increasing complexity has necessitated the adoption of techniques and methods of system analysis. A system has been defined by Dooge (1973) as "any structure, device, scheme, or procedure, real or abstract, that interrelates in a given time reference , an input, cause or stimulus, of matter, energy, or information, and an output, effect, or response, of information, energy, or matter". Systems analysis is thus concerned with describing the way in which a system converts one or more input(s) to one or more output(s). The structure and physical laws of the system which converts input to output will be referred to as the process. Complex systems may be broken down into subsystems, each having their own input-output characterizations. The smallest subdivision of a system or subsystem, again an input-output element, will be referred to as a component.

With the increasing application of digital computers in the management of water resources systems, there has been a widespread interest in the development of methodologies which utilise this computing capability in a real-time framework. Real-time has become a familiar term amongst engineers and scientists involved in the management of water-resource systems and is a term which has been used for many years in the computer/electronics field. It is defined as the performance of a computation during the actual time the related physical process transpires in order that results of the computation can be used in guiding the physical process (IEEE, 1972)*. Thus real-time forecasting and control refers to the short-term operational management of the water resource system.

** A large proportion of the writer's research presented herein is based on linear filter, estimation and modern control theories. Rapid developments in these fields have led to a profusion of published papers which is typified by the procedings of the 3rd IFAC (International Federation for Automatic Control) conference on 'Identification and System Parameter Estimation' which contains no less than 150 papers. With this flood of literature has come an extensive (and confusing) vocabulary of technical terminology. Thus in the presentation of this thesis particular attention will be paid to clear definition of the terminology used. The IEEE (Institute of Electrical and Electronics Engineering) standard definitions will be used as a reference where necessary (IEEE, 1972).*

The objective of control is to manipulate some of the system inputs in order to bring about some process performance, which in a management sense will result in a more effective or optimal operation of the system. A control system consists of a controller associated with one or more processes to perform specific functions according to a control law (IEEE, 1972). The controller is the part of the control system that implements the control law. It is interesting to note that control when referring to water quality in the past has implied control by more effective design of water treatment facilities or control by legislation, rather than control by the optimal performance of the water treatment facilities themselves. Clearly, if a system is to be controlled (in real-time) it is necessary that some or all of the variables which describe the system 'state' are known. The state of a system is defined in general terms as the set of variables (observable quantities or measurements) required to describe the behaviour of that system (a more rigorous definition is given in Section 3.1). The state of a water resource system is determined by variables characterising the system such as reservoir levels, pumping rates and streamflow and rainfall quantities. These variables are known as the state variables. Real-time control of a water resource system requires knowledge of the system state (from measurements), a dynamic operating policy (the control law) which is updated in response to measurements of the system state variables, and a model of the system from which forecasts of certain state variables (e.g. streamflow) can be obtained.

The design of a real-time forecasting and control water resource management system can be separated into (a) the design of monitoring and computer systems and (b) the development of mathematical models and computer techniques for use in decision making, and their arrangement within a management framework.

(a) Monitoring and Computer Systems

Monitoring of the water resource system entails the on-line and automatic collection of data so that the system state is known at the time of forecasting or controlling any of the state variables.

The computer system, apart from performing the necessary computations must monitor the system state and handle the system control variables. The design of a monitoring and computer system for real-time water resource management may involve some or all of the following aspects:

(i) the design of the raingauge network for the on-line measurement of rainfall. In the future this may be extended to alternative forms of rainfall measurement using remote sensing techniques (i.e. radar and satellite imagery);

(ii) the selection of sites for the on-line measurement of river levels and water quality variables;

(iii) the specification of a computer system which is required to handle data acquisition, forecasting and control computations, and if necessary communication with other computers;

(iv) the specification of a data communication network linking one or more computers to the measurement sensors and control hardware.

(b) Computer Techniques and Management Framework

The above steps specify some of the 'hardware' requirements for real-time management of a water resource system. To complement the hardware the following 'software' and decision-making framework for the implementation of a particular scheme may be required:

(v) the development of real-time rainfall forecasting models which may be based on mathematical modelling of atmospheric processes or statistical forecasting techniques. In the future remote sensing is likely to play an important part in providing data for these models;

(vi) the selection and/or development of mathematical models for real-time flow forecasting from rainfall;

(vii) the selection and/or development of mathematical models for real-time river flow routing;

(viii) the selection and/or development of mathematical models for forecasting one or more water quality variables;

(ix) the design of automatic controllers, which based on forecasts of hydrological variables and the demands made on the water resource system, adjust the system control variables such as sluice gate settings and pumping rates to achieve desirable system states;

(x) the formulation of a management or decision-making framework within which forecasting models and controllers are linked. This will involve the selection of some criteria for global optimality (short term) of the system as each of its components may have its own optimal solution.

(xi) the formulation of some framework within which the real-time operation of the water resource system satisfies the medium-term operation of the system.

This thesis is based mostly upon research into computer algorithms for real-time flow forecasting and real-time control ((vi) and (ix) above respectively). Real-time flow forecasting is mainly concerned with the forecasting of flood events while a real-time controller may be employed in making short-term management decisions for various operational functions of the water-resource system such as reservoir releases and groundwater abstraction rates.

The recent interest in real-time hydrological forecasting and control has been stimulated by the development and availability of streamflow and rainfall gauging networks linked directly to computers which perform automatic data acquisition. On-line is often used in this context and an on-line computer pertains to the direct connection of measurement devices (or data terminals) to a computer in order to offer a real-time computational capability. On-line, is however, also used in another context. An 'on-line method' (IEEE, 1972) is a numerical procedure (computer algorithm) which does not require the reprocessing of past data once they have already been used. In other words, past data do not have to be stored within the computer central memory making such a method ideal for real-time operation where computer storage may be an important limitation. To avoid confusion, this definition will not be used here. Instead the term 'recursive method' will be used. Research hydrologists have realised that there is a wide range of theories and recursive methods from the fields of applied statistics and control engineering which can be applied to real-time hydrological forecasting and control.

During the past few years many rainfall-runoff models have been developed specifically for real-time forecasting using these methods. The models have two important properties which separate them from the classical catchment model. Firstly the models are formulated within a stochastic framework which may take into account noise arising from observation error, errors associated with parameter estimates (or state estimates) and errors due to an imperfect model. Statistical tests may be used to give an indication of the reliability of the forecasts. Secondly, statistically efficient recursive estimation procedures are used to estimate the model parameters or system states in real-time. These recursive estimation procedures are formulated to allow for updating of the forecasts using on-line measurements of both rainfall and runoff. The result of using these models (instead of the classical catchment model) usually results in a dramatic reduction of model implementation effort and computer storage requirement, together with a significant increase in the standard of real-time forecasting.

The development of strategies for the short-term management of a large water resource system is a very complex procedure because of the large number of variables considered, and the many uncertainties affecting the inputs and outputs of the system components. Thus while real-time models have and are being developed with some success for hydrological forecasting, there is little research being carried out in the field of real-time (automatic) control of water resource systems. This is probably because engineers tend to assume control as being a large multi-dimensional and complex problem. This need not be so, as a complex water resource system may be reduced to a number of 'subsystems' which can be individually controlled in real-time by automatic control computers. The justification for automatic control of the water resource system lies in both an increase in the standard of service and a future reduction in costs. A more sophisticated operating system has seldom resulted in a reduction in costs; however, standardization of control hardware and computer software packages will dramatically reduce the capital cost of control systems, while the cost of employing the equivalent manpower is steadily rising. Automation

would ensure that there were fewer people working more efficiently. Also, when very large quantities of data are being handled (inevitable if water quality is being monitored) computerized data handling is an absolute essential. Therefore the control computer could also be serving the subsidiary role of routine data aquisition.

2. OPERATIONAL MANAGEMENT OF WATER RESOURCE SYSTEMS

2.1 COMPLEX WATER RESOURCE SYSTEMS

Large modern water resource systems are generally characterized by the simultaneous operation of multiple facilities to meet multiple objectives. These facilities or system components may comprise storage reservoirs, regulating reservoirs (or multi-purpose reservoirs), pumping stations, generating plants and water treatment plants. The tasks which must be carried out during the planning and design phase of a proposed water resource system have been outlined in section 1.2. The objectives of system planning and design are to select the combination of variables (system components) together with suitable operating policies which will maximise the net benefits in accordance with the design criteria. These design criteria may be derived from technical, economic and social constraints and the benefits may be tangible (directly quantifiable in monetary terms) or intangible (not directly quantifiable in monetary terms e.g. preservation of wild life). The large number of constraints and objectives that are associated with a multi-component and multipurpose water resource system generally results in an almost infinite number of possible system component combinations and operating policies. Except for trivial systems, the selection and sizing of system components, and the determination of optimal operating policies are performed as separate tasks. Often, a further task is the scheduling (in time) of system component construction and commissioning.

Problems relating to these three tasks can be solved by linear programming (Dorfman, 1962), dynamic programming (Young, 1967), stochastic dynamic programming (Dudley and Burt, 1973), discrete differential dynamic programming (Heidari et al, 1971), simulation (Hufschmidt and Fiering, 1966), queuing theory (Moran, 1959), chance-constrained dynamic programming (Askew, 1974), branch compression techniques (Meier and Beightler,1967), Markov models (Gablinger and Loucks, 1970) and various combinations and modifications of these techniques. A survey paper on the planning and design of

water resource systems using many of the above techniques (with particular reference to British practice) has been given by Cole (1975).

The problem of operating a water resource system in real-time consists of assigning appropriate values to the system control variables such that optimal system operation is achieved. The system control variables may be groundwater pumping rates, reservoir releases, abstraction rates, effluent discharge rates etc. Recent studies in water resource system management have generally been concerned with the determination of medium-term optimal policies for surface water and groundwater storage allocation. The most complex problems arise with the determination of policies for multipurpose reservoir systems. These policies are used to determine reservoir retention levels and reservoir storage releases.

Discussion of the operational management of a water resource system can best be made with reference to a real system. Two examples will be given here; the Central Valley Project in California, and the River Dee System in North Wales.

The Central Valley Project is a well documented large multi-functional system (Becker et al, 1976), and serves to illustrate the complexity associated with operating such a system. On the other hand the River Dee System is a much smaller system. The River Dee System is of particular interest here, because it has been the subject of a general study into real-time flow forecasting for river regulation (Lowing, Price and Harvey, 1975)*.

Central Valley Project

The Central Valley Project shown in fig. 2.1 consists of nine storage facilities (reservoirs or lakes), seven power plants, eight pumping plants, two pumping-generation plants, four fish spawning

* *The River Dee research program (CWPU, 1977) included comprehensive data collection and archiving. These data, which are held at the Water Research Centre, Medmenham, are used for both simulated flow forecasting and river regulation studies in Chapters 5 and 6 of this thesis.*

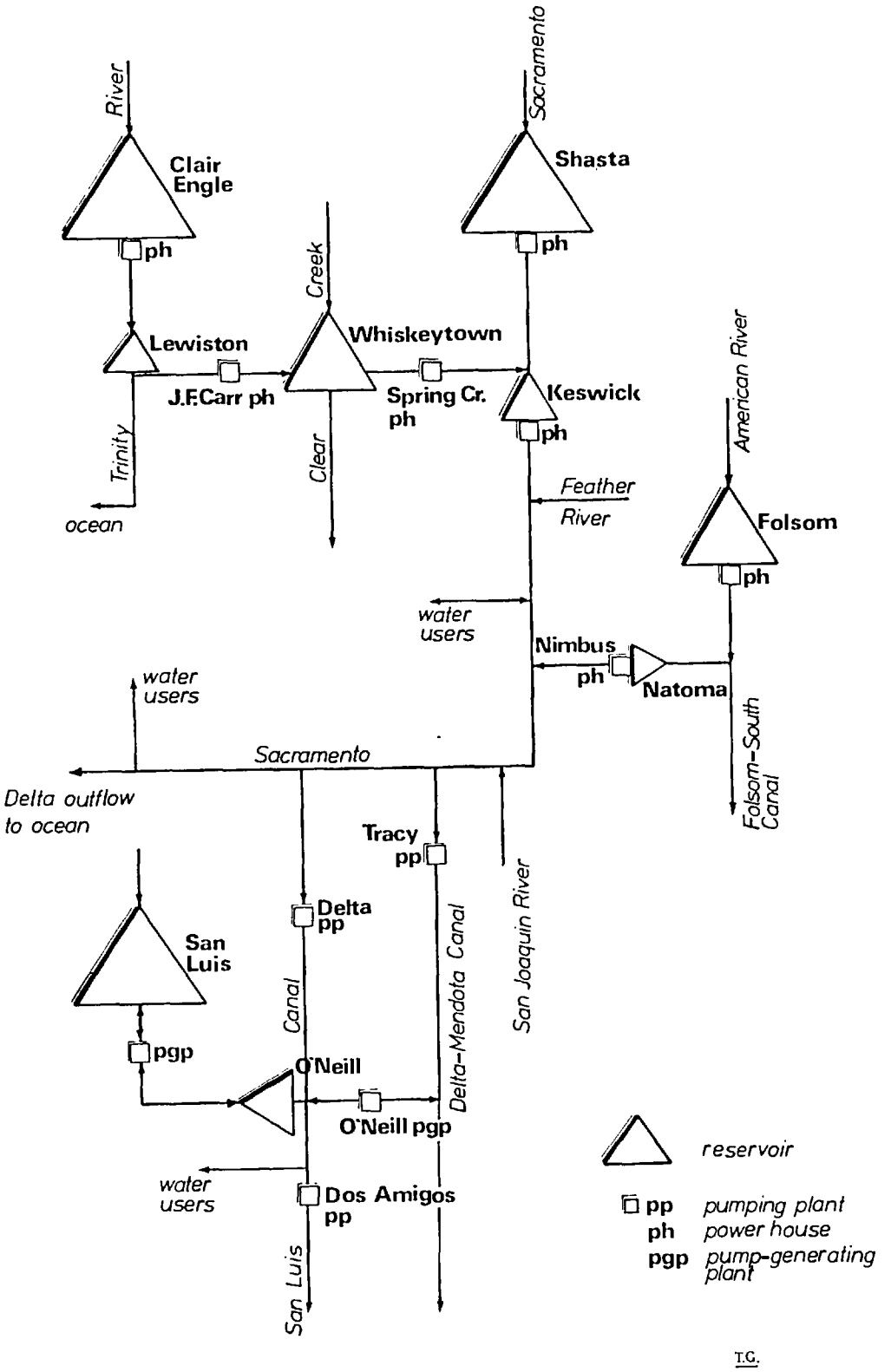


Fig. 2.1 Principal operating system components for the Central Valley Project, California, prior to 1976. Further storage facilities have since been added or are under construction.

facilities and nine canals. The reservoirs and their storage capacities are given in table 2.1. Shasta, Clair Engle, Folsom and San Luis are the major storage reservoirs within the system while Lewiston,Whiskeytown, Keswick, Natoma and O'Neill are essentially regulating reservoirs. Shasta and Folsom in addition to their main functions (water supply and power generation) are used for flood control purposes. Winter flows in the Trinity River are stored in Clair Engle for later release through Trinity powerhouse.

Part of the Trinity River water is diverted by gravity tunnel through Judge Francis Carr power plant,into Whiskeytown and then through Spring Creak power plant into Keswick Reservoir. Releases are made from Whiskeytown (whose storages are maintained within the constraints of a rule curve) down Clear Creek to satisfy downstream requirements. The major portion of Folsom dam releases are re-regulated at Nimbus Dam by releasing through the Nimbus power plant. The releases then combine with the Sacramento River which flows down to the Sacramento-San Joaquin Delta. Release constraints at Nimbus Dam have to be imposed to satisfy a daily Delta water balance (mainly for salinity control) which must be computed from tributary inflows, abstractions, and releases made from Keswick four days previously. Furthermore, releases are constrained within stated tolerances due to power requirements.

It is desired that the system operation be optimized on an hourly basis for maximum electrical energy, so that additional constraints are maximum and minimum loadings and efficiency characteristics of the individual generating units. San Luis and O'Neill releases for water supply and power output are made to formulae agreed between the state and the power companies. Either power generation or pumpback modes are available. San Luis is also used for low flow augmentation (summer irrigation demand).

Reservoir	Capacity $10^6 \times m^3$	Functions
Shasta Lake	5,600	I/FC/P/MI/WQ/N/R/F
Clair Engle	3,000	I/P/MI/WQ/N/R/F
San Luis Reservoir	2,500	I/P/MI/R
Folsom Lake	1,240	I/FC/P/MI/WQ/R/F
Whiskeytown Lake	292	REG/I/P/MI/R/F
O'Neill Forebay	70	REG/I/P/MI/R
Keswick Reservoir	29	REG/P/R/F
Lewiston Lake	18	REG/P/R/F
Lake Natoma	11	REG/P/R/F

Functions
I - irrigation water supply, FC - flood control, P - power generation, MI - municipal and industrial water supply, WQ - river water quality, N - navigation, R - recreational, F - fish protection, REG - river regulation.

Table 2.1 Reservoirs of the Central Valley Project, California and their functions

The operational management procedures for the Central Valley Project have developed over the years on an ad-hoc basis and depend largely on the skill and experience of the operators. Before 1966 the management decisions were generally made by one man. This was possible because there was minimal coordination between the management objectives, namely flood control, irrigation, navigation, water quality and power generation. As the project expanded, placing a greater demand on the available water resources and environmental concern became an important factor, management began to improve its effectiveness using new techniques (such as dynamic programming) together with better data collection and communication networks.

Today, the operation of the Central Valley Project is directed from the Central Valley Project Control Centre, Sacramento, with data communication lines to field divisions and their respective facilities (Fults and Hancock, 1974). Information is received at the control centre and analysed to provide orders relating to reservoir operations, power generation requirements, pumping rates etc. Flow forecasts to aid decision making are provided by the National Weather Service, Sacramento. In addition, 10 on-line water quality measurement stations linked to the control centre are located in the Delta. These are primarily used for salinity monitoring and forecasting, though any water quality problems which may arise can also be anticipated, and the necessary action taken at short notice.

The major operational task is the daily scheduling of hydro-power generation. The general procedure is for the system manager to forward a proposed schedule of hourly releases for the next day to the power authorities. The power authorities then assess the anticipated power requirements (the bulk of which is generated by thermal units). The projected hourly hydroelectric power demand for the next day is then determined by maximising the generating efficiency of the thermal and hydroelectric generating plants. The hourly demand projection is then submitted back to the system manager. Finally this demand is used by the system manager to provide a schedule of hourly water releases through the individual generating units. This is clearly an inefficient procedure because the optimisation of power generation and reservoir operation is not performed concurrently.

In an attempt to increase the effectiveness of the operational management of the Central Valley Project, Fults and Hancock (1972) have developed a series of mathematical models for providing optimal decision-making information. The procedure requires the use of three models which are respectively dependent on a monthly, a daily and an hourly time-step. The monthly and daily models are based on the minimisation of the loss in potential energy of the stored waters in the reservoirs resulting from any release policy of one

year duration for the monthly model and of one month duration for the daily model. The hourly model is based on the maximization of power generation over a 24 hour period. The three levels of models are necessary to reduce the computational burden. The models are linked as follows. The output from the monthly model (monthly data) is the input for the daily model, while the daily data is used as input to the hourly model. Referring to section 1.2, the monthly and daily models are used to determine the medium-term strategy (noting that the daily model produces the release policy for one month and not one day), while the real-time strategy is actually provided by the hourly model. Computational and institutional difficulties at present preclude the use of these models in real-time, though it seems inevitable, with the continuing expansion of the Central Valley Project, that these (or similar) models will become the principal tools used in the decision-making process.

River Dee System

The River Dee water resource system shown in fig. 2.2 consists of three major regulating reservoirs (with controlled releases), Llyn Celyn, Llyn Tegid and Llyn Brenig which are primarily used to maintain prescribed flows upstream of major water supply abstraction points near Chester. The reservoirs have been described by Lambert and Cameron (1975) as follows:

(a) Llyn Tegid (Bala Lake) is used for short term retention of flood runoff for subsequent controlled releases. Releases are also made for low flow augmentation in which case they are closely controlled with adjustments being made on a day to day basis. The retention levels in Llyn Tegid may be controlled by releasing from Llyn Celyn. During April to September inclusive, a minimum lake level must be maintained for intensive amenity use.

(b) Llyn Celyn is the principle storage reservoir utilised every year for low-flow augmentation of the Dee. Storage is also provided for short to medium term retention of flood runoff, with

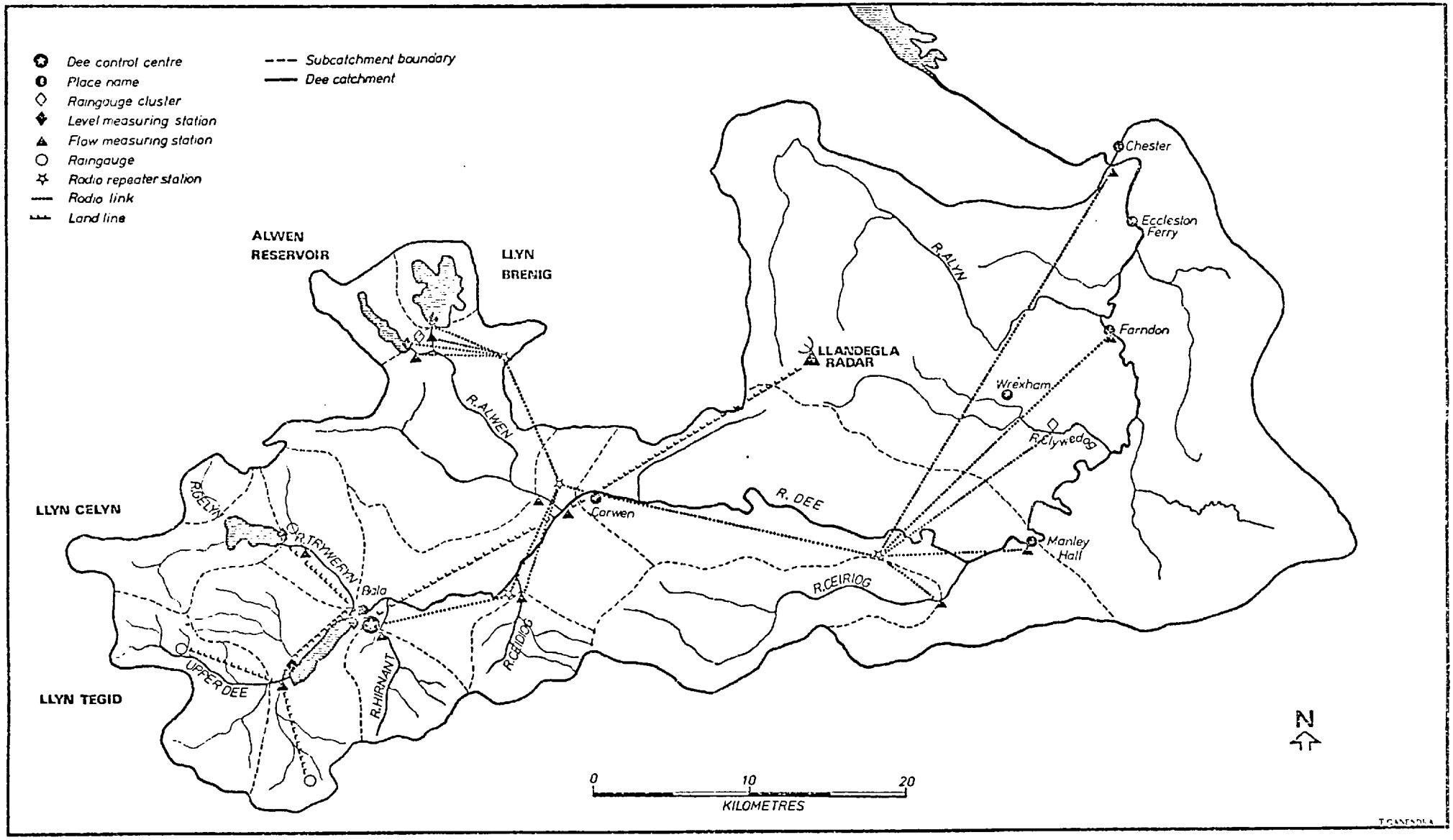


Fig. 2.2 The River Dee Water Resource System, North Wales. The on-line data collection network is centred around the Dee control centre at Bala.

a varying retention level depending upon the time of year. Llyn Celyn is also used for hydro-electric power generation, however priority is given to flood control or low flow river regulation.

(c) Llyn Brenig will be used for assisting low flow augmentation of the River Dee, and for amenity use. (Llyn Brenig commenced filling in August 1975 and was 44% full in March 1977).

In addition to the above three reservoirs, direct water supply is provided by Alwen Reservoir. The reservoirs and their respective contributing catchment areas are shown in table 2.2.

Reservoir	Capacity $m^3 \times 10^6$	Catchment Area km^2	Functions
Alwen Reservoir	14.5	25.5	Direct supply
Llyn Tegid	17.6	266.8	Flood control, river regulation, recreation
Llyn Celyn	80.7	59.9	River regulation, flood control, hydro-power
Llyn Brenig (Stage 1)	60.0	22.2	River regulation

Table 2.2 Reservoirs on the Dee Catchment

Reservoir operation is dependent on rule curves, such as the 1975 Llyn Celyn rule curve shown in fig. 2.3. Here available storage above the normal retention level is used for temporary detention of floodwater. The reservoir rule curves represent the medium-term strategy discussed in section 1.2. They have been derived with the global objective of achieving a compromise between flood control and low flow augmentation objectives.

The River Dee System has been the subject of a general study from the control rule aspect (Lambert and Cameron, 1975). Included in that study was the investigation of potential benefits of real-time forecasting to operational requirements in river regulation schemes and the use of radar-derived rainfall estimates as input to a hydrological model.

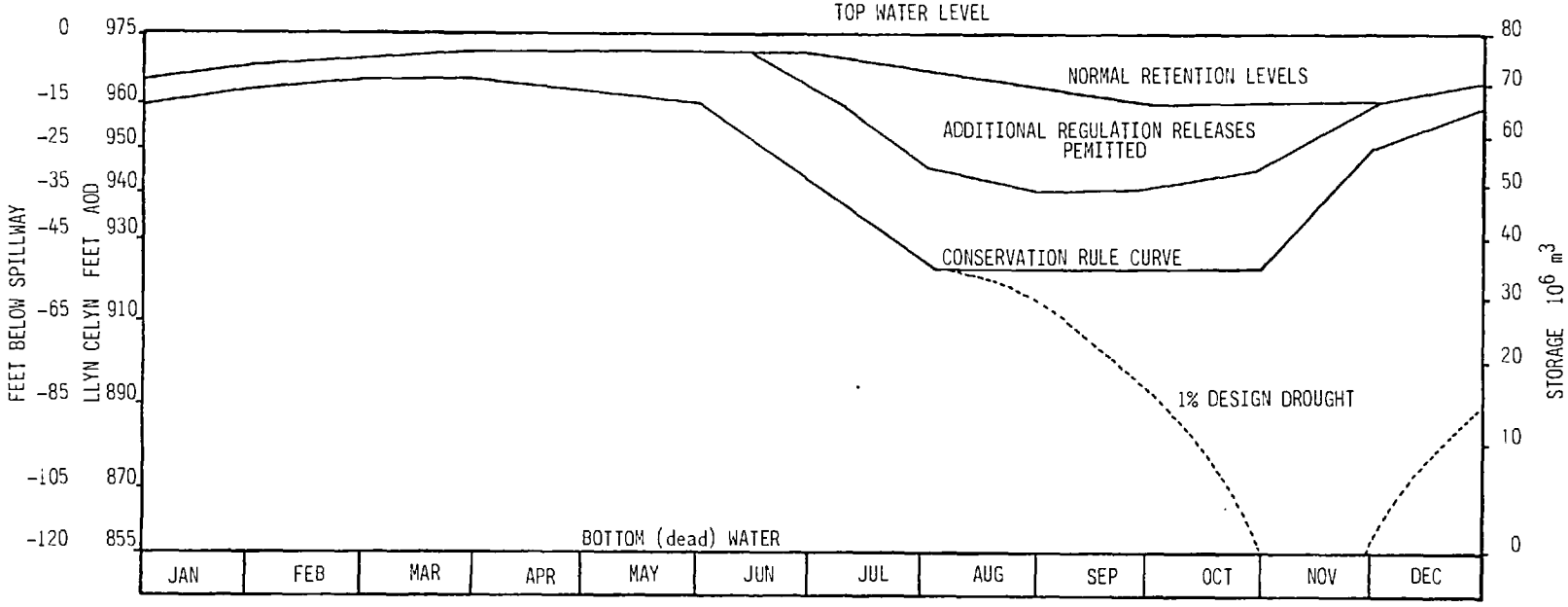


Fig. 2.3 Llyn Celyn rule curve

A simulation model of the River Dee basin has been developed to provide real-time forecasts of river flows (Lowing, Price and Harvey, 1975). The model consists of subcatchment models of the Ceiriog, Hirnant, Upper Dee, Celyn, Alwen and the Ceidiog subcatchments (Lambert, 1972). The outputs of the subcatchment models are used as inputs to a model of the main channel (River Dee) between Bala and Manley Hall. To take into account the variation in the Bala to Manley Hall travel time of between 9 hours for flood peaks to 19 hours for low flows, the main channel is modelled using the variable parameter diffusion method (Price, 1973). Runoff from ungauged catchments is estimated by reference to the nearest gauged catchment. The forecasted flows at Manley Hall are finally routed down to Farndon (the measured stage at Farndon relates to discharge at Eccleston) using a combination of two linear reservoirs and two linear channels in series (Cole, 1976).

As shown in fig. 2.2 the River Dee System has an extensive on-line data collection system. Hydrometric outstations are connected by landline or UHF radio links to a control centre at Bala. Interrogations of the raingauge and water level gauges are performed automatically by a mini computer (PDP 11/40) every half hour and the data displayed on a mimic panel or on a teletype. The inclusion of a weather radar in the on-line hydrometric network was an important development in the operational management of the River Dee System (the radar has now been moved to Titterstone, Shropshire to serve the Severn-Trent Water Authority). The weather radar was situated at Llandegla and provided quantitative rainfall measurement as input data to the forecasting model or alternatively the data could be displayed in the form of an isohyetal map on a colour television screen. The radar was calibrated on-line using data from the raingauge clusters indicated in fig. 2.2.

During low flow regulation periods, releases are made from the regulating reservoirs so that if possible the prescribed flow at Eccleston is maintained with a minimum wastage (excess release). The long travel time between Bala and Eccleston requires that flow

forecasts at Eccleston are required for a 36 hour lead time in order to anticipate the correct releases. Release patterns are determined daily or twice daily as necessary. This is a trial and error procedure, requiring multiple runs of the Dee simulation model. However in the absence of surface runoff and anticipated rainfall, and with the river in a state of recession, there is little element of hydrological forecasting required other than to calculate the speed with which a change in release is propagated downstream. In such circumstances a master recession curve technique (CWPU, 1977) is used instead of the simulation model.

While the need for a real-time model to assist in low flow augmentation becomes less during a drought as the drought becomes more severe (and recession rates become smaller or negligible), the converse is true when considering flood alleviation and the severity of floods. The real-time model installed at Bala is used to predict Llyn Tegid and Llyn Celyn retention levels and flows in the main channel. Flood warnings are issued by reference to the Llyn Tegid levels, for it is the town of Bala which is prone to the most damage if flooding occurs. When a flood producing situation is imminent, steps are taken to limit any increase in the Llyn Tegid retention levels and to ensure that there is sufficient retention storage available. This usually means an increase in the release from the Bala sluices coupled with a reduction of flow from Celyn to Tegid. During this stage, bank-full conditions between Bala and Corwen must not be exceeded. If a bank overflow condition is inevitable (in which case a flood warning for the upper Dee would already have been issued), then a warning that flooding of the A5 trunk road at the point where it crosses the Dee at Corwen must be issued. The on-line, half-hourly acquisition of hydrological data is used to obtain accurate information on the system state. This together with the real-time flow forecasting model can be used to aid decision making at the Dee control centre.

2.2 A REAL-TIME CONTROL STRATEGY

Two levels of time dependency (short-term and medium-term) have been identified for the operational management of a water resource system. A proposed methodology for the operational management of large complex water resource systems is presented in this section. The approach is based on the decomposition of the whole system into a number of subsystems, which can be individually controlled in real-time. Concurrently, the system as a whole can be controlled optimally in the medium-term, though this does not ensure global optimality in the real-time. However the flexibility of the approach is such that the system may be operated optimally in real-time if necessary. The approach has become theoretically feasible, with the availability of on-line forecasting models and control algorithms which adapt to changes in the system state, and applicable in practice due to the widespread availability of digital computers, the necessary data transmission media and on-line hydrometric instrumentation.

A large water resource system such as the Central Valley Project described in the previous section includes a large number of system components which have to be controlled in real-time. The global optimization of the system as a whole on say an hourly basis presents a formidable computational task. With such large systems, there is a tendency for 'centrality' of system management. This means that all the information available about the system, and the calculations based upon this information are available or take place at a single location. However the vast volume of data that may have to be handled could introduce an unacceptable level of unreliability. The computational requirement may also exceed the available central computer resources. For a large water resource system, geographical separation of the system components and reliability of communication links to some extent favour a 'decentralized' decision making framework. With the present day computer technology making possible low cost small microprocessor based computers, a distributed computer system is both technically

and economically feasible.

The subsystem

A reduction in the complexity of a water resource management problem may be affected by dividing the system in question into a number of inter-dependent subsystems. An example of a hypothetical subsystem is given in fig. 2.4.

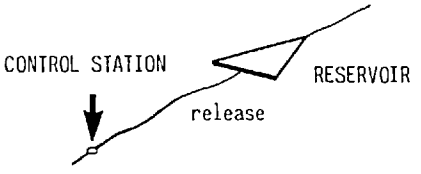


Fig. 2.4 A subsystem : River regulation

The control variable for the above subsystem is a sluice gate, so that subsystem control refers to the actual manipulation of the sluice gate to alter the release. The control of the subsystem is assumed to be based on a 'satisficing' approach (Jamieson, 1972) where priority is given to a primary objective which must be satisfied by real-time control. Secondary objectives need only to be considered if the primary objective is already satisfied to a predetermined standard. In fig. 2.4 it is further assumed that downstream abstraction requires that the river is maintained at a desired reference level at some point along the river. This point will be referred to as the control station, while the desired level will be referred to as the set-point (IEEE, 1972). The objective of control is then to make releases from the reservoir so that the difference between the observed level and the set-point is minimised. If the reservoir is being used for supply purposes, the releases which result in flows exceeding the set-point will constitute wastage of reservoir storage.

If the reservoir is also used for flood alleviation, sufficient storage capacity must be reserved for the temporary detention of flood-waters. In the event that the available storage is allowed to become less than the flood storage requirement, releases must be scheduled to ensure a reservoir drawdown rate which will enable this storage requirement to be maintained. This is usually determined by reference to downstream bank-full constraints. The problem is similar to that of flow augmentation, except that the set-point is set to a maximum value corresponding to the maximum allowable river discharge.

Thus the real-time control of the subsystem shown in fig. 2.4 for both flood control and low flow augmentation can be achieved by reference to the downstream set-point. In effect the set-point can also be manipulated to adjust the retention level of the reservoir.

Satisficing approach

The organization of the above hypothetical real-time control procedure is into a primary control objective of maintaining a downstream requirement at the control station and a secondary control objective of maintaining a desired reservoir level. The desired reservoir level can be pre-determined on the basis of a medium-term strategy and presented in the form of a set of control rule curves as described in Section 2.1. The primary control objective (referring to the actual reservoir release as the control) is to satisfy a downstream requirement, while the secondary objective is to allow 'following' of the reservoir rule curves. The control procedures which depend on the reservoir level in relation to the reservoir rule curves (i.e. a low reservoir level is equivalent to the reservoir level being below the minimum reservoir level control-rule curve) are outlined below:

(a) Reservoir level normal, low flow augmentation required	Release for water supply downstream. Control objective is to meet set-point flow.
(b) Reservoir level normal, low flow augmentation not required	Release water to maintain normal reservoir level. Control objective is to meet desired reservoir level.
(c) Reservoir level low, low flow augmentation may or may not be required	Release for water supply downstream only when required. Control objective as for (a) but if necessary set-point is lowered and/or abstraction rate is less.
(d) Reservoir level high	Release for water supply downstream. Control objective as for (a) but set-point is raised, and if flooding is imminent, set-point is set to maximum.

So while the primary control objectives in (a), (c) and (d) are to satisfy a downstream objective, the downstream objective itself may be determined by reference to the reservoir level. The control objective in (b) is secondary as the primary control objective is already satisfied. For real-time operations, forecasts of inflows into the reservoir which will aid in the anticipation of reservoir level changes are required.

Operating policy

The real-time control should be based upon forecasts of the future system state. In fig. 2.4, changes in the system state are mainly due to stochastic variations in reservoir inflows and deterministic downstream abstraction rates. Optimal (in a broad sense) management of this subsystem depends on correctly allocating values to the set-point. This may be based on a medium-term strategy with the option of adopting a real-time strategy during emergency situations. The medium-term strategy can comprise a set-point rule curve (or decision table) which is related to the observed reservoir level and the time of year. Switching from control modes (a) to (d) above should, with the aid of reservoir inflow forecasts, be accomplished in real-time.

Hierarchical decentralized control

Using the above approach, the management framework of a complex water resource system conveniently allowing for both short and medium-term time dependency can be developed. Here, while the control variables are specified in real-time, the desired subsystem outputs (set-points) can be specified at the medium-term level or in real-time. For water resource subsystems, the control variables may include pumping rates and reservoir releases while the set-points may include reservoir levels, groundwater table levels and river levels.

By dividing the system into subsystems, the management organization (as shown in fig. 2.5) is characterized by two separate entities, the regional centre and the subsystem field station. A distributed computer system can easily be developed for this organizational framework. Medium-term planning can then be carried out at the regional centre with a large in-house computer system capable of handling problems requiring extensive computational facilities as well as 'expert level' manpower. From the regional centre the policies (set-points, rule curves, etc.) can be transmitted to subsystem field stations, each of which would have its own dedicated mini/micro computer system. As shown in fig. 2.6, data can be acquired and sub-system control computations performed in real-time by the subsystem mini/micro computer. If the control decisions are then executed automatically by the computer, the staff requirement at the subsystem field stations will be minimal. Computer communications between subsystems could if necessary be achieved via the regional centre. With recursive forecasting and control algorithms which do not require the re-processing of past data, the data-base for all subsystems need only be held at the regional centre. In conclusion the objective of the regional centre is to achieve global optimal management of the water resource system by issuing medium-term directives for the control of the subsystems. Clearly the global real-time system operation will be sub-optimal because the subsystems are controlled individually in real-time. Thus the interactions between the subsystems at the real-time level are

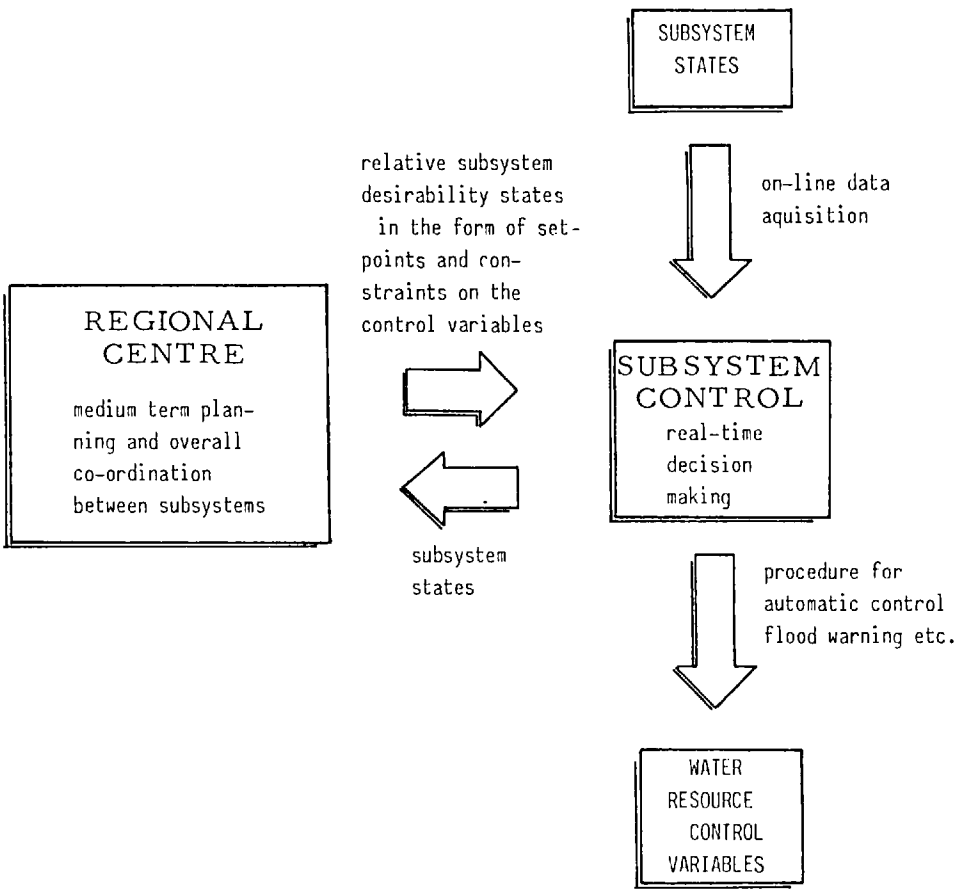


Fig. 2.5 Organization of the operational management system

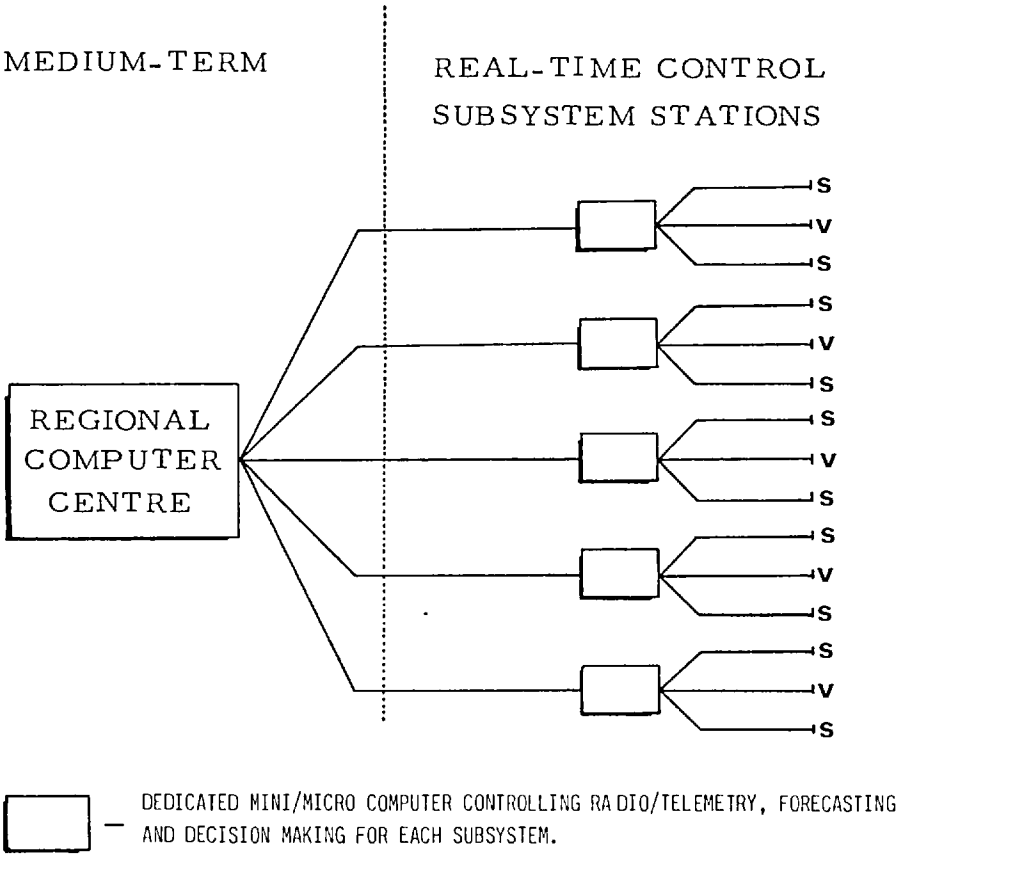


Fig. 2.6 Hierarchy of distributed computers

not considered. This approach does however allow interaction between subsystems at a medium-term level, and the facility for the control monitoring of selected system states in real-time. This facility could be reserved for emergency situations when, if necessary any subsystem could be controlled in real-time from the regional centre by issuing directives (set-point etc.) in real-time instead of at the medium-term level.

The management of the water resource system as described above is based on a hierarchy of control decisions, as shown below.

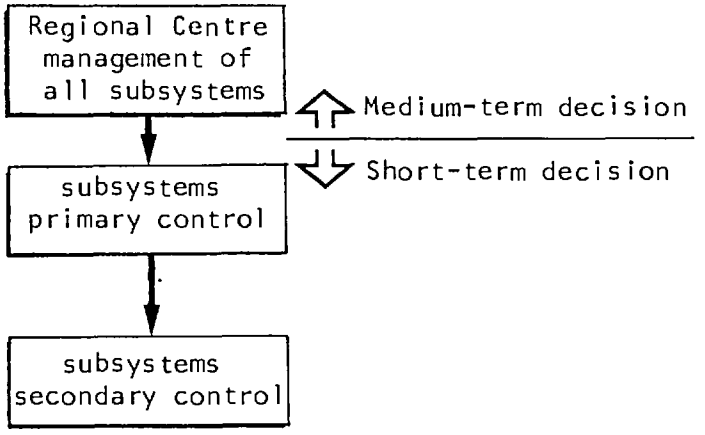


Fig. 2.7 Hierarchy of water-resource system management

Generally, there may be several (hierarchial) levels of secondary control within a sub-system. In other words, if the most important secondary objective is satisfied, the next most important secondary objective forms the subsystem control objective.

Control Hardware

The control hardware required for the real-time control system as shown in fig. 2.6 consists of a large central computer, sub-system mini or micro-computers, data transmission media, remote sensoring devices (e.g. telemetering raingauges) and mechanical control devices (e.g. reservoir release valves). Of particular interest in this study is the recent rapid development of digital computer technology.

Present day digital computers may be categorized into (a) main-frame computers, (b) mini-computers and (c) micro-computers. The rapid development of the mini-computer during the last five years has resulted in some difficulty in clearly defining the separation between (a) and (b) while the present rapid development of micro-computers has resulted in some difficulty in clearly defining the separation between (b) and (c). The computer classifications may be outlined as follows:

(a) Mainframes — Mainframe computers are large computers characterised by large memories (internal computer storage) and fast central processing units. The memory of this class of computer is generally in excess of 64K bytes and they are always associated with high cost in relation to mini-computers.

(b) Mini-computers — Mini-computers range from desktop machines which are optimized for use by a single user to large systems based on a time-sharing and a multi-user environment. The memory size may range from 4K bytes (desk-top machine) and upwards to 1M byte (super-mini). The cost may range from £5,000-£100,000.

(c) Micro-computers — A micro-computer is based on a micro-processor (which is essentially a large scale integrated circuit chip). Micro-computers have, until recently been associated

with limited input-output facilities, low-level language programming, and a small memory capacity. Micro-computers have now been developed, particularly for personal computer applications, which can be complemented by a wide range of peripheral equipment, as well as supporting languages such as Basic and Fortran. The cost of a micro-computer is in the region of £100-£1,000.

It is assumed that in general the micro-computer will be able to provide the computing capability required for a subsystem. While the software development costs for this type of computer are very high, standardization of forecasting, control and data handling 'packages' could virtually eliminate these capital cost components. If however, the number of tasks to be performed by the computer is very large, a mini-computer may have to be employed. The more powerful computer may be required in particular if a very large number of water quality variables have to be sampled within a short period of time. The regional computer centre will almost certainly require a mainframe computer or a super-mini computer of mainframe capability.

Limitations of hydrological forecasts on the real-time strategy

One of the fundamental problems associated with the real-time forecasting and control of a water resource system as described above is the relatively short lead-time with which accurate forecasts of stream-flow can be made. If rainfall is observed in real-time, but not forecast, the maximum forecast lead-time for which exact runoff forecasts can theoretically be obtained is equal to the pure time delay between the rainfall and resultant runoff. This can be shown by considering a storm with rainfall occurring in two separate events, A and B as shown by the hyetograph in Fig. 2.8.

The shaded portion of the hydrograph indicates the runoff due only

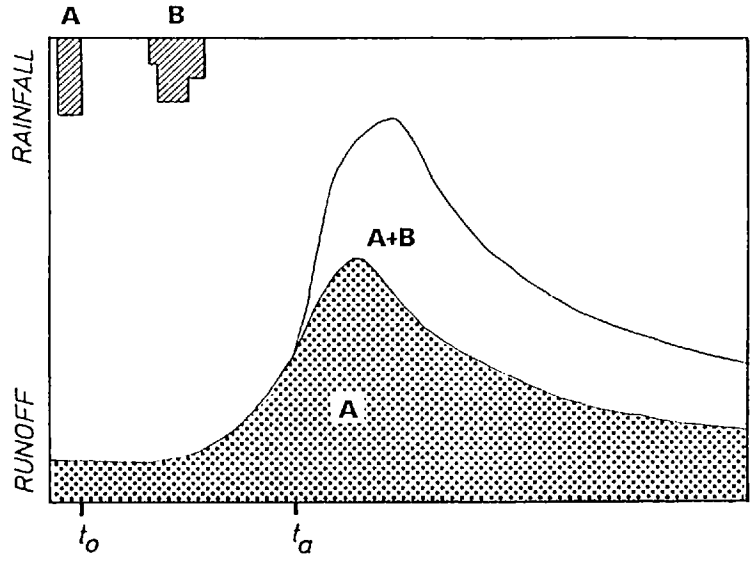


Fig. 2.8 Runoff response from two rainfall events A and B

to rainfall event A while the remaining portion of the hydrograph shows the runoff occurring as a result of both events A and B. Now if at time t_o, a forecast is to be made using knowledge of rainfall event A (but not B), then the maximum lead-time is $t_a - t_o$, and any forecast beyond t_a will be an under-estimation as B is unknown. Accurate forecasts of the runoff beyond this limit would have to rely on some procedure for obtaining quantitative precipitation forecasts.

Little work has been done on the application of real-time weather forecasting to runoff modelling, though it seems likely that in the near future quantitative precipitation forecasting will form an integral part of the flow forecasting procedure. At present there are several meteorological centres in the world providing quantitative precipitation forecasts. The forecasts which are usually stated in depth and probability of occurrence, generally refer to average rainfall over a very large area. Therefore these rainfall forecasts may be unrepresentative of the rainfall actually falling within the catchment area in question (assuming a small to medium sized catchment of less than 1500 km^2 in area).

In the near future however, numerical weather prediction models coupled with information from weather radar and satellite imagery will probably be able to give sufficiently accurate forecasts for inputs to streamflow forecasting models. This in turn will allow a better real-time control of the water resource system which is in general dependent on forecasts of the system states.

In this study it will be assumed that quantitative rainfall forecasts are not available, and that observations of rainfall are obtained from a network of on-line raingauges.

Raingauge network design

The design of a network of radio/telemetering raingauges is fundamental to a real-time flood forecasting system. The formulating of a model of the rainfall-runoff process which can provide forecasts with minimum square error should take full account of the fact that measurements of streamflow are available up to the present time, and consequently the necessity to define accurately mean areal rainfall over a catchment as model input becomes less important. However, the location of gauges and the method used to define model input from their measurements assumes a more important role in relation to the accurate forecasting of the rising limb and peak of the hydrograph.

Bras and Rodriguez-Iturbe (1976) have investigated the effect of raingauge network density and configuration on the simulation of the rising limb and peak of the hydrograph on a hypothetical catchment of area 212 km^2. Using an idealized model based on the kinematic wave equations, it was found that estimates of the mean square error for the peak of the hydrograph were reduced when gauges were located in the upstream reaches of the catchment, but this increased the error of estimation for the rising limb of the hydrograph. Station location was found to be more important than the number of stations, although a reduction in mean square error was still observed when up to eight gauges were included in the network. It was noted though, that the falling limb of the hydrograph could be accurately reproduced using only a few gauges. An updating facility was not included in the rainfall-runoff model.

In a recent exploratory study, O'Connell et al (1977) establish by direct statistical inference a causal relation between rainfall and runoff. The model was formulated as a linear autoregressive moving average representation of the rainfall-runoff process. This allowed an updating facility based on observed streamflows to be included for real-time forecasting. The model was applied to the 33.9 km^2 Hirnant Catchment, which has a dense raingauge network. It was found that the use of those gauges having the strongest causal relation with river flow may provide the basis of a better forecasting model than the areal average computed using all gauges. The results also suggested that one or two gauges should be adequate for real-time flow forecasting in small catchments provided they sample the area contributing to the rising limb and peak of the hydrograph.

In general the storage effect of the catchment does alleviate the network design problem in that rain falling on more remote parts need not be sampled, the falling limb being adequately modelled by an autoregressive flow component. This only pertains to real-time flow forecasting, where the facility exists to reset continually previously forecast flow values to their measured value in real-time, for use in subsequent forecasts.

The problem of instrument failure may prove to be the more important criterion affecting gauge density requirement. The reliability of radio/telemetry installations to transmit gauge measurements must also be considered in network design. All automatically performed computational procedures relating to forecasting and control of the subsystem must incorporate some detection and error recovery procedure, when failure of on-line instrumentation occurs resulting in spurious measurements being received by this data acquisition system.

3. LINEAR FILTERING AND PREDICTION

3.1 FILTERING AND PREDICTION THEORY

Much of the recent research into the application of estimation theory to real-time forecasting and control of water resource systems is based on a technique developed by Kalman (Kalman, 1960 , Kalman and Bucy, 1961) and first applied to space vehicle trajectory control. Known as Kalman filtering, this technique is a modern version of Gauss' least squares method. The Kalman filter encompasses two main features which separate it from the classical method and these are: (a) errors in the observations and/or errors associated with modelling and estimation are accounted for, and (b) the problem is formulated within a particular framework of equations referred to as a state space representation. The Kalman filter is a recursive method, making it particularly suitable for real-time applications.

Again, giving a more rigorous definition to the state (see section 1.2), the state of a deterministic system is the smallest set of variables such that the knowledge of the variables at time $t = t_o$, called state variables, together with the input for $t \geq t_o$, completely determines the behaviour of the system for $t \geq t_o$. Thus the state of the system separates the future from the past as the state contains all the information required to determine the response to the present input. For stochastic systems however the future behaviour of a system cannot be uniquely determined by the state alone, and instead the probability distributions of the state variables for $t \geq t_o$ are uniqely determined by the state at $t = t_o$. An example of a simple stochastic system is the Markov process.

Conventionally, in the systems approach, a linear system will imply linearity in the sense that the principle of superposition holds, i.e. if $y_1(t)$ and $y_2(t)$ are the system outputs corresponding to the inputs $u_1(t)$ and $u_2(t)$ then the input $u_1(t) + u_2(t)$ will produce the output $y_1(t) + y_2(t)$. Non-linearity will imply that the principle

of superposition does not hold. A time-invariant system is one whose input-output relationship does not depend on the time at which the input is applied, i.e. if $y_1(t)$ is the system output corresponding to the input $u_1(t)$, then the system output $y_1(t+\tau)$ corresponds to the input $u_1(t+\tau)$ where τ represents a shift in the time domain. The converse is true for a time-variant system. Furthermore while the system state is usually continuous (the state variables vary continuously in time), the state space representation requires that the state variables are known only at finite time increments and hence the system model is assumed discrete (the state variables vary at discrete time intervals).

The dynamic behaviour of a stochastic system is often given by the following nth order linear discrete stochastic vector difference equation

$$X(t) = \Phi X(t-1) + \Gamma U(t-1) + W(t-1) \qquad (3.1)$$

where Φ is the $n \times n$ state transition matrix, Γ is the $n \times p$ control transition matrix, $X(t)$ is the n-vector of the system states, $U(t)$ is the p-vector of deterministic input (or control) variables and $W(t)$ is a vector of $N(0,Q)$ Gaussian white noise variables known as the process disturbances. The above system is time-invariant as the elements of the transition matrices are not functions of time. A time-variant system could be similarly modelled, but with time dependent Φ and Γ matrices. When Φ and Γ are independent of time the process is also known as a stationary process. In order to obtain information on the system state, measurements of the output must be made. The measurement equation is given by

$$Z(t) = HX(t) + V(t) \qquad (3.2)$$

where $Z(t)$ is the m-vector of measurements (or system outputs), $V(t)$ is the m-vector of $N(0,R)$ Gaussian white measurement noise variables, and H is the $m \times n$ measurement matrix. The measurement equation (eqn. 3.2) together with equation 3.1, known as the system or state

equation, gives a state space representation of a linear time-invariant dynamic system.

Having defined the system model, the problem that remains to be solved is that of the state estimation. The estimate of the state $X(t+k)$ at time $t+k$ based on measurements up to time t is denoted by $\hat{X}(t+k|t)$. If $k = 0$ the estimation problem is called filtering, if $k > 0$ it is called prediction and if $k < 0$ it is called smoothing or interpolation. For a system model and measurement model specified by equations 3.1 and 3.2, Kalman (1960) has given a sequential procedure known as Kalman filtering for the estimation problem in which the quadratic loss function

$$J = E\{\tilde{X}(t|t)\tilde{X}(t|t)^T\} \tag{3.3}$$

is minimised where $E\{.\}$ is the expectation operator and

$$\tilde{X}(t|t) = \hat{X}(t|t) - X(t). \tag{3.4}$$

The estimate of the state in this case will be optimal in the sense that the expected value of the mean square error (error variance) of the state estimates will be a minimum. The procedure is outlined below.

(i) Predict the state estimate

$$\hat{X}(t+1|t) = \Phi\hat{X}(t|t) + \Gamma U(t) \tag{3.5}$$

(ii) Predict the error covariance of the state estimates

$$P(t+1|t) = \Phi P(t|t)\Phi^T + Q \tag{3.6}$$

(iii) Obtain the Kalman gain matrix

$$K(t+1) = P(t+1|t)H^T[HP(t+1|t)H^T + R]^{-1} \tag{3.7}$$

(iv) The filtered state estimate is then given by

$$\hat{X}(t+1|t+1) = \hat{X}(t+1|t) + K(t+1)[Z(t+1) - H\hat{X}(t+1|t)] \tag{3.8}$$

(v) The updated error covariance of the state estimate is

$$P(t+1|t+1) = [I - K(t+1)H]P(t+1|t) \tag{3.9}$$

The initial condition assumptions are given by

$$P(0) = E\{\tilde{X}(0)\tilde{X}(0)^T\} \tag{3.10}$$

$$X(0) = \hat{X}(0) \tag{3.11}$$

and it is assumed that the covariance

$$E\{W(t)\ V(k)\} = 0 \tag{3.12}$$

for all t and k. The above Kalman filter specifies how the new noisy measurement Z(t+1) is combined with $\hat{X}(t+1|t)$ to provide an updated estimate of the state $\hat{X}(t+1|t+1)$.

The attractive features of the Kalman filter are:

(i) the state space formulation is flexible and can be adapted to a wide range of applications;

(ii) system noise and measurement noise are handled separately;

(iii) the recursive calculation procedure is particularly amenable to implementation on a mini or micro computer;

(iv) optimal (in the mean square error sense) estimates and predictions are obtained;

(v) the covariance matrices of the predictions and the updated state estimates are provided.

Furthermore the properties of the innovation sequences defined by

$$\nu(t+1) = Z(t+1) - H\hat{X}(t+1|t) \tag{3.15}$$

can be used to provide tests for the performance of the filter. For an optimal filter where P(0), ϕ, Γ, H, Q and R are correctly specified, the innovation sequences should be white Gaussian noise uncorrelated with the measurement vector (Mehra, 1969), i.e.

$$E\{\nu(t)\nu(t+k)^T\} = 0 \quad ; \quad k \neq 0$$

$$E\{\nu(t)Z(k)^T\} = 0 \quad ; \quad t > k \tag{3.14}$$

Clearly the Kalman filter is not a model in itself but an estimation procedure used in conjunction with a system model which is an imperfect representation of the process being modelled. The system state is introduced for mathematical convenience and may or may not have a physical interpretation. The state space formulation given by equations 3.1 and 3.2 is one example from a wide range of possible system models to which Kalman filtering is applicable. The state space equations may be of more complex form than equations 3.1 and 3.2 in that the transition matrices may be time variant, system and model noise coefficient matrices may be introduced, a more general form of the loss function (eqn. 3.3) may be used and the system model may be non-linear. Conversely simplified forms of equation 3.1 and 3.2 may be used. The reader is referred to Jazwinski (1970) or Gelb (1974) for a more detailed exposé of Kalman filtering.

There are three basic approaches to real-time flow forecasting using the Kalman filtering method. The first two approaches assume that the rainfall-runoff process can be formulated as an ARMAX* process model given by

$$q(t) + a_1 q(t-1) + \ldots a_n \, q(t-n) = b_0 r(t-b) + b_1 \, r(t-b-1) + \ldots$$

$$\ldots + b_n r(t-b-n) + \xi(t) \qquad (3.15)$$

where at time t, $q(t)$ is the catchment discharge, $r(t)$ the rainfall, $\xi(t)$ is some noise process and n is the order of the model. For notational convenience it is assumed that the number of r and q terms is the same, though in practice they may, and generally will be, of different order (this can easily be accommodated by assuming that the appropriate coefficients are zero). The pure time delay (transport lag) between the rainfall and runoff is given by b. Note that if $q(t)$ and $r(t)$ are replaced by $q'(t)$ and $r'(t)$, the storm runoff and excess precipitation, then equation 3.15 is merely a stochastic formulation of the well-known discrete form of the convolution integral (Dooge, 1973) given by

---

* *The term ARMAX is used to denote an autoregressive moving average (ARMA) model with exogenous deterministic inputs (Young and Whitehead, 1977).*

$$q'(t) = \sum_{i=1}^{m} h(i)r'(t-i) \tag{3.16}$$

where h(i) represents the m ordinates of the unit hydrograph. Equation 3.15 is also a parsimonious parameterization of a discrete convolution integral model, in that the impulse response of both models are the same but equation 3.15 contains fewer parameters. The third approach does not require that the rainfall-runoff process be represented by an ARMAX model. The three approaches are outlined below.

Type 1

The measurements (assumed to be noisy) are elements of the measurement vector H(t) (eqn. 3.2), while the state vector X(t) consists of the parameters of the model to be estimated. The state transition matrix must be determined a priori, an example being the identity matrix, resulting in the simple random walk model

$$X(t+1) = X(t) + W(t) \tag{3.17}$$

The above equation which assumes time variant parameters is the system equation while the measurement equation is given by

$$Z(t) = H(t)X(t) + V(t) \tag{3.18}$$

where X(t), H(t) and X(t) are given by (from equation 3.15)

$$Z(t) = q(t) \quad ; \quad H(t) = \begin{bmatrix} q(t-1) \\ \cdot \\ \cdot \\ \cdot \\ q(t-n) \\ r(t-b) \\ \cdot \\ \cdot \\ r(t-b-n) \end{bmatrix} \quad ; \quad X(t) = \begin{bmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_n \\ b_o \\ \cdot \\ \cdot \\ b_n \end{bmatrix} \tag{3.19}$$

The noise sequences W(t) and V(t) are white Gaussian noise sequences distributed as N(0,Q) and N(0,R) respectively. The Kalman filter

procedure is used to estimate the parameters (state vector) of the model.

Type II

In the second approach the parameters of the model adopted (i.e. eqn. 3.15) are elements of the state transition matrix while the state vector contains the unknown true values of the noisy measurements (the output vector Z(t) contains the noisy measurements). A separate estimation procedure is required to estimate the model parameters themselves, which define the state transition matrix. The measurement matrix H is usually taken as the identify matrix.

Type III

In these applications of Kalman filtering to real-time flow forecasting, conceptual models are formulated in the state space framework in order to facilitate efficient estimation of parameters and the handling of modelling and measurement errors. The transition and/or measurement matrices are often non-linear functional matrices in which case the (non-linear) extended Kalman filter (Jazwinski, 1970) must be employed. This algorithm linearises the system state around the previous estimate of the state variable so that the linear Kalman filter approach can be applied.

3.2 PARAMETER ESTIMATION

It has been shown that the general linear filtering and prediction problem can be solved by the Kalman filter technique. In practical hydrological applications of this method, the difficulties that usually prevail are that the system model and noise parameters are unknown. Basically, before applying the estimation procedure, the overall nature of the system operation must be characterized by

(i) the definition of the equations and associated parameters which describe the system (this is the model of the system);

(ii) the definition of the inputs which apply to the system.

The above analysis is known as system identification. This subject has received a great deal of attention in the literature. System identification pertaining to recursive methods is reviewed extensively by Åström and Eykhoff (1971); Söderström, Ljung and Gustavsson (1974); and Young, Shellswell and Neethling (1971).

Considering again the nth order ARMAX model equation 3.15 in a more general form

$$y(t) + a_1 y(t-1) + \ldots a_n y(t-n) = b_1 u(t-1) + b_2 u(t-2) + \ldots$$
$$\ldots + b_n u(t-n) + e(t) + c_1 e(t-1) + \ldots + c_n e(t-n) \qquad (3.20)$$

where $u(t)$ and $y(t)$ are the input and output respectively, and $e(t)$ is a sequence of $N(0,\sigma^2)$ Gaussian white noise. The above equation can be described in terms of polynomials in the backward shift operator z^{-1} where

$$\begin{aligned} z^{-1}y(t) &= y(t-1) \\ z^{-2}y(t) &= y(t-2) \\ &\vdots \\ z^{-n}y(t) &= y(t-n) \end{aligned} \qquad (3.21)$$

and the polynomial operators in z^{-1}, $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ are given by

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

$$B(z^{-1}) = b_1 z^{-1} + b_2 z^{-2} + \ldots + b_n z^{-n}$$

$$C(z^{-1}) = 1 + c_1 z^{-1} + \ldots + c_n z^{-n} \qquad (3.22)$$

Again it is assumed for convenience that the polynomials are of the same order, though this is not necessarily so. Equation 3.20 is thus given by

$$A(z^{-1})y(t) = B(z^{-1})u(t) + C(z^{-1})e(t) \qquad (3.23)$$

The identification problem in this case (having already specified the class of model) is to determine the model order n and to estimate the coefficients of the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$. These coefficients will be referred to as the model parameters.

Three recursive methods which have been applied to the estimation of the parameters of real-time flow forecasting (ARMAX) models are

(i) Recursive least squares

(ii) Instrumental variables (IV)

(iii) Approximate maximum likelihood (AML)

Derivations of these methods are to be found in Young, Shellswell and Neethling (1971), while a brief resumé is given below.

Recursive least-squares

For the recursive least squares to yield consistent (asymtotically unbiassed) estimates, that is $\hat{\theta}$ converges to θ in probability,

$$\hat{\theta} \rightarrow \theta$$

$$N \rightarrow \infty \qquad (3.24)$$

where θ and $\hat{\theta}$ are the true and estimated parameter vectors respectively, the noise term must be uncorrelated with the observations $y(t)$. Furthermore, if $y(t)$ is assumed to be a sequence of noisy observations, then the parameter estimates will be consistent if the noise term $C(z^{-1})e(t)$ is not autocorrelated (i.e. $C(z^{-1}) = 1$).

The estimator results from the minimisation of the loss function

$$V = \sum_{i=1}^{N} \varepsilon(i)^2 \tag{3.25}$$

where $\varepsilon(i)$ is the residual error between $y(t)$, the observed output and $\hat{y}(t|t-1)$, the predicted output at time t given the measurement up to time t-1,

$$\varepsilon(t) = y(t) - \hat{y}(t|t-1) \tag{3.26}$$

Defining the measurement and parameter vectors ϕ and θ respectively as

$$\phi(t) = [-y(t-1) \ \ldots \ -y(t-n), \ u(t-1) \ \ldots \ u(t-n)]$$

$$\theta^T = [a_1, \ \ldots \ a_n, \ b_1, \ \ldots \ b_n] \tag{3.27}$$

The recursive least squares equations for estimating θ are given by

$$\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1)\varepsilon(t+1) \tag{3.28}$$

where the gain $K(t)$ is updated by

$$K(t+1) = \frac{P(t) \ \phi^T(t+1)}{1 + \phi(t+1) \ P(t) \phi^T(t+1)} \tag{3.29}$$

and $P(t)$, the parameter estimation error covariance (standardised with respect to $\varepsilon(t)$) is updated by

$$P(t+1) = P(t) - \frac{P(t)\phi^T(t+1)\phi(t+1) \ P(t)}{1 + \phi(t+1) \ P(t)\phi^T(t+1)} \tag{3.30}$$

The residual error is given by

$$\varepsilon(t+1) = y(t+1) - \phi(t+1)\,\hat{\theta}(t) \tag{3.31}$$

The derivation of this recursive least squares algorithm is given in Appendix D.

These equations are clearly similar to the Kalman filter equations and the recursive least squares method is in fact equivalent to the Kalman filter for the time-invariant system model.

$$\theta(t) = \theta(t-1)$$

$$y(t) = \phi(t)\,\theta(t) + \varepsilon(t) \tag{3.32}$$

This is the approach taken in the Type I Kalman filter approach given by equations 3.17 and 3.18 where

$$X(t) = \theta(t) \qquad H(t) = \phi^{T}(t)$$

$$Z(t) = y(t) \qquad V(t) = \varepsilon(t) \tag{3.33}$$

For the case where $C(z^{-1}) \neq 1$ a recursive form of the generalised least-squares method (Clark, 1967) can be used. This method is not reported here.

Instrumental Variables

The instrumental variable (IV) algorithm in its non-recursive form has been known for some time in the statistical field (Kendall and Stewart 1961). The least-squares algorithm results in an inconsistent estimator when the noise term is correlated with the measurements. This problem is overcome by the instrumental variables algorithm. In this method an IV vector (or matrix) is introduced which is composed of elements chosen to be highly correlated with the unobservable noise-free process parameters of $A(z^{-1})$ and $B(z^{-1})$ but totally uncorrelated with the additive noise components given by $C(z^{-1})e(t)$ in equation 3.23. When combined with a

procedure similar to the least squares approach this method yields consistent estimates of the coefficients of $A(z^{-1})$ and $B(z^{-1})$. Young (1970) has given a recursive form for the IV method where $\hat{\theta}(t+1)$ and $\varepsilon(t+1)$ are given by equation 3.28 and 3.31 respectively, while the K vector and P matrix are given by

$$K(t+1) = \frac{P(t)\ z(t+1)}{1 + \phi(t+1)\ P(t)z(t+1)} \tag{3.34}$$

and

$$P(t+1) = P(t) - \frac{P(t)z(t+1)\phi(t+1)\ P(t)}{1 + \phi(t+1)P(t)z(t+1)} \tag{3.35}$$

where $z(t+1)$ is the vector of instrumental variables. The vector of instrumental variables is usually generated by an 'auxiliary' model of the form

$$z(t)^T = [z(t), \ldots z(t-n), u(t), \ldots u(t-n)] \tag{3.36}$$

$$A'(z^{-1})z(t) = B'(z^{-1})u(t) \tag{3.37}$$

where $A'(z^{-1})$ and $B'(z^{-1})$ are polynomials in the backward shift operator.

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

$$B'(z^{-1}) = b_o' + b_1'z^{-1} + \ldots + b_n'z^{-n} \tag{3.38}$$

the coefficients of which are chosen so as to ensure the desired properties of the instrumental variables. The IV method does not give the coefficients of the $C(z^{-1})$ polynomial; however if $\hat{y}_{IV}(t)$ is the predicted output given by the IV method, the noise component $\xi(t)$ of the process can be determined by

$$A(z^{-1})\hat{y}_{IV}(t) = B(z^{-1})u(t) \tag{3.39}$$

$$\xi(t) = y(t) - \hat{y}_{IV}(t) \tag{3.40}$$

$$\xi(t) = C(z^{-1})e(t) \tag{3.41}$$

Clearly a model of the noise process can be identified (eqn. 3.41) and combined with equation 3.39 to give the composite model in equation 3.23 (Young, Shellswell and Neethling, 1971). Frequently the approximate maximum likelihood (AML) method (Young, 1968) is used for estimating the parameters of the noise process and this approach is known as the instrumental variables - approximate maximum likelihood method (IV-AML).

Approximate maximum likelihood

Considering again the recursive least squares method, if measurements of the noise e(t) were possible, the recursive least squares algorithm (eqns. 3.28 to 3.31) could be used with the following measurement and parameter vectors:

$$\phi(t) = [-y(t-1), \ldots -y(t-n), u(t-1), \ldots u(t-n), e(t-1), \ldots \ldots e(t-n)]$$

$$\theta^T = [a_1, \ldots a_n, b_1, \ldots b_n, c_1, \ldots c_n] \tag{3.42}$$

The model is given by

$$y(t) = \phi(t)\,\theta + e(t) \tag{3.43}$$

(this is the same as for equation 3.31), so that an approximation to e(t) can be obtained by the residual error

$$e(t) \simeq \varepsilon(t) \doteq y(t) - \phi(t)\,\hat{\theta}(t-1) \tag{3.44}$$

This approach is known as the approximate maximum likelihood (AML) method. Though it has been interpreted as a maximum likelihood method it is not actually based on this method (Kendall and Stewart, 1961). The maximum likelihood estimator is obtained by maximising the log likelihood function

$$L(\varepsilon|\theta) = -\frac{1}{2\sigma_\varepsilon^2}\sum_{i\ 1}^{N}\varepsilon^2(i) - \frac{N}{2}\ln\sigma_\varepsilon^2 - \frac{N}{2}\ln 2\pi \qquad (3.45)$$

where σ_ε^2 is the variance of $\varepsilon(t)$ and N is the sample size. This function leads to a statistically consistent estimation scheme. The log likelihood function can be maximised (as a function of the parameters and residual variance) using gradient or hill climbing methods. The similarity in the AML and the maximum likelihood methods lies in the assumption that

$$C(z^{-1})\varepsilon(t) = A(z^{-1})y(t) - B(z^{-1})u(t) \qquad (3.46)$$

(see eqn. 3.23) and that the loss function

$$V = \sum_{i=1}^{N}\varepsilon(i)^2 \qquad (3.47)$$

is minimised (Åström and Bohlin, 1966). The AML parameter estimates have also been found to compare favourably with the maximum likelihood estimates, thus indicating the statistical efficiency of the AML method (Young, Shellswell and Neethling, 1971).

3.3 MODEL STRUCTURE ESTIMATION

So far it has been assumed that the model structure is known. The orders of the polynomials $A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$ of the ARMAX model, equation 3.23, specify the model structure. For hydrological systems, this information is not known a priori. Hence the fitting of an ARMAX model to hydrological data must include a model structure identification stage. Two methods have readily been used for this task; a technique based on obtaining an approximation to the systems impulse response function and due to Box and Jenkins (1970), and a trial and error method based on tests of the residuals of the fitted model. These methods are briefly outlined below.

Auto and cross-correlation analysis

Box and Jenkins (1970) have shown that the auto and cross-covariances of the input and output time series, u(t) and y(t) respectively, can be used to infer properties of the underlying process. The lag τ autocovariance function $\gamma_{yy}(\tau)$, of y(t) with mean μ_y,

$$\gamma_{yy}(\tau) = E\{(y(t)-\mu_y)(y(t+\tau) - \mu_y)\}, \tag{3.48}$$

gives the covariance between the values of the variate separated by τ timesteps. Similarly for a bivariate sequence, the covariance between y(t) and u(t) separated by a lag of τ timesteps and where μ_u is the mean of u(t), is given by the cross-covariance function

$$\gamma_{yu}(\tau) = E\{(y(t)-\mu_y)(u(t+\tau)-\mu_u)\} \tag{3.49}$$

For comparison purposes it is usual to standardize the auto and cross-covariance functions so that the magnitude of their coefficients are not dependent on the magnitudes (scale) of the variables. The standardized autocovariance function is known as the autocorrelation function $\rho_{yy}(\tau)$, and defined as

$$\rho_{yy}(\tau) = \frac{\gamma_{yy}(\tau)}{\gamma_{yy}(0)} \tag{3.50}$$

The corresponding cross-correlation function $\rho_{yu}(\tau)$ is defined as

$$\rho_{yu}(\tau) = \frac{\gamma_{yu}(\tau)}{\sqrt{(\gamma_{yy}(0)\gamma_{uu}(0))}} \tag{3.51}$$

The estimation of these functions (sample autocorrelation and cross-correlation functions) and their associated variances is treated in Appendix E.

The first stage in the Box-Jenkins method entails the interpretation of the autocorrelation functions of the input and output series, u(t) and y(t), to determine whether either of the series exhibit trends (non-stationarity). This is done by observing that if the coefficients of the autocorrelation function fail to damp out except for large lags, the series is non-stationary. The converse is true for stationary series.

Non-stationarity of the series can be removed by differencing. Thus if y(t)is the non-stationary series, a new series $\nabla y(t)$ can be formed from the differencing operation

$$\nabla y(t) = y(t) - y(t-1) \tag{3.52}$$

If the autocorrelation function still fails to damp out, further differencing may be carried out to form the series $\nabla^n y(t)$ where

$$\nabla^n y(t) = \nabla^{n-1} y(t) - \nabla^{n-1} y(t-1) \tag{3.53}$$

and n is the degree of non-stationarity. In the discussion that follows it will be assumed that the series u(t) and y(t) are stationary.

In general an ARMAX model such as equation 3.23 can be represented by the equivalent model

$$y(t) = v_1 u(t) + v_2 u(t-1) + \ldots + \zeta(t) \tag{3.54}$$

where $v_1, v_2, \ldots$ are the impulse response weights and $\zeta(t)$ is a noise term. If the input $u(t)$ is a sequence of white noise variables, an approximation to the impulse response is given by

$$v_\tau = \frac{\gamma_{uy}(\tau)}{\sigma_u^2} = \frac{\rho_{uy}(\tau)\sigma_y}{\sigma_u} \qquad (3.55)$$

If however the input is not a white noise sequence, it can be 'prewhitened' where the series $u(t)$ is lineary transformed to the white noise sequences $e_u(t)$. The two series are related by the auto-regressive moving average (ARMA) prewhitening filter

$$A'(z^{-1})u(t) = B'(z^{-1})e_u(t) \qquad (3.56)$$

Thus before the impulse response (eqn. 3.55) can be estimated, the parameters of equation 3.56 must be estimated. The orders of the polynomials $A'(z^{-1})$ and $B'(z^{-1})$ can be obtained by inspection of the autocorrelation and partial auto-correlation functions of the series $u(t)$. The general procedure is given below

(a) Order of the AR process

For an AR process model such as

$$u(t) = \alpha_1 u(t-1) + \alpha_2 u(t-2) + \ldots + \alpha_n u(t-n) + e(t) \qquad (3.57)$$

where $e(t)$ is a white noise sequence, the autocorrelation function of the series $u(t)$ satisfies a difference equation with the same autoregressive parameters, i.e.

$$\rho_{uu}(\tau) = \alpha_1 \rho_{uu}(\tau-1) + \alpha_2 \rho_{uu}(\tau-2) + \ldots \alpha_n \rho_{uu}(\tau-n)$$

$$\tau > 0 \qquad (3.58)$$

(Box and Jenkins, 1970). The set of n equations obtained by replacing τ by 1,2, ... n are known as the Yule-Walker equations. If a further subscript is added to the AR parameters to indicate the order of the AR process model (equation 3.57), then, for an nth order process the set of equations

$$\begin{bmatrix} 1 & \rho(1) & . & . & \rho(n-1) \\ \rho(1) & 1 & . & . & \rho(n-2) \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \rho(n-1) & \rho(n-2) & . & . & 1 \end{bmatrix} \begin{bmatrix} \alpha_{n1} \\ \alpha_{n2} \\ . \\ . \\ . \\ 1 \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ . \\ . \\ . \\ \rho(n) \end{bmatrix} \qquad (3.59)$$

is obtained. The solution of the above matrix equation gives $\alpha_{n1}, \alpha_{n2}, \cdots \alpha_{nn}$. The partial autocorrelation function is given by the coefficients $\alpha_{11}, \alpha_{22}, \alpha_{33}, \ldots$ which are obtained from the solution of the Yule-Walker equations for assumed AR process models of orders 1,2,3, ... If the autoregressive process is actually of order n, the partial autocorrelation function α_{jj} will be non-zero for j less than or equal to n and zero for j greater than n, that is, the partial autocorrelation function has a 'cut-off' beyond lag n. This property can be used to determine the order of the AR process.

(b) Order of the MA process

The autocovariance function of the MA process model

$$u(t) = e(t) + \beta_1 e(t-1) + \ldots + \beta_m e(t-m) \qquad (3.60)$$

where e(t) is a white noise sequence distributed as $N(0,\sigma^2)$ is given by

$$\gamma_{uu}(\tau) = E\{(e(t) + \beta_1 e(t-1) + \ldots + \beta_m e(t-m))(e(t) + \beta_1 e(t-1) + \ldots + \beta_m e(t-m))\} \qquad (3.61)$$

The variance of this process is

$$\gamma_{uu}(0) = (1 + \beta_1^2 + \beta_2^2 + \ldots \beta_m^2)\sigma^2 \qquad (3.62)$$

and its autocorrelation function is given by

$$\rho_{uu}(\tau) = \begin{cases} \dfrac{\beta_\tau + \beta_1\beta_{\tau+1} + \ldots + \beta_{m-\tau}\beta_m}{\rho_{uu}(0)} & \tau = 1,2, \ldots m \\ 0 & \tau > m \end{cases} \qquad (3.63)$$

Thus the autocorrelation function for a MA process has a cut-off at lag m, a property which can be used to determine the order of the MA process.

For a mixed (ARMA) process the orders of the separate AR (n terms) and MA (m terms) processes can be obtained by reference to both the autocorrelation function and the partial autocorrelation function of the input u(t). The autocorrelation function will consists of damped exponentials and/or damped sine waves after the first (n-m) lags.

Finally, estimates of the parameters of an ARMA process model are obtained (separately for the MA and the AR parameters) using approximate methods based on the autocovariances (Box and Jenkins, 1970, Appendix A6.2). These parameter estimates, though statistically inefficient, suffice as parameters for a prewhitening filter.

Having determined the polynomials $A(z^{-1})$ and $B(z^{-1})$, the (correlated) input series u(t) is prewhitened using the filter

$$e_u(t) = \frac{A'(z^{-1})}{B'(z^{-1})}\, u(t) \qquad (3.64)$$

where $e_u(t)$ is the transformed series. Similarly, the output y(t) is also transformed to a new series y'(t), i.e.

$$y'(t) = \frac{A'(z^{-1})}{B'(z^{-1})}\, y(t) \qquad (3.65)$$

Corresponding to equation 3.55, the approximate system impulse response weights are then given by

$$v_\tau = \frac{\gamma_{e_u y'}(\tau)}{\sigma_{e_u}^2} \tag{3.66}$$

In practice the above procedures would be carried out using the estimated auto and cross-correlation functions (the sample auto and cross-covariances) presented in Appendix E. The impulse response function obtained can be used to select the appropriate ARMAX model structure. Unfortunately this is not an easy task and one that is best learned by experience . Box and Jenkins (1970, p 349) gives representative examples of impulse response functions for several ARMAX models.

Residual Diagnostic Checking

The second method for determining the appropriate ARMAX model structure is based on diagnostic checks applied to the 1-step ahead prediction error (residual) $\varepsilon(t)$ of the prediction model, where

$$\varepsilon(t) = y(t) - \hat{y}(t|t-1) \tag{3.67}$$

On the presupposition of a particular ARMAX model structure, the prediction model itself is applied to a set of test data to yield a sequence of residuals $\varepsilon(t)$. Diagnostic checks can consequently be applied to these residuals to verify model adequacy. The selection of the appropriate ARMAX model structure can thus be made through the comparison of residual diagnostic checks for prediction models with competing ARMAX model structures.

A fundamental requirement for model adequacy is that $\varepsilon(t)$ is not autocorrelated, i.e.,

$$\rho_{\varepsilon\varepsilon}(\tau) = 0 \quad ; \quad \tau \neq 0 \tag{3.68}$$

If $\varepsilon(t)$ is autocorrelated, the information in one residual can be used to infer something about the next residual, indicating that there is scope for improvement of the prediction model. Model adequacy also requires that the residual is not cross-correlated with past values of the input series $u(t)$, i.e.

$$\rho_{u\varepsilon}(\tau) = 0; \quad \tau \geq 0 \tag{3.69}$$

If $\varepsilon(t)$ is cross-correlated with $u(t)$, some of the information in the input series $u(t)$ has not been used by the model in obtaining the prediction, again suggesting inadequacy of the prediction model.

The course of action for identifying a suitable ARMAX model structure is firstly to select several prediction models of the same type but with different ARMAX model structures (different orders of the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$). These models are then calibrated over test data and sequences of 1-step ahead prediction residuals generated. The sample (estimated) auto and cross-correlations $r_{\varepsilon\varepsilon}(\tau)$ and $r_{\varepsilon u}(\tau)$ (See Appendix E) are then examined and the model with the residuals best satisfying equations 3.68 and 3.69 is selected. This procedure then is one of trial and error where the performances of different ARMAX models are compared, and the best one selected.

The problem that remains is to decide when the sample auto and cross-correlations are to be assumed zero. Box and Jenkins (1970) show that the upper bound for the standard error of the sample auto-correlation function $r_{\varepsilon\varepsilon}(\tau)$ is $1/\sqrt{N}$. This may be interpreted as; based on normality, if the first 20 values of $r_{\varepsilon\varepsilon}(\tau)$ are computed, then on average 19 out of 20 values can be expected to lie between $\pm 1.96/\sqrt{N}$ (the 95% confidence interval). This possibility of spurious coefficients spotlights one of the difficulties in interpreting the sample autocorrelation function. Furthermore, while $1/\sqrt{N}$ produces an upper bound for the standard error or $r_{\varepsilon\varepsilon}(\tau)$, for low lags the standard error of $r_{\varepsilon\varepsilon}(\tau)$ may be less than $1/\sqrt{N}$, and particularly for lag one, a value of $r_{\varepsilon\varepsilon}(1)$ inside the range $\pm 1.96/\sqrt{N}$ may still be significantly different from zero. Box and Pierce (1970) show that the statistical distribution properties of the sample autocorrelations at low lags depend heavily on the estimated parameters of the model. Furthermore ripples in the sample autocorrelation function at low lags can arise because of high induced correlation between these estimates.

The sample cross-correlation function for two white noise sequences

will also tend to zero for large N with a standard error of $1/\sqrt{N}$ (Box and Jenkins, 1970). However if the input u(t) is not a white noise sequence, and assuming that $\varepsilon(t)$ is, the sample cross-correlation coefficients $r_{\varepsilon u}(\tau)$ may have approximately the same auto-correlation function as does the input signal u(t). Thus for the two series which are not cross-correlated, large cross-correlation coefficients may exist which are actually spurious. This effect is eliminated however if the cross-correlations are computed with a prewhitened input.

Portmanteau Tests

Rather than considering the autocorrelation coefficients individually, an indication of model adequacy can be gained by considering the first K autocorrelation coefficients taken as a whole. Box and Pierce (1970) have shown that if a fitted AR model is appropriate

$$Q = N \sum_{i=1}^{K} r_{\varepsilon\varepsilon}{}^2(i) \qquad (3.70)$$

is approximately distributed as $\chi^2(K-m)$ where N is the number of data used in obtaining the sample autocorrelations and m the number of AR terms. A hypothesis test of model adequacy (the Q-test) is then obtained by referring the computed value of Q to a table of percentage points of χ^2 with K-m degrees of freedom. This test can further be extended to ARMA models by the following argument. If the outputs of an AR process and an ARMA process are both generated by the same white noise sequence e(t), and if an AR and an ARMA model are then consequently fitted to the respective output sequences, the autocorrelations of their residuals will both be similar. A correctly fitted ARMA model will have an auto-correlation function distributed as if the model had been an AR model with m+n AR terms, where m and n are the number of AR and MA terms respectively. Thus for an ARMA model the hypothesis test for model adequacy is obtained by comparing the computed value of Q to the table of percentage points of χ^2 with K-m-n degrees of freedom. To extend further this test to the ARMAX model (eqn. 3.23) the Q-test is applied to the

noise-model component of the model (Box and Jenkins, 1970). Equation 3.23 can be written as

$$A(z^{-1}) = B(z^{-1}) + \zeta(t) \tag{3.71}$$

where $\zeta(t)$ is the noise model given by

$$\zeta(t) = C(z^{-1})e(t) \tag{3.72}$$

The Q-test consists then of comparing Q with the percentage points of χ^2 for K-ℓ degrees of freedom, where ℓ is the number of parameters in the noise model.

As in the Q-test the S-test for the cross-correlation between the input and the prediction error can be used as a test of model adequacy. The S-test is given by

$$S = N \sum_{i=1}^{K} r_{\varepsilon u}^2 (i) \tag{3.73}$$

which is approximately distributed as χ^2 with K n-m degrees of freedom. The number of parameters in the polynomials $A(z^{-1})$ and $B(z^{-1})$ is given by n and m respectively. However, to avoid the problem of the autocorrelation of the input sequence influencing the cross-correlation function, the input sequence should first be prewhitened.

The above residual diagnostic checks are based on the examination of the auto and cross-correlation coefficients individually, or as a whole by means of the Q and S-tests. These checks are used to aid the selection of a model with a trial and error approach.

A further criterion which can be used in the model selection procedure is the loss function

$$V = \sum_{t=1}^{N} \varepsilon(t)^2 \tag{3.74}$$

where $\varepsilon(t)$ is the prediction error residual. Here, competing models are selected, and for a given sequence of test data, corresponding sequences of $\varepsilon(t)$ are generated. The model resulting in the lowest value of V is assumed to be the model with the appropriate structure.

3.4 SUMMARY

An outline of some of the linear filtering and prediction techniques which have recently been applied to real-time flow forecasting has been presented in this chapter. The method central to modern filter theory is the Kalman filter. The wide adaptability and computational elegance of the Kalman filter is largely due to the state-space formulation. Many rainfall-runoff models can be formulated as state-space models, and these can then be classified as being either of type I, II or III as described in Section 3.1.

For deterministic systems, the state contains the information required to describe the system response to one or more inputs. However, for stochastic systems, probabilistic concepts have to be introduced to describe the state. The Kalman filter provides a statistically efficient method for recursively estimating the system state, while taking into account the system noise and measurement noise. The algorithm also provides the updated state estimates and the error covariance matrix of the state estimates, while the properties of the innovation sequence (see eqn. 3.14) can be used to evaluate the performance of the filter.

Before applying the Kalman filter estimator, the system behaviour must be characterised in mathematical terms, and this is known as system identification. System identification, when the system model is assumed to be an ARMAX model, involves a model structure estimation stage and a parameter estimation stage. Three recursive parameter estimation techniques which have recently been applied to rainfall-runoff modelling were outlined, namely recursive least squares, instrumental variables and approximate maximum likelihood. The similarities between the recursive equations of these algorithms and the recursive equations of the Kalman filter are clearly apparent.

Two model structure estimation techniques were described in Section 3.3. These were the Box and Jenkins auto and cross-correlation

analysis and a residual diagnostic checking method. The Box and Jenkins analysis, which is applicable to a single input-single output process only, involves the use of the sample auto and cross-correlation functions of the bivariate time series to infer properties of the underlying process. In particular, non-stationarity in a time series can be detected and removed as necessary. An approximation to the impulse response of the system can be obtained from the sample cross-correlation function between the prewhitened input and the output. The appropriate ARMAX model structure can be estimated from the system impulse response.

The residual diagnostic checking approach involves the comparison of the residual error sequence obtained from a set of test data using a selection of proposed ARMAX models (each with a different model structure). The criteria for an appropriate model are that the residual sequence should not be autocorrelated and should not be cross-correlated with the input sequence. In practice, the sample auto and cross-correlation functions may be tested by referring the individual auto and cross-correlation coefficients to the 95% confidence interval, or by considering the first K coefficents as a whole by using the portmanteau tests described. Competing ARMAX models may also be selected simply using a criterion of minimum square residual error.

Given a particular ARMAX model structure, the residual diagnostic tests can be used for verifying the performance of the parameter or state estimator algorithm.

4. REAL-TIME CONTROL

4.1 FEEDBACK AND FEEDFORWARD CONTROL

The field of control theory is both extensive and specialised and therefore it is not practicable to give a comprehensive outline of deterministic and stochastic modern control theory in this presentation. Introductory treatments of deterministic control theory are found in Bryson and Ho (1969) and Athans and Falb (1966), while stochastic control theory is treated by Åström (1970). In this chapter only a simple single-input, single-output process control problem will be considered. In relation to this problem several types of stochastic controller will be introduced, finally leading to the self-tuning regulator which forms the basis of the present study.

Consider a process *M* with a single input and single output as shown in fig. 4.1.

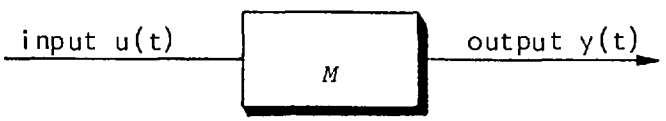


Fig. 4.1 Single-input, single-output process to be controlled

The process input u(t) is the control variable, which is to be adjusted so that the output y(t) is equal to a set-point. The set-point value is the target or reference for the output. When this set-point is constant the above control is referred to as a regulator problem, and when the set-point is variable it is known as a servo problem. In a controlled process the function of the controller can be defined as the collection of all necessary information (measurements) from the system being controlled and the use of these data to manipulate some of the system inputs (in this case only one) in order to bring about some desired process performance.

The controller design is based upon two general principles: the feedforward principle and the feedback principle.

The feedforward principle is employed to cancel the effect of disturbances on the process behaviour. As shown in fig. 4.2, the

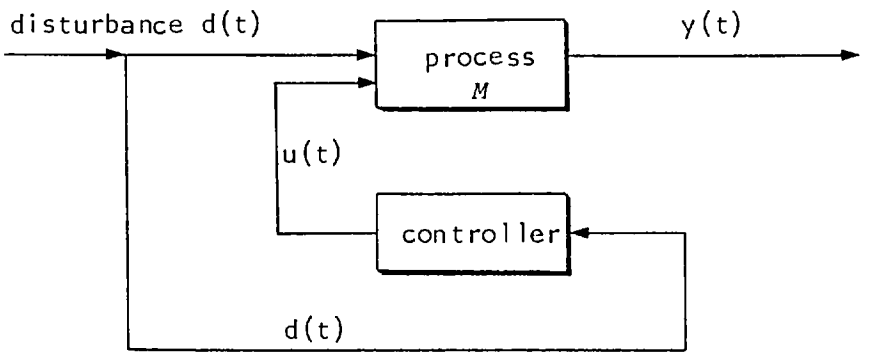


Fig. 4.2 Feedforward control

observed disturbance is used by the controller to adjust the input u(t) accordingly, so that in principle the effect of the disturbance d(t) on the process is eliminated and y(t) is set equal to the set-point. The operation of a feedforward controller suffers from the following disadvantages:

(i) it requires accurate measurement of the disturbance;

(ii) it requires accurate modelling of the process as the output error (deviation of y(t) from the set-point) is not used in correcting any modelling errors;

(iii) there is no self-adjusting action, i.e. if the process changes the controller will be incorrect.

The feedback controller, contrary to the feedforward controller, is designed to cancel the effect of the unobserved disturbance d(t) on the process output y(t) as shown in fig. 4.3.

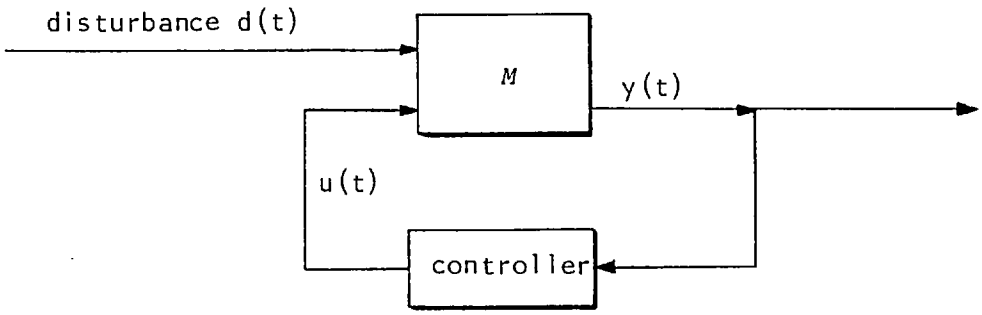


Fig. 4.3 Feedback controller

This controller uses the observed process output y(t), or the error in the output(difference between the output and the set-point) to determine the input u(t) required to minimise the output error in the light of unobservable disturbances acting on the process. Clearly the feedback controller does not suffer from disadvantages (i) and (ii) of the feedforward controller, and furthermore the feedback controller could easily be designed to be adaptive, i.e. if the process changes,the controller will accordingly adjust its performance in order to minimise the output error.

If the process being controlled has both observable disturbances and an observable output, the best controller to use would be a feedback-feedforward controller. The control action can then take into account disturbances (which can be regarded as uncontrollable process inputs) and past errors in the output.

4.2 STOCHASTIC ADAPTIVE CONTROL

Frequently controllers are needed which should not require manual adjusting and should be able to handle changes in system dynamics and/or noise characteristics. Controllers which fall into this category are referred to as adaptive, self-organizing, self-adjusting or self-optimizing controllers. The term adaptive will be used forthwith.

The basic functions that are common to most adaptive controllers are:

(i) the estimation of unknown system parameters;

(ii) the evaluation of a performance index from measurement, i.e. the mean square error between output and set-point;

(iii) the derivation of a control input;

(iv) the on-line modification of control parameters.

In Section 3.1 it has been shown that the rainfall runoff process can be formulated as an ARMAX process model. Wood (1979) has also shown that this type of model can also be applied to real-time river flow routing. A class of controllers based on the ARMAX model (which may be in a state space formulation) are known as stochastic adaptive controllers, of which an excellent survey has been given by Wittenmark (1975). A stochastic adaptive controller takes into account variations in the parameters of a stochastic model of the process to be controlled. A block diagram which can represent many stochastic adaptive controllers is shown in fig. 4.4.

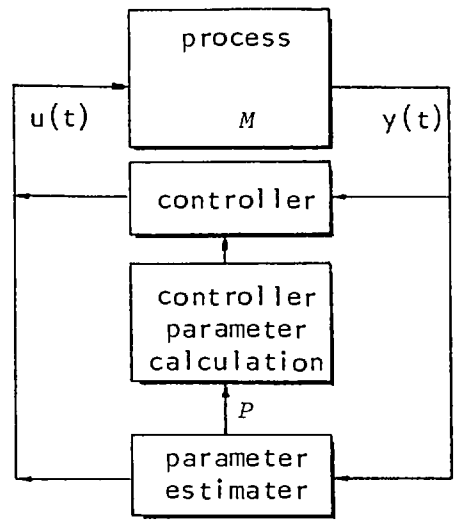


Fig. 4.4 Block diagram of the stochastic adaptive controller

The input to the process is u(t), determined by the controller, while y(t) is the process output. The process which is assumed to be unknown is modelled by an ARMAX model and the estimator attempts to find the parameters of this model. The information vector P contains the information (i.e. the model parameters) used to calculate the parameters of the controller. The feedback controller then uses the past observed output(s) and the estimated controller parameters to calculate the next process input. The above control procedure relies either on the 'certainty equivalent principle' (Bar-Shalom and Tse, 1974) or the 'separation principle' (Witsenhausen, 1971), and the corresponding controllers are referred to as certainty-equivalent controllers or cautious controllers, respectively. The certainty equivalent principle is valid if (a) for a known process the optimal controller can be obtained in terms of the known process parameters, and (b) the certainty-equivalent controller can be obtained for unknown process, by substituting the unknown true process parameters with estimated process parameters. In fig. 4.4 this means that the information vector P, which is transferred from the estimator to the controller, is the estimated parameter vector $\{\hat{\theta}\}$, i.e. $P = \{\hat{\theta}\}$. The certainty equivalent controller does not take into account that the estimated parameters may not be equal to the true parameters.

The separation principle is valid if the parameters of the process are first estimated, and based on these parameter estimates and their errors, the parameters of the controller are then estimated. Thus there is a separation between the estimation of the process parameters and of the controller parameters. The resulting controller is known as a cautious controller as the controller is aware of errors in the parameter estimates. In this case the information vector P is taken as $P = \{\hat{\theta}, p\}$, where p is a measurement of the uncertainty in the parameter estimates $\{\hat{\theta}\}$.

The behaviour of a controller will depend largely on the loss function (performance index or cost function) to be minimised. If the loss function only takes into account the previous measurements and does not assume that further information will be available, then the resulting controller is non-dual in the sense of Feldbaum (1960). The minimisation of the loss function one step ahead (independent of future observations) will give a non-dual controller. However if the loss function is dependent on future observations (in terms of the probability distribution of the future observations given information up to the actual time), the controller will be dual. A dual controller minimising its loss functions several steps ahead has the possibility of making control actions which may so disturb the process as to obtain improved parameter estimates. The dual controller has to make a compromise between parameter estimation and control when good parameter estimation requires large changes in control inputs and good control requires only small changes in control inputs.

There are two conditions for which a dual controller may be required.

(i) If the parameters of the process are rapidly time varying, a cautious non-dual controller may assume that the parameters are always poorly estimated and consequently the control will also be poor.

(ii) If the parameters of the process to be controlled are initially unknown a dual controller may be required to obtain good parameter estimates rapidly.

It was found however that for the hydrological simulation experiments carried out, the processes were slowly time varying (relative to the chosen sampling interval), and that good process parameter estimates were quickly obtained. Furthermore dual controllers are more complex than non-dual controllers, so that in practice when dual controllers are employed they are usually suboptimal. The possible choice of controllers was thus limited to non-dual types; a further choice between a certainty-equivalent or a cautious controller was then necessary.

Cautious controllers based on ARMAX models have two major disadvantages (Wittenmark, 1975) which would be of significance when considering control problems encountered in real-time water resource systems management. These are as follows:

(i) The problem becomes increasingly complex as the delay (time lag) between the input and output increases.

(ii) If the parameter estimates are uncertain then the control input u(t) may become very small so that the system is not greatly excited resulting in even poorer estimates. Eventually the controller may 'turn-off' (Åström and Wittenmark, 1971).

Also, cautious controllers based on state space models are difficult to realise because state and parameter estimates must be obtained simultaneously. Based on the above arguments, for this study the choice of controller to be employed was limited to certainty-equivalent, non-dual types.

4.3 THE SELF-TUNING REGULATOR

A class of stochastic non-dual certainty-equivalent controller has been developed by Åström and his co-workers (Åström and Wittenmark, 1973) which are designed to control processes with time-invariant but unknown parameters. Known as the self-tuning regulator, variations of this algorithm have been both the subject of extensive theoretical investigations and practical application (Wittenmark, 1973; Wellstead, 1976). Self-tuning regulators have been applied successfully to several different industrial processes requiring control by on-line computer. Two well-documented applications of the self-tuning regulator are the following:

(i) the control of the moisture content in a paper manufacturing process to improve the economy of production and quality of the paper (Borisson and Wittenmark, 1973);

(ii) the control of ore input to an ore crusher in order to maintain a high constant output and thus a high production level (Borisson and Syding, 1974).

A more general class of regulators based on minimum variance control laws which are derived on the assumption that the process parameters are known, will first be considered. The process to be controlled is described by the ARMAX model

$$A(z^{-1})y(t) = B(z^{-1})u(t-k) + C(z^{-1})e(t) \qquad (4.1)$$

where $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ are polynomials in the backwards shift operator (see eqn. 3.21) given by

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n} \\ B(z^{-1}) &= b_1 z^{-1} + \ldots + b_n z^{-n} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \ldots + c_n a^{-n} \end{aligned} \qquad (4.2)$$

and u(t) is the input or control variable, y(t) the output and e(t) a sequence of Gaussian white noise variables distributed as $N(0,\sigma^2)$. The model has a delay of k time steps. Again, for notational convenience the polynomials are assumed to have the same number of parameters. The objective of the control is to minimise the variance of the output about zero (i.e. y(t) is set to zero at each time step) and the loss function (which is the performance index) to be minimised is

$$V_1 = E\{y^2(t)\} \tag{4.3}$$

The optimal strategy (which is the control law) when the parameters of $A(z^{-1}),B(z^{-1})$ and $C(z^{-1})$ are known is obtained by introducing the following identity (Åström, 1970):

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-k}G(z^{-1}), \tag{4.4}$$

where

$$\begin{aligned} F(z^{-1}) &= 1 + f_1 z^{-1} + \ldots + f_{j-1} z^{-k+1} \\ G(z^{-1}) &= g_o + g_1 z^{-1} + \ldots + g_{n-1} z^{-n+1} \end{aligned} \tag{4.5}$$

It is shown in Appendix C that the control law which minimizes V is given by

$$u(t) = \frac{-G(z^{-1})}{B(z^{-1})F(z^{-1})}\, y(t) \tag{4.6}$$

This optimal control law is known as the minimum variance strategy. At time t the control u(t) is a linear function of the observed outputs up to and including time t, that is, y(t), y(t-1), ... , and of all the past control signals u(t-1), u(t-2), ... If the parameters of the model (eqn. 4.1) are known, using the identity equation (eqn. 4.4) the parameters of the polynomials $F(z^{-1})$ and $G(z^{-1})$ (eqns. 4.5) can be determined, and hence the control can be evaluated from the control law (eqn. 4.6). It is shown in Appendix C that the control error, which is the magnitude of the output, is given by

$$y(t) = F(z^{-1})e(t) \qquad (4.7)$$

where σ is the standard deviation of e(t).

For hydrological applications, the processes must be assumed unknown, so the control cannot be directly calculated from known parameters. It is possible however to use the certainty-equivalence principle and estimate the parameters of the process, assume that the estimated parameters are the true parameters, and so calculate the control from equation 4.6. A block diagram of this class of regulator is shown in fig. 4.5.

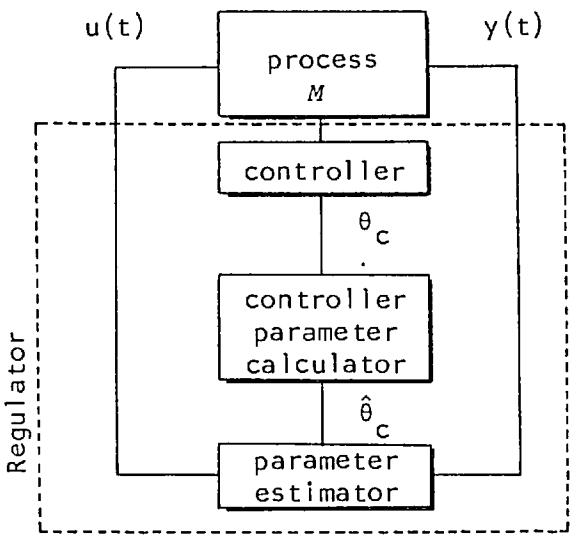


Fig. 4.5 Block diagram of the regulator

The regulator is composed of three parts: a parameter estimator, which estimates the unknown parameters of the process *M*, a controller (control law) which determines the control input, and a parameter calculator which relates the estimated process parameters $\hat{\theta}$ to the controller parameters θ_c. By the certainty-equivalence principle, the parameters θ_c are assumed to be known without error. As outlined in Section 3.2 several estimation algorithms are available for solving the parameter estimation problem. The simplest of these, the recursive least squares algorithm is used for the self-tuning regulator. When the control inputs are

generated by a feedback control they are correlated with the disturbances e(t), and it is not obvious that all the parameters of the model can be determined. As an example, consider the process represented by the first order process model

$$y(t) + ay(t-1) = bu(t-1) + e(t) \tag{4.8}$$

with a unit input delay (k=1). A feedback control law which sets the output y(t) to zero (i.e. the set-point is zero) is given by

$$u(t) = \omega\, y(t) \tag{4.9}$$

If the parameters a and b are known the parameter for the regulator would be $\omega = a/b$. This corresponds to the minimum variance control law. When the parameters a and b are unknown the estimated parameters a and b replace the true parameters such that $\alpha = \hat{a}/\hat{b}$. If least squares is used to obtain the parameter estimates, the loss function given by

$$V_{LS} = \sum_{t=1}^{N} [y(t) + ay(t-1) - bu(t-1)]^2 \tag{4.10}$$

is minimised with respect to a and b for N time steps. However as the input is generated by a feedback law (eqn. 4.9) the parameters a and b cannot both be estimated. This can be shown by multiplying both sides of equation 4.9 by a factor f such that

$$fu(t) - \omega f y(t) = 0 \tag{4.11}$$

Subtracting this expression from the summation term in equation 4.10 gives a new loss function

$$V_{LS} = \sum_{t=1}^{N} [y(t) + (a + \omega f)y(t-1) - (b + f)u(t-1)]^2 \tag{4.12}$$

which when minimised with respect to a and b give solutions $\hat{a}$ and $\hat{b}$ on a linear manifold (vector space) with the same value for V_{LS}. To overcome this problem the parameter b is fixed (Wittenmark, 1973) at a pre-determined value, and only the parameter a is estimated.

An alternative to estimating the process model parameters and then computing the parameters of a control law (which for all but the simplest cases will require the solution of a set of linear equations) is to estimate the parameters of the regulator directly. For the least-squares method to give consistent estimates, the assumption $C(z^{-1}) = 1$ in equation 4.1 must be made. By using the identity given by equation 4.4 with $C(z^{-1}) = 1$, $A(z^{-1})$ can be eliminated from equation 4.1 to give

$$y(t) - G(z^{-1})y(t-k) = B(z^{-1})F(z^{-1})u(t-k) + \varepsilon(t) \qquad (4.13)$$

where $\varepsilon(t) = F(z^{-1})e(t)$. If m and ℓ are the number of parameters in the polynomials $G(z^{-1})$ and $B(z^{-1})F(z^{-1})$ respectively equation 4.13 can be written as

$$y(t) + a_1y(t-k) + \ldots + a_my(t-k-m+1)$$

$$= b_1u(t-k) + b_2u(t-k-1) + \ldots + b_\ell u(t-k-\ell+1) + \varepsilon(t) \qquad (4.14)$$

The minimum variance control law is then simply

$$u(t) = -\frac{1}{b_1}[a_1y(t) + \ldots + a_my(t-m+1) - b_2u(t-1) - \ldots$$

$$\ldots -b_\ell u(t-\ell+1)] \qquad (4.15)$$

Hence the parameters of the model given by equation 4.14 (which is in fact a k-step ahead predictor of y(t)) can be estimated by the least squares method and used to compute the control input u(t) given by equation 4.15. This is known as the self-tuning regulator. As shown above, one parameter must be fixed to avoid possible difficulties with the parameter estimation. In this case the parameter b_1 is assumed to be given, and Åström and Wittenmark (1973) have shown that the choice of this parameter is not critical.

The parameters of equation 4.14 are estimated using data up to time t, i.e. the loss function V is minimised where

$$V = \sum_{k=1}^{t} \varepsilon(t)^2 \qquad (4.16)$$

The recursive least-squares algorithm is used, and parameter estimates are obtained at each time step with only moderate computations and computer storage requirements. The penalty for estimating the parameters of equation 4.14 instead of the actual model parameters (eqn. 4.1) is that k more parameters must be estimated.

The basis of the self-tuning principle is that the estimated parameters of the regulator will converge in the limit to the parameters of the minimum variance regulator which could be obtained had the parameters of the process to be regulated been known. For the case where $C(z^{-1}) = 1$, the least squares technique results in a consistent estimator, however when $C(z^{-1}) \neq 1$ the self-tuning regulator also gives good results. The self-tuning property depends on a delicate balance of the inconsistency in the least squares estimate and the fact that the control law is determined from an erroneous model (Åström, 1974). The self-tuning regulator which has the correct number of parameters converges to the optimal regulator when the following relationships hold (Appendix B).

$$\gamma_{yy}(\tau) = E\{y(t)y(t+\tau)\} = 0 \quad ; \tau = k, \ldots k+m-1$$

$$\gamma_{yu}(\tau) = E\{y(t+\tau)u(t)\} = 0 \quad ; \tau = k, \ldots k+\ell \qquad (4.17)$$

The convergence properties in fact apply if there are too many parameters (Åström and Wittenmark, 1973) in the self-tuning regulator; however, by the principle of parsimony, in practice an excess of parameters may result in poor regulation.

4.4 SUMMARY

The objective of this chapter has been to present a control problem in a very simple setting so that the basic concepts of stochastic control could be developed with very little mathematical sophistication. The treatment of the problem was limited to systems with one control input and one output, and the problem to be solved was to find an admissible control law that minimises the variance of the output about the set-point. The feedback and feedforward principles, which are fundamental to control were firstly outlined. It was concluded that for systems with both observable disturbances (uncontrollable inputs) and an observable output, the best controller would be a feedback-feedforward controller.

Stochastic adaptive controllers discussed in Section 4.2 may depend on either the certainty equivalent principle or the separation principle. Both these controllers rely on the estimation of the parameters of the system model. These parameters are used to calculate the parameters of the certainty equivalent controller without taking into account parameter estimation errors. The cautious controller (using the separation principle) on the other hand takes into account errors associated with model parameter estimation.

Stochastic adaptive controllers may also be either dual or non-dual controllers. Dual controllers assume that future observations will be available and hence control inputs may be so generated as to produce system perturbations that ensure a system response which will lead to good parameter estimation. Non-dual controllers do not assume that further information will be available.

From the above available choice of controllers, the certainty equivalent non-dual class of controller was selected, as these are the simplest type, and they were later found to perform adequately when applied to simulated water resource subsystem control problems. The self-tuning regulator, which falls into this class, was then described. The self-tuning regulator consists of a recursive least-squares parameter estimator for the process model, and a minimum

variance control law. The estimated parameters (divided by b_1) give the parameters of the control law. The performance of the self-tuning regulator may also be easily evaluated by using the statistical properties given in equations 4.17.

Two of the important features of the self-tuning regulator are that it requires minimal computer memory and processing and that the parameters are recursively estimated at each time step. The self-tuning regulator is also closely related to the self-tuning predictor to be described in the next chapter.

5. REAL-TIME STREAMFLOW FORECASTING

5.1 RAINFALL-RUNOFF MODELS

The development of mathematical models of catchment behaviour has paralleled the development of the digital computer, and has for more than twenty years commanded a dominating role in hydrological research. A mathematical model of catchment behaviour is inevitably a simplified representation of a highly complex system and may be broadly classified under two categories; Conceptual and Black-box.

Conceptual models characterize the dynamic behaviour of the hydrological system with a mechanistic approach (Harris, 1976) in which the model and its calibration is due to both (a) the approximation of the internal physical mechanisms and laws governing the system dynamics and (b) the reference to data pertaining to the system response. The data may be input-output (rainfall-runoff) data necessary for model calibration and verification or parameters required to describe the physical and dynamic characteristics of the catchment. In principle, the user of a conceptual model may have to hand a valuable aid in understanding the physical behaviour of the system. Unfortunately, in terms of mathematical description, runoff from a natural catchment is so complex that a conceptual model based on continuum mechanics (hydraulics) would require an amount of data which in practical circumstances could not be obtained. The necessary computations may demand a very large memory allocation and an extensive execution time on a mainframe computer. An example of this type of model is given by Freeze (1972). Conceptual models however are usually simplified into an arrangement of a relatively small number of elements, each of which is itself a simple representation of a physical process, such that the overall effect is to simulate the catchment behaviour. A well-known example of this type of model is the Stanford Watershed Model (Crawford and Linsley, 1966).

Black-box models, often referred to as empirical models, infer the system input-output relationship from observed data, and hence unlike the conceptual model, the internal physical representation is bypassed. A black-box model often applied in catchment hydrology is the regression model (Yevdjevich, 1964).

Following Clarke's (1973) classification of hydrological models, conceptual and black-box catchment models may further be subdivided into stochastic or deterministic models. Stochastic models take into account random variables having probabilistic distributions, while the variables of the deterministic model are regarded as free from random variation.

Under the umbrella of the above four classes of models there is a plethora of catchment models for the hydrologist to select for his particular application. This may be either one of simulation or of real-time forecasting. Simulation models are used to (a) determine catchment behaviour for selected rainfall input sequences (design storms) for planning and design purposes, (b) to reconstruct missing flow data for which period rainfall data is available, or (c) to identify the effect of land use change on catchment behaviour.

Simulation models, having been calibrated, convert a rainfall input to a runoff output without reference to observed flow (reference to observed flow may be required during the calibration and model verification stage). Real-time forecasting models are used to forecast runoff in an operational role. Simulation models can also be adapted for real-time flow forecasting (e.g. WMO, 1975) though this may be unsatisfactory for some or all of the following reasons.

(i) There may be as many as thirty parameters which must be calibrated using empirical trial and error or automatic parameter optimization procedures involving multiple simulation runs, with the objective of reproducing as closely as possible a set of historic data. On-line parameter estimation is clearly infeasible in such cases,

and consequently in the long term a poorly calibrated real-time model may result if catchment characteristics are slowly changing with time.

(ii) Cumulative errors may be allowed to build up if the model does not have a correction facility which uses the past forecast errors (obtained from observed streamflow) in updating the forecast.

(iii) The computer storage requirement may necessitate a large computer as conceptual models generally use non-recursive methods requiring the re-processing of historic data at each time step. Furthermore, if a small mini-computer is being used, the program run-time may be a significant portion of the forecast lead-time, resulting effectively in a reduction of the lead-time.

(iv) It may not be possible to give some confidence limits (in a statistical sense) to streamflow forecasts.

The real-time models, hereafter referred to as 'models' to be described overcome some or all of the above problems and can be categorised as follows:

(i) stochastic linear black-box models which use efficient recursive estimation techniques such as the Kalman filter;

(ii) simple conceptual models formulated within a Kalman filter estimation procedure (which takes into account noise due to measurement and/or modelling error) to allow for (a) the incorporation of a forecast error correction facility so that a previous forecast can be updated using the latest information and/or (b) the on-line estimation of parameters. These models may be non-linear.

5.2 REAL-TIME MODELS

Some of the more widely used techniques of linear filtering and parameter estimation have been outlined in sections 3.1 and 3.2. A survey of the literature pertaining to the development of on-line river flow forecasting models using filtering techniques is now presented.

One of the first applications of the state space approach to streamflow forecasting is due to Toyoda, Toriumi and Inone (1969) who investigated the stability and adaptability of a state-space runoff prediction model as applied to the Sagar River in Japan. Hino (1970) used an analytical solution to the Wiener Hopf equation* for linear rainfall runoff processes to obtain optimum unit hydrographs which were then used for prediction purposes. Later Hino (1973) used a Kalman filter (TYPE I, see section 3.1) to estimate recursively ordinates of the unit hydrograph which were then used for prediction purposes. In his work the filter was used for parameter estimation but not for actually deriving the forecast. Clarke (1971) showed that an ARMAX model can be used for flow forecasting and in particular how the variances of the forecasts can be calculated. Kashyap and Rao (1973) considered several heuristically derived recursive prediction algorithms, where the parameters were estimated using least squares. They did not however use rainfall input so only runoff observations were used in deriving the forecasts. Their basic approach is very similar to that of the self-tuning predictor (Wittenmark, 1974) which has been applied to flow forecasting (with rainfall input) by the writer (Ganendra, 1976). Whitehead and Young (1974) have applied the IV-AML method to flow forecasting in conjunction with a stochastic water quality model for part of the Bedford-Ouse River System. The IV-AML method has also been used by Lorent (1976), in which the baseflow and surface runoff components were separated, modelled independently and then combined to produce a forecast. Todini and Bouillot (1975) used a Kalman filter with a TYPE II state-space model formulation in which the model parameters were estimated by the IV-AML method.

** Wiener's (1949) classical formulation of the linear stationary filtering problem later to be solved by Kalman (1960).*

An improvement over this model was possible (Todini, Szöllösi-Nagy and Wood (1976) by introducing a threshold which is used to split a single rainfall input into two or more separate inputs (Todini and Wallis, 1977). The resultant model response is piecewise linear and dependent on antecedent rainfall. Furthermore the state-space model formulation incorporated a rainfall prediction component to enable the model to be used for a larger forecast lead-time. Todini, Szöllösi-Nagy and Wood (1976) have also used a Kalman filter with a TYPE I state-space model formulation, where parameter time-variance is modelled by a random walk model. The maximum likelihood method (Åström and Bohlin, 1966) was used by Katayama (1976) to estimate the parameters (non-recursively) of a TYPE II state-space model where the state was updated using a Kalman filter. Finally a Kalman filter and a TYPE III formulation was used by Moore and Weiss (1976) to estimate the parameters of a simple conceptual model (Mandeville, 1975) and simultaneously provide flow forecasts. Duong, Winn and Johnson (1975) also used a Kalman filter with a TYPE III state-space formulation for estimating the parameters of the Prasad model (1967).

5.3 THE SELF-TUNING PREDICTOR

The model to be described is based on a minimum mean square error predictor which assumes that the rainfall runoff process can be represented by a linear time-invariant stochastic difference equation (see Section 3.1). It will first be shown that the predictor can be obtained if the parameters of the process are known (Åström, 1970). Consider first the nth order stochastic process model,

$$A(z^{-1})y(t) = C(z^{-1})e(t) \tag{5.1}$$

where $y(t)$ is the output and $e(t)$ is a Gaussian white noise sequence distributed as $N(0,\sigma^2)$.

The polynomials $A(z^{-1})$ and $C(z^{-1})$ are given by

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n} \\ C(z^{-1}) &= 1 + c_1 a^{-1} + \ldots + c_n a^{-n} \end{aligned} \tag{5.2}$$

and for convenience are assumed to be of the same order n. 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 k-step ahead prediction error is given by

$$\varepsilon(t+k) = y(t+k) - \hat{y}(t+k|t) \tag{5.3}$$

The minimum mean square error predictor, i.e. the predictor that minimises the loss function

$$V = E\{\varepsilon(t)^2\} \tag{5.4}$$

can be obtained, as for the case of the self-tuning regulator (Section 4.2), by introducing the identity

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-k}G(z^{-1}) \tag{5.5}$$

where

$$F(z^{-1}) = 1 + f_1 z^{-1} + \ldots + f_{k-1} z^{-k+1}$$

$$G(z^{-1}) = g_o + g_1 z^{-1} + \ldots + g_{n-1} z^{-n+1} \qquad (5.6)$$

It is shown in Appendix A that the predictor that minimises V is given by

$$\hat{y}(t+k|t) = \frac{G(z^{-1})}{A(z^{-1})F(z^{-1})}\ \varepsilon(t) \qquad (5.7)$$

The derivation of the above optimal predictor is in fact similar to the derivation of the minimum variance control law derived in Appendix C. The expected k-step ahead prediction error variance is

$$E\{\varepsilon(t)^2\} = \sigma^2(1 + f_1^2 + \ldots f_{k-1}^2) \qquad (5.8)$$

If the parameters of the process were known, the optimal predictor could readily be calculated. For processes with unknown parameters the parameters of the predictor can be directly estimated without reference to the process parameters. The penalty of using the predictor (eqn. 5.7) directly is that it contains k-1 more parameters than the system model it represents. 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, the predictor obtained in the limit will be the minimum mean square error predictor. Wittenmark (1974) has shown that in fact the self-tuning predictor is a special case of the self-tuning regulator (see Section 4.3). This is evident by considering the predictor to be a regulator with $\hat{y}(t+k|t)$ as the control input to the process where the variance of the output $\varepsilon(t)$ is to be minimised. Consequently the results of theoretical and experimental investigations of the self-tuning regulator are applicable to the self-tuning predictor.

Considering the self-tuning predictor as a streamflow prediction model, the system is assumed to be represented by the nth order ARMAX model given by

$$A(a^{-1})q(t) = B(z^{-1})r(t) + C(z^{-1})e(t) \tag{5.9}$$

where $q(t)$ is the streamflow (output) and $r(t)$ the rainfall (input) while $e(t)$, $A(z^{-1})$ and $C(z^{-1})$ are as in equation 5.1. The noise sequence $e(t)$ accounts for unknown disturbances, poor modelling and measurement noise. The polynomial $B(z^{-1})$ is defined as

$$B(z^{-1}) = b_o + b_1{}^{-1} + \ldots + b_n{}^{-n} \tag{5.10}$$

The optimal k-step ahead predictor for the system equation 5.9 is given by

$$\hat{q}(t+k|t) = \frac{B(z^{-1})r(t)}{A(z^{-1})} + \frac{G(z^{-1})}{A(z^{-1})F(z^{-1})}\varepsilon(t) \tag{5.11}$$

which can be rearranged to give

$$\hat{q}(t+k|t) = (1-A(z^{-1})F(z^{-1}))\hat{q}(t+k|t) + B(z^{-1})F(z^{-1})r(t) + G(z^{-1})\varepsilon(t) \tag{5.12}$$

Introducing the vector notation of section 3.2 and defining

$$\begin{aligned}(1-A(z^{-1})F(z^{-1})) = A(z^{-1}) &= \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots + \alpha_n z^{-n} \\ B(z^{-1})F(z^{-1}) = B(z^{-1}) &= \beta_1 + \beta_2 z^{-1} + \ldots + \beta_m z^{-m+1} \\ G(z^{-1}) = C(z^{-1}) &= \gamma_1 + \gamma_2 z^{-1} + \ldots + \gamma_\ell z^{-\ell+1}\end{aligned} \tag{5.13}$$

where n, m and ℓ are simply the number of terms in the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ respectively, and the data and parameter vectors are given by

$$\begin{aligned}\phi(t) &= [\hat{q}(t+k-1|t-1), \ldots \hat{q}(t+k-n|t-n); r(t), \ldots r(t-m+1); \\ &\qquad \varepsilon(t), \ldots \varepsilon(t-\ell+1)] \\ \theta^T &= [a_1, \ldots a_n; b_1, \ldots b_m; c_1, \ldots c_\ell]\end{aligned} \tag{5.14}$$

so that the predictor equation 5.12 can be written as

$$\hat{q}(t+k|t) = \phi(t)\theta \tag{5.15}$$

The problem now is to estimate the parameter vector θ at each sampling point using the latest data acquired. For the case where $C(z^{-1}) = 1$, the recursive least-squares parameter estimator (Appendix D) will give consistent parameter estimates and result in self-tuning. The estimator is thus given by the following equations which are evaluated at each time step:

$$\varepsilon(t) = q(t) - \hat{q}(t|t-k) \tag{5.16}$$

$$P(t) = P(t-1) - \frac{P(t-1)\phi^T(t-1)\phi(t-1)P(t-1)}{1 + \phi(t-1)P(t-1)\phi^T(t-1)} \tag{5.17}$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\hat{\theta}^T(t-1)\varepsilon(t) \tag{5.18}$$

The P-matrix is the error covariance matrix of the parameter estimates, standardized with respect to the noise variance, i.e.

$$\sigma^2 P(t) = E\{(\theta-\hat{\theta}(t))(\theta-\hat{\theta}(t))^T\} \tag{5.19}$$

The parameters θ are assumed time invariant and the parameter estimates $\hat{\theta}(t)$ above are shown as a function of time merely to indicate the parameter estimates at time t, t-1, ... etc. This algorithm avoids direct matrix inversion (the inverse term in equation 5.17 is a scalar) and as with other recursive methods the past data do not have to be stored for subsequent calculations. The estimated optimal (minimum mean square error) predictions, by the certainty-equivalent principle (see Section 4.2), are then computed as if the values of the estimated parametres were the true ones. Hence the predictor for q(t+k) is given by

$$\hat{q}(t+k|t) = \phi(t)\hat{\theta}(t) \tag{5.20}$$

The above estimator and predictor when combined are known as the

self-tuning predictor (Wittenmark, 1974). Analagous to the self-tuning regulator, the self-tuning predictor which has sufficient parameters converges to the optimal predictor when the covariances

$$\gamma_{\varepsilon\varepsilon}(\tau) = E\{\varepsilon(t)\varepsilon(t+\tau)\} = 0 \quad \tau = k, \ldots k+\ell-1$$

$$\gamma_{\varepsilon\hat{q}}(\tau) = E\{\varepsilon(t+\tau)\hat{q}(t+k|t)\} = 0 \quad \tau = k, \ldots k+n \qquad (5.21)$$

Following the innovations property (Mehra, 1969) the residuals $\varepsilon(t)$ should constitute a sequence of Gaussian white noise (i.e. $\varepsilon(t)$ should also be uncorrelated for all τ, since the model has then abstracted all the information in the data and knowledge of one residual gives no information about the next.

If the process being modelled is slowly time varying, the least-squares estimator can be modified to include exponential weighting of the past data (Söderström, Ljung and Gustavsson, 1974). In this case the loss function at time t is taken as

$$V(t) = \sum_{i=0}^{t} \alpha^{t-i} \varepsilon(i)^2 \qquad (5.22)$$

giving more weight to recent data and less to past. The parameters in this case will be estimated mainly on the weighting of the most recent data. The factor α, $0 << \alpha < 1$ is known as the forgetting factor. Exponential weighting may be interpreted as a constraint which prevents the elements of the P-matrix from becoming too small. The parameter estimates will then be assumed to have not yet converged so that the estimator will be sensitive to changes in the system dynamics. As a rough guide, about 200 past data values are remembered for $\alpha = 0.99$ while only 40 are remembered for $\alpha = 0.95$.

To implement the self-tuning predictor, initial parameter values must first be assumed. It is possible to accelerate the convergence rate of the parameters (if the initial parameter values are poor) by applying a time dependent forgetting factor,

$$\alpha(t) = \alpha(t-1)\alpha_o + (1-\alpha_o) \qquad (5.23)$$

where α_o is a constant, $0 << \alpha_o < 1$. Equation 5.23 allows $\alpha(t)$ to tend to one at a rate dependent on α_o. This forgetting profile ensures that the parameters will initially (when $\alpha(t)$ is small) converge rapidly to their proper values, while as t becomes larger the forgetting factor $\alpha(t)$ will tend to one by which time presumably the covariances $\gamma_{\varepsilon\varepsilon}$ and $\gamma_{\varepsilon\hat{q}}$ will satisfy equations 5.21.

The algorithm

The self-tuning predictor with exponential weighting of the past data is given by the following recursive equations which are evaluated at each time-step:

$$\text{(STEP 1)} \quad \varepsilon(t) = q(t) - \hat{q}(t|t-k) \tag{5.24}$$

$$\text{(STEP 2)} \quad \alpha(t) = \alpha(t-1)\alpha_o + (1-\alpha_o) \tag{5.25}$$

$$\text{(STEP 3)} \quad P(t) = \frac{P(t-1)}{\alpha(t)} - \frac{P(t-1)\phi^T(t-1)\phi(t-1)P(t-1)}{\alpha(t)^2+\alpha(t)\phi(t-1)P(t-1)\phi^T(t-1)} \tag{5.26}$$

$$\text{(STEP 4)} \quad \hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\phi^T(t-1)\varepsilon(t) \tag{5.27}$$

$$\text{(STEP 5)} \quad \hat{q}(t+k|t) = \phi(t)\hat{\theta}(t) \tag{5.28}$$

The self-tuning predictor algorithm remains the same if more than one input is included in the model. The additional variables and corresponding parameters can be included in the respective vectors $\phi(t)$ and θ. Also, if a delay b is applied to an input, the variable in question is delayed by b in $\phi(t)$. As an example, if there are two rainfall inputs $r_1(t)$ and $r_2(t)$ (which may be from two raingauges located at different parts of the catchment) and $r_2(t)$ is delayed by b timesteps (which may be due to its more remote location than $r_1(t)$, in relation to the catchment outfall), the data vector can be given by

$$\phi(t) = [\hat{q}(t+k-1|t-1), \ldots ; r_1(t), \ldots ; r_2(t-b), \ldots ; \varepsilon(t), \ldots]. \tag{5.29}$$

The procedure for selecting the initial conditions for this algorithm are outlined below.

Initial Conditions

To initiate the above algorithm it is necessary to have some starting values of the parameters $\hat{\theta}(0)$, the matrix $P(0)$ and the forgetting factor $\alpha(0)$ (and α_o). If these parameters are completely unknown, the algorithm can be initiated with $\theta(0) = 0$ and $P(0) = \mu \times I$, where μ is a large number. Typically μ is equal to between 10 and 10^4 and as a rough guide, 10 times the variance of the output $q(t)$ is sufficient. If the parameters of the system model (eqn. 5.9) are approximately known, they can be used to compute $\hat{\theta}(0)$ from equation 5.13. Subsequently μ can be chosen to be between 10^{-4} and 10^{-1}, reflecting more confidence in the parameter estimates. In general, the choice of initial values, $\hat{\theta}$ and $P(0)$, is not critical. If the estimates converge properly, the initial values will have no influence on the parameters estimation asymptotically (Söderström, Ljung and Gustavsson, 1974). A too small value of μ can however give rise to a slow rate of convergence, though this can be eliminated by exponential forgetting. To ensure rapid convergence $\alpha(0)$ should be set to between 0.95 to 0.99, with the constant α_o set at between 0.98 to 0.99 (the larger α_o is, the faster $\alpha(t)$ will converge to 1.0). If the system is slowly time varying $\alpha(0)$ can be set to between 0.98 and 0.999 while α_o is set to 1.0. The selection of $\alpha(0)$ and α_o must be made to ensure both rapid parameter convergence and allow for time variance of the parameters. The sample auto and cross-correlation functions $r_{\varepsilon\varepsilon}(\tau)$ and $r_{\varepsilon\hat{q}}(\tau)$ (which should satisfy eqns. 5.21) can be used as a guide in the selection procedure.

It was found that, despite the poor initial parameter estimate used (i.e. $\theta(0) = 0$), the parameters of the 1-step ahead predictor ($k = 1$) always converged to give reasonable predictions. For $k > 1$ an instability condition may occur before reasonable estimation of the parameters are obtained. Failure of the computational procedure at this stage (with the numerical value of a variable exceeding the machine capability) can be averted by applying a maximum and minimum constraint on $\hat{q}(t+k|t)$, i.e.

If $\hat{q}(t+k|t) > q_{max}$; then $\hat{q}(t+k|t) = q_{max}$

and

If $\hat{q}(t+k|t) < q_{min}$; then $\hat{q}(t+k|t) = q_{min}$

The values of q_{max} and q_{min} can be taken as twice the value of the largest observed flow and zero respectively.

Selection of polynomial orders

The choice of the number of terms in the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ of the self-tuning predictor is critical. These polynomials are not the same as the polynomials for the system model (i.e. $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$), therefore the Box and Jenkins method for model structure estimation outlined in Section 3.3 does not give the structure of the predictor directly. However the orders of the polynomials $A(a^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ which may be known a priori or obtained by the Box and Jenkins method can be used to obtain the orders of the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$. Noting that the polynomial $F(z^{-1})$ has k terms and that $G(z^{-1})$ is of order n-1 if $A(a^{-1})$ is of order n, the respective number of terms in the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ can be computed from equations 5.13. Alternatively, competing predictors (with different structures) may be applied to test data and the best model selected by analysis of the residuals as indicated in section 3.3 (or by the requirement that the covariances $\gamma_{\varepsilon\varepsilon}$ and $\gamma_{\varepsilon\hat{q}}$ for the optimal predictor satisfy equations 5.21).

5.4 FLOW FORECASTING FOR THE BROSNA CATCHMENT

The Brosna Catchment covers some 1180 sq km of mostly flat grazing land in the centre of Ireland. The River Brosna, a tributary of the Shannon, is perennial with maximum flows of about 100 cumecs occurring during February. Glacial deposits do provide some relief, and there are some peatbogs and woodland areas. The geology consists of carboniferous limestone covered by boulder clay. Rainfall is uniformly distributed throughout the year and of relatively low average intensity.

The data have been compiled from 5 years of discharge measurements (14 September 1960 to 26 September 1965) at Ferbane and a centrally sited continuous recording raingauge. Areal mean rainfall was computed using 18 daily raingauges distributed throughout the catchment, 3 hourly data being evaluated with reference to the central recording gauge.

Model structure estimation

An approximation to the impulse response for the rainfall-runoff relationship was first obtained using the Box and Jenkins method discussed in section 3.3. The sample autocorrelation functions of the rainfall (r(t)) and runoff (q(t)) series and the sample partial autocorrelation function of r(t) for a sample size N = 1400 (3-hourly data starting 9 am, 31 August 1960) are shown in fig. 5.1. The dashed lines indicate the 95 per cent confidence intervals within which the sample auto and cross-correlation functions can be assumed to be zero (Appendix E). The sample partial autocorrelation function $\alpha_{\tau\tau}$ of r(t) cuts off after lag one while the sample autocorrelation function $r_{rr}(\tau)$ of r(t) tails off after lag one, suggesting that the rainfall series can be represented by a first order AR model. The solution to the corresponding Yule-Walker equation (eqn. 3.58) gave the model

$$r(t) = 0.3r(t-1) + e_r(t) \qquad (5.30)$$

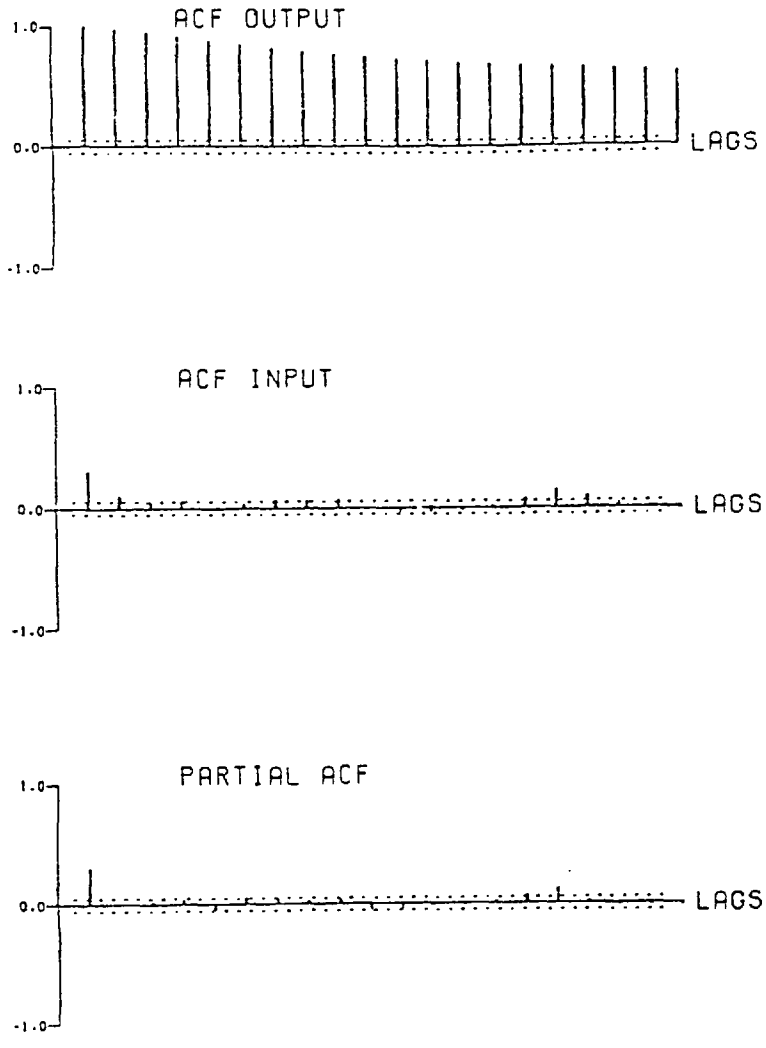


Fig. 5.1 *BROSNA DATA: Sample autocorrelation function of output (flow) and input (rainfall), and partial autocorrelation function for input. Sample size N = 1400 (95% confidence limits ±1.96/√N are shown.*

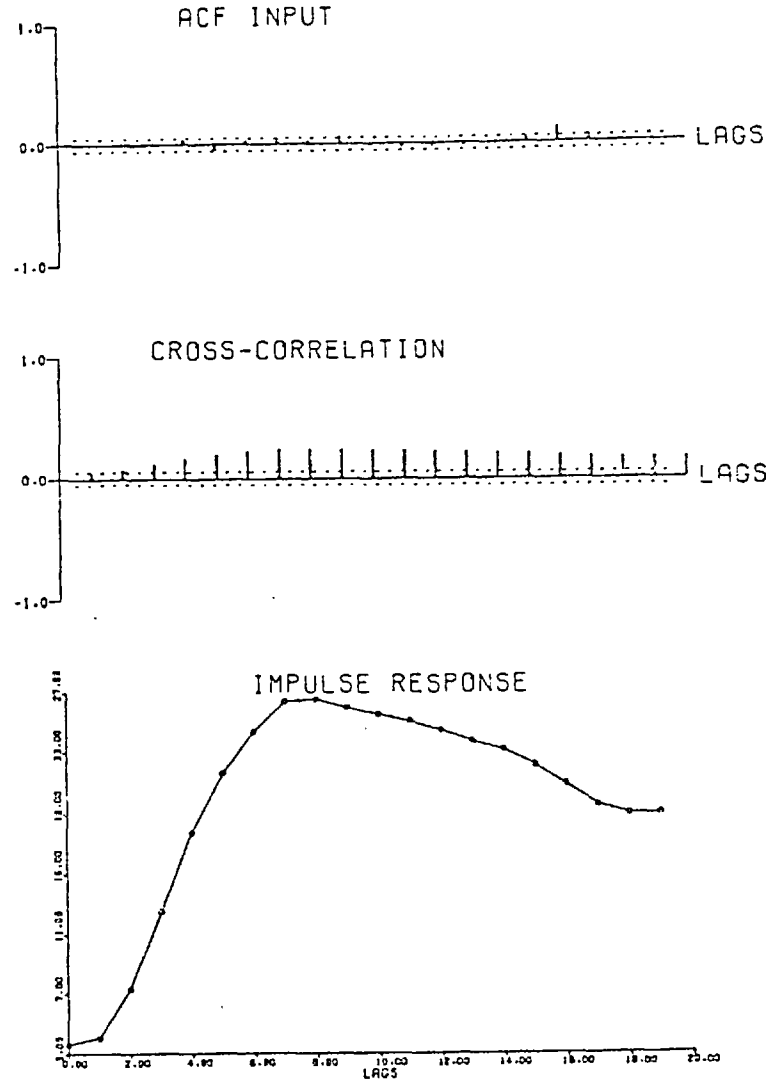


Fig. 5.2 *BROSNA DATA: Sample autocorrelation of prewhitened rainfall input, sample cross-correlation function for prewhitened input and runoff output, and approximate impulse response. Sample size N = 1400.*

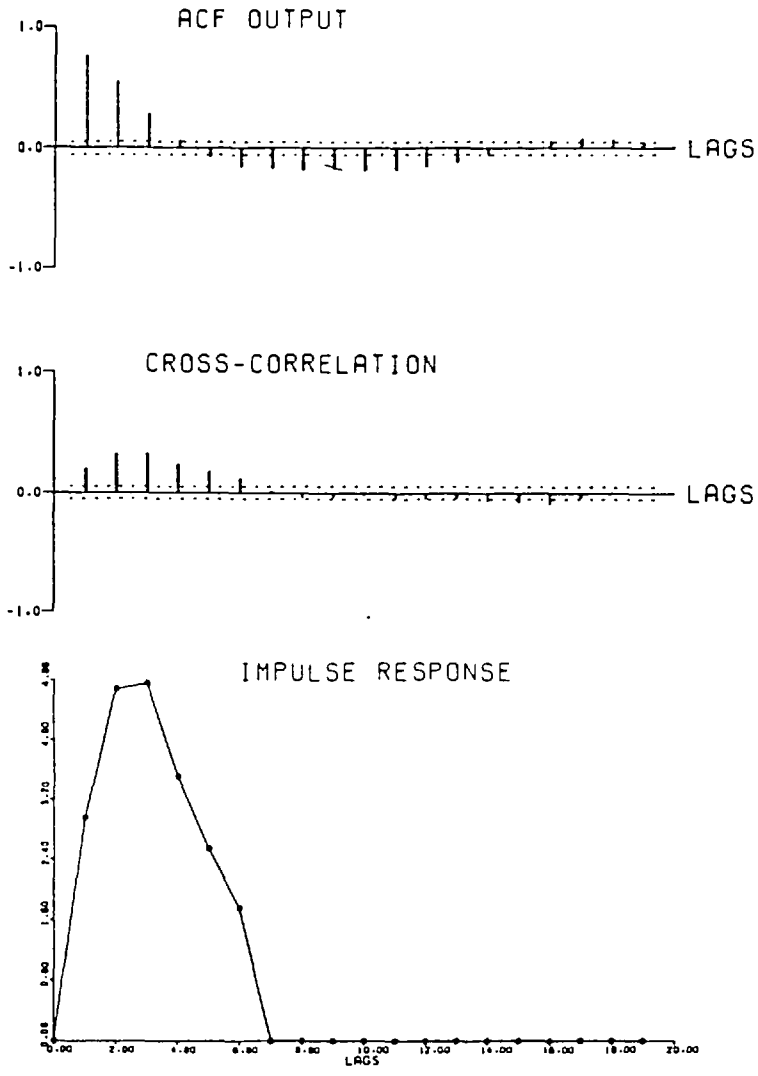


Fig. 5.3 BROSNA DATA: Sample autocorrelation function of differenced output (flow), sample cross-correlation function for pre-whitened input and differenced output (flow), and approximate impulse response. Sample size N = 1400.

where $e_r(t)$ is the white noise sequence constituting the prewhitened rainfall series (input) when the following linear transformation is made.

$$e_r(t) = r(t) - 0.3r(t-1) \tag{5.31}$$

The sample autocorrelation function of the prewhitened input is shown in fig. 5.2. The output (runoff) series was then similarly transformed (replacing the input by the output in equation 5.31) and the cross-correlation between this series and the prewhitened input series estimated. Dividing these cross-correlation estimates by the estimated variance of the prewhitened input gives the approximate impulse response function (see eqn. 3.66) as shown in fig. 5.2.

It is clearly evident that the sample cross-correlation function shown in fig. 5.2 has been dominated by the non-stationarity of the output $q(t)$. This non-stationarity is revealed by the slow decay rate of $r_{qq}(\tau)$. Figure 5.3 shows the sample autocorrelation function of the differenced runoff series $r_{\nabla q \nabla q}(\tau)$, the sample cross-correlation function between $e_r(t)$ and the correspondingly transformed $\nabla q(t)$ series, and the resulting estimated impulse response.

Thus while a well defined impulse response can be obtained if the output is differenced, a clear impulse response function was difficult to determine with the non-stationary series. The problem with forecasting a differenced runoff series $\nabla q(t)$ is that the forecasted variable has to be transformed back to the forecast runoff $\hat{q}(t|t-k)$. This can lead to amplification of the prediction error (Young, Shellswell and Neethling, 1971). The sample autocorrelation function is dependent on the data sample size and sampling interval. If the sample autocorrelation function is used to determine whether the time-series in question is stationary or not, then any stationarity (or non-stationarity) assumptions will also be based on the

sample size and sampling interval. The data sample size used in estimating the parameters of the predictor is dependent on the memory (or forgetting factor $\alpha(t)$) of the estimator. For the Brosna runoff data (3-hourly), the apparent non-stationarity indicated by the sample autocorrelations could be removed by reducing the sampling size. Thus, if $\alpha(t)$ is correctly chosen, it can be assumed that the data corresponding to the memory of the estimator are stationary. It was also found that any attempt to reduce the sample size used in estimating the autocorrelations (from N = 1400) resulted in a sample autocorrelation function for the rainfall series that was difficult or impossible to interpret (the lag one autocorrelations were small while spurious autocorrelations occurred at larger lags).

As a result of the above analysis it was concluded that the Box and Jenkins method for model structure identification was unreliable when applied to the Brosna data. Alternatively, analysis of the prediction error series (as outlined in Section 3.3) was found to be both a simple and reliable method for model structure identification. The following procedure was adopted for selecting the orders of the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ of the self-tuning predictor:

(i) a number of competing predictors were selected (with different n, m and ℓ values);

(ii) using the selected predictors, sequences of predictions $\hat{q}(t+k|t)$, and prediction errors $\varepsilon(t+k)$ were generated;

(iii) from these, the predictors which best satisfied equations 5.21 were selected;

(iv) finally the best predictor was selected on a minimum mean square error requirement (and hence giving n, m and ℓ).

While the difficulty in estimating the impulse response precluded its use in estimating n, m and ℓ, the first few ordinates of the

impulse response were good indicators of the input delay parameter b. The impulse response shown in fig. 5.2 indicates that b = 1 for a 1-step ahead predictor (k = 1) if the forecast lead time Δt is 3 hours, and the cross-correlation coefficients (fig. 5.2) are assumed to be zero for lags 0 and 1. In other words there is no relationship between q(t) and r(t) or q(t) and r(t-1).

Parameter convergence

The rapid convergence of the parameters of the self-tuning predictor (to those of the optimal predictor) is apparent in fig. 5.4, which shows the observed and 1-step ahead predicted runoff, the observed rainfall and the prediction error. The data are sampled starting at t = 0. The initial conditions $\phi(0) = 1 \times 10^4$ and $\theta(0) = 0$ were used with the result that the 1-step ahead prediction error rapidly decreases to a reasonable value after four time steps. Values for n, m, ℓ, α_o, $\alpha(0)$, k and the sampling interval are given in fig. 5.4. An input delay of 0 was used. The parameter p has not yet been defined and is not used when p = 0.

The P-matrix is proportional to the error covariance matrix of the parameter estimates, and is positive definite and symmetric. An indication of the magnitude of the matrix elements is thus given by the trace of the matrix, tr|P| (the sum of the modulus of the diagonal elements). The behaviour of the P-matrix as characterised by tr|P| is shown in fig. 5.5 for the case where $\alpha_o = 0.99$ and $\alpha(0) = 0.95$ and for $\alpha_o = \alpha(0) = 1.0$ which corresponds to no exponential weighting of data. Two observations can be made:

(i) exponential weighting prevents tr|P| from decreasing too rapidly so that a larger sample size is processed before good parameter estimates are assumed;

(ii) large changes in the parameters can be recognized by a large decrease in tr|P|, which occurs before the 10th time step and just after the 50th time step. The changes correspond to the storm hydrographs shown in fig. 5.4.

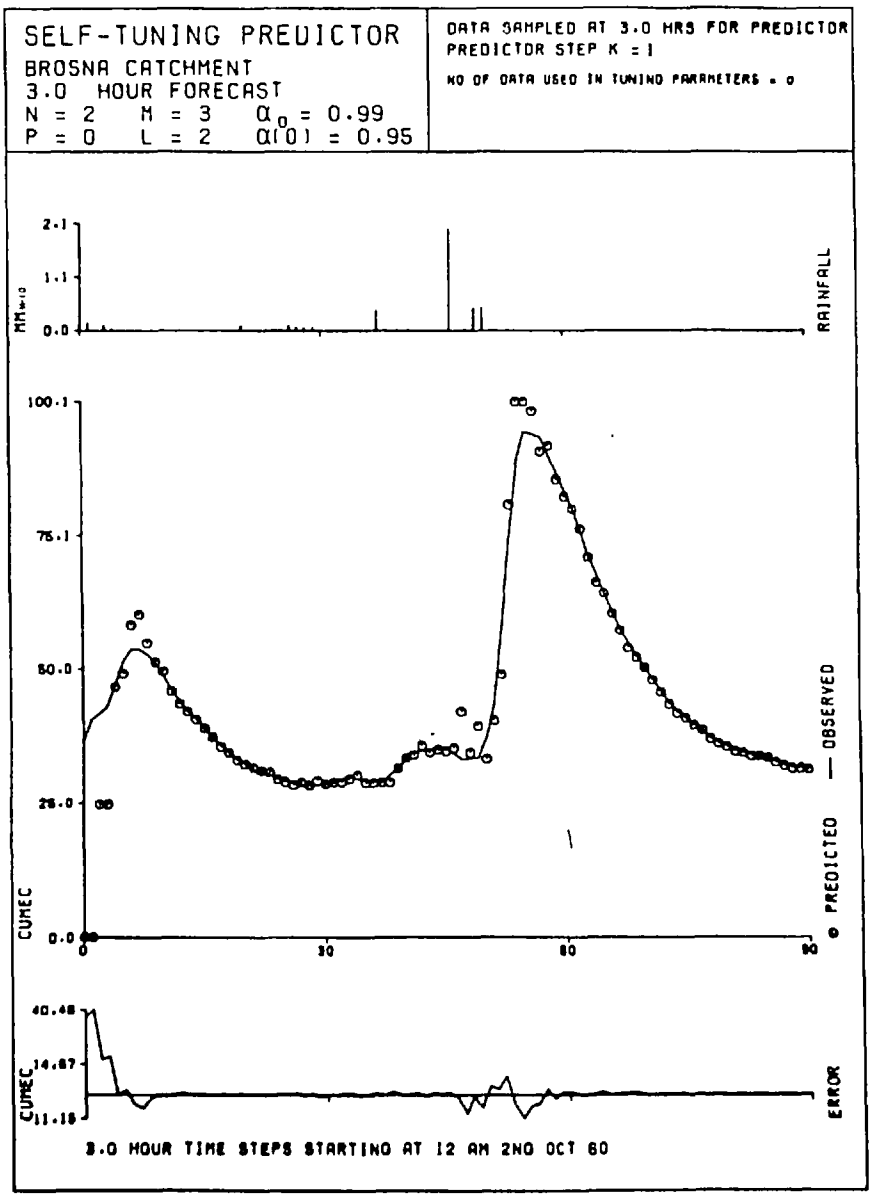


Fig. 5.4 Observed and 1-step ahead hydrograph demonstrating predictor performance and rapid convergence of parameter estimates. Starting conditions at $t = 0$ (corresponding to the first plotted point) where $\phi^T(0) = \hat{\theta}(0) = 0$ and $P(0) = I \times 10^4$. The 1-step ahead prediction error is also shown.

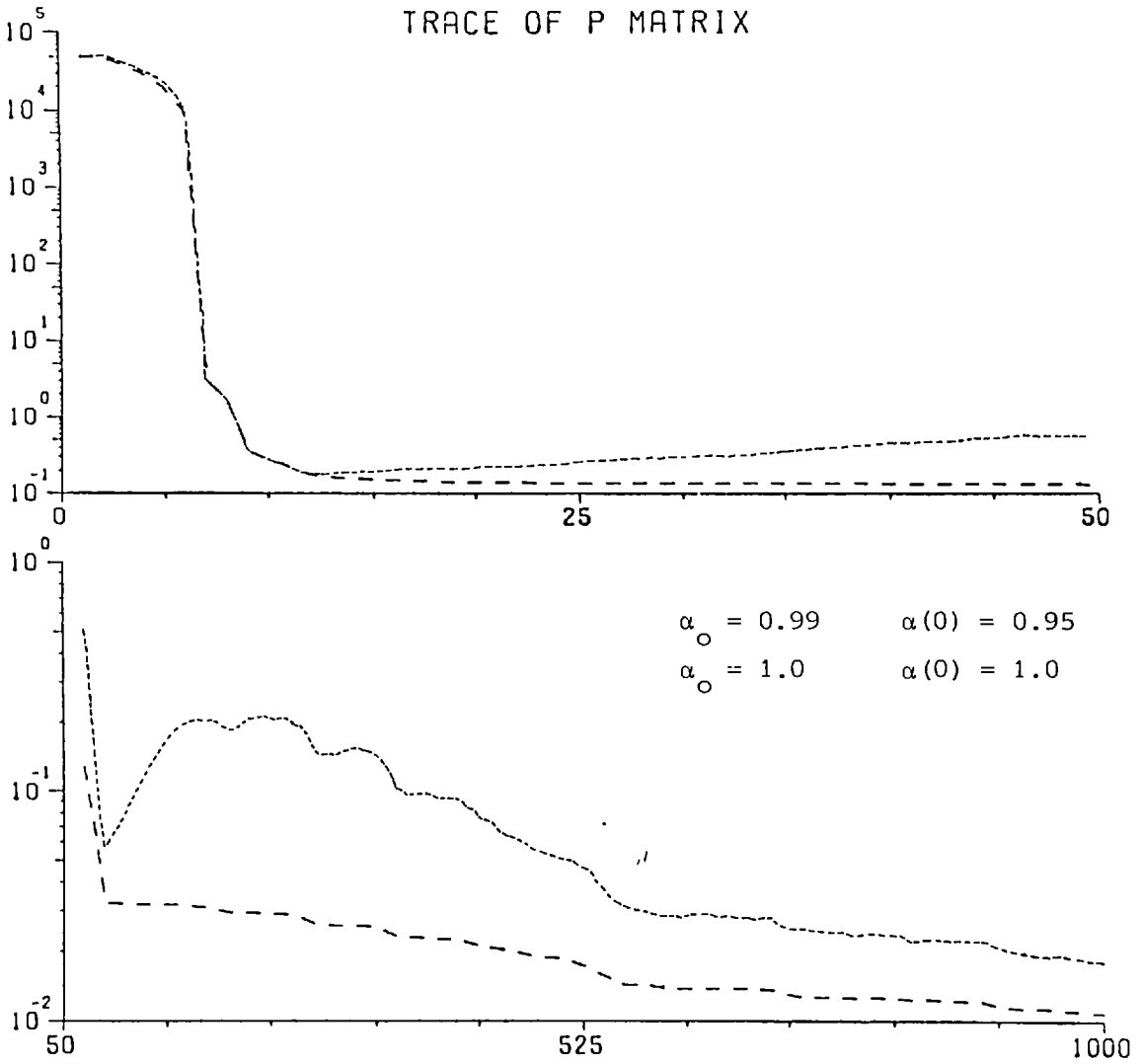


Fig. 5.5 *Parameter convergence shown by $tr|P|$ assuming $P(0) = I \times 10^4$ and $\phi^T(0) = \hat{\theta}(0) = 0$. The first 90 data points correspond to fig. 5.4. The effect of exponential weighing ($\alpha(0) = 0.95$ above) is to constrain the elements of the P-matrix from becoming too small.*

The variations of the parameters $\hat{\theta}$ for the above example are shown in fig. 5.6. Note that $\hat{\theta}(1)$ corresponds to a_1 and $\hat{\theta}(2)$ to a_2 etc. It is characteristic of self-tuning algorithms that the parameters tend to change in unison as shown. The changes occur during large perturbations of the system (storm events) and especially during the initial tuning of the parameters. This does not mean that the parameter estimates are poor, but that there exists a finite parameter space for which the minimum variance requirements are satisfied. The change in parameters corresponding to the two storm events shown in fig. 5.4 (and also shown in fig. 5.5) is particularly obvious in a_1, a_2, c_1 and c_2 (shown as $\hat{\theta}(1)$, $\hat{\theta}(2)$, $\hat{\theta}(6)$ and $\hat{\theta}(7)$). It is also evident that b_1 is a redundant parameter as it tends to 'wander' very close to zero. This indicates that there is a delay of one time step (b=1) between the rainfall and runoff, as suggested by the approximate impulse response shown in fig. 5.2.

The effect of exponential weighting on parameter estimates is shown in fig. 5.7. The model n = 2, m = 2, ℓ = 1, b = 1 was used with 1.0, while for α_o = 1.0 and $\alpha(0)$ = 0.98, the exponential weighting is constant. The ability of the parameters to adapt rapdily to system changes is demonstrated in fig. 5.7(b). The danger of using a large forgetting factor is that while the parameters may change in accordance with the system dynamics, they may oscillate around the true parameter values without actually converging asymptotically. Excessive parameter variation is reflected in a high prediction error variance. In general, the best compromise was found with values of α_o = 0.99 and $\alpha(0)$ = 0.95, thus assuring rapid parameter convergence without excessive parameter fluctuations persisting.

If the mean streamflow

$$\bar{q} = \frac{1}{N} \sum_{t=1}^{N} q(t) \qquad (5.32)$$

is large in relation to the fluctuations in the streamflow, it may

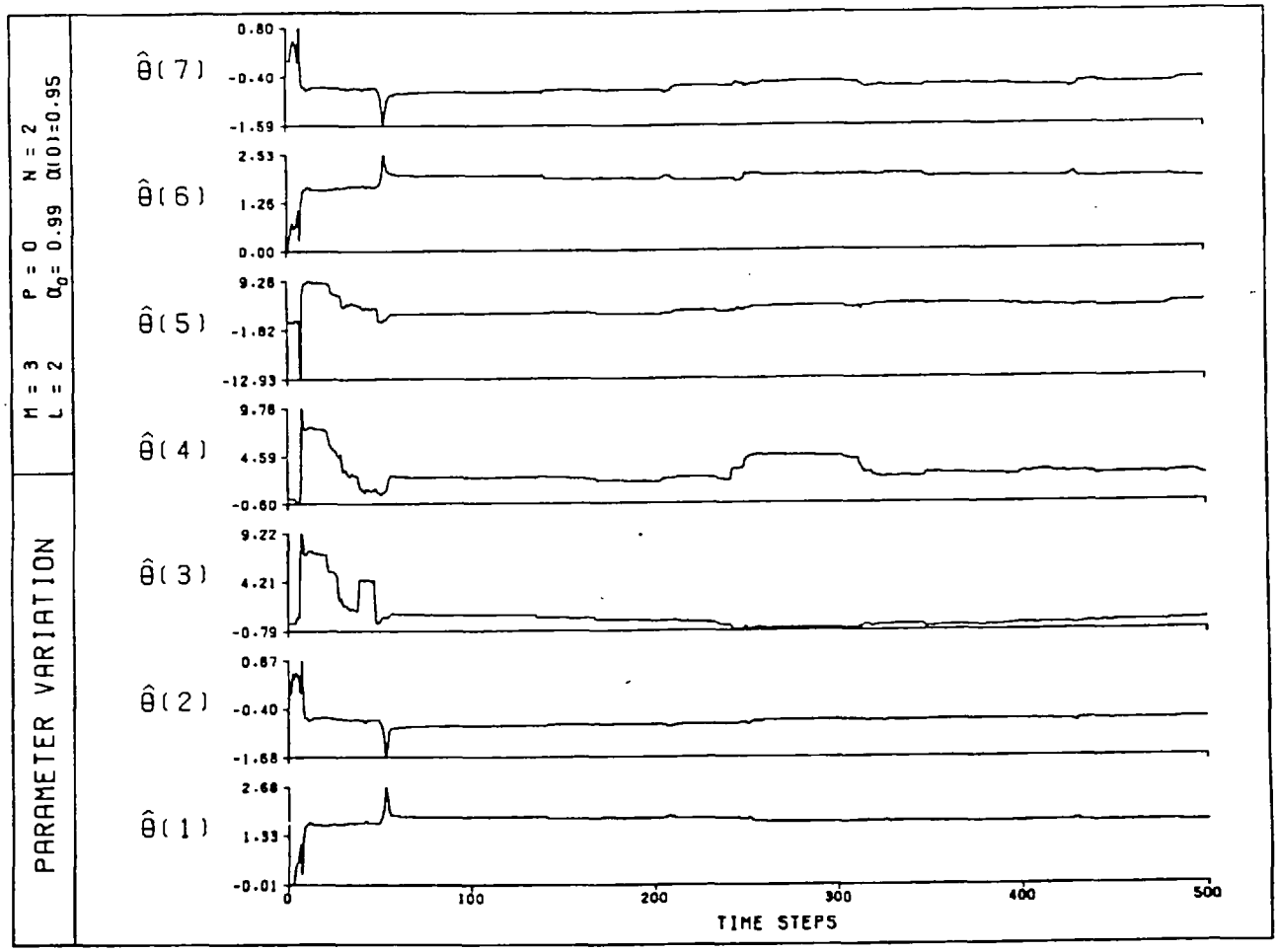
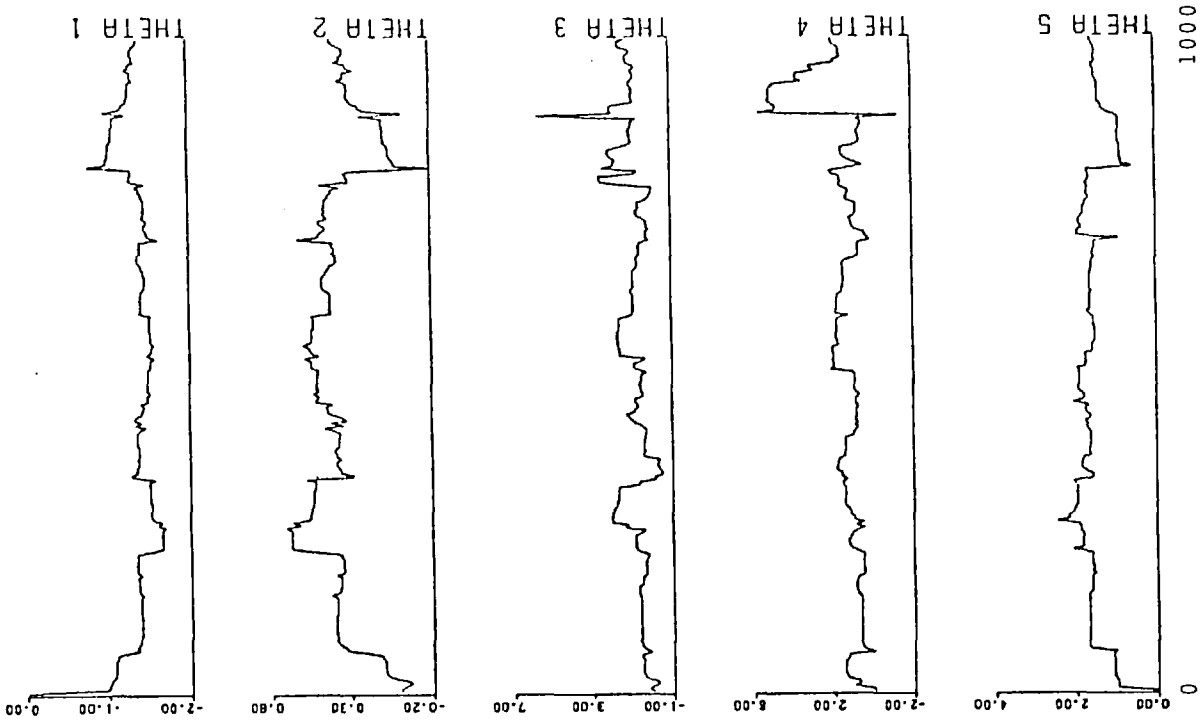


Fig. 5.6 Recursive parameter estimates of the self-tuning predictor. The first 90 data points correspond to fig. 5.4. $\hat{\theta}(1)$ *to* $\hat{\theta}(7)$ *as shown correspond to* a_1, a_2, a_3, b_1, b_2, c_1 *and* c_2 .



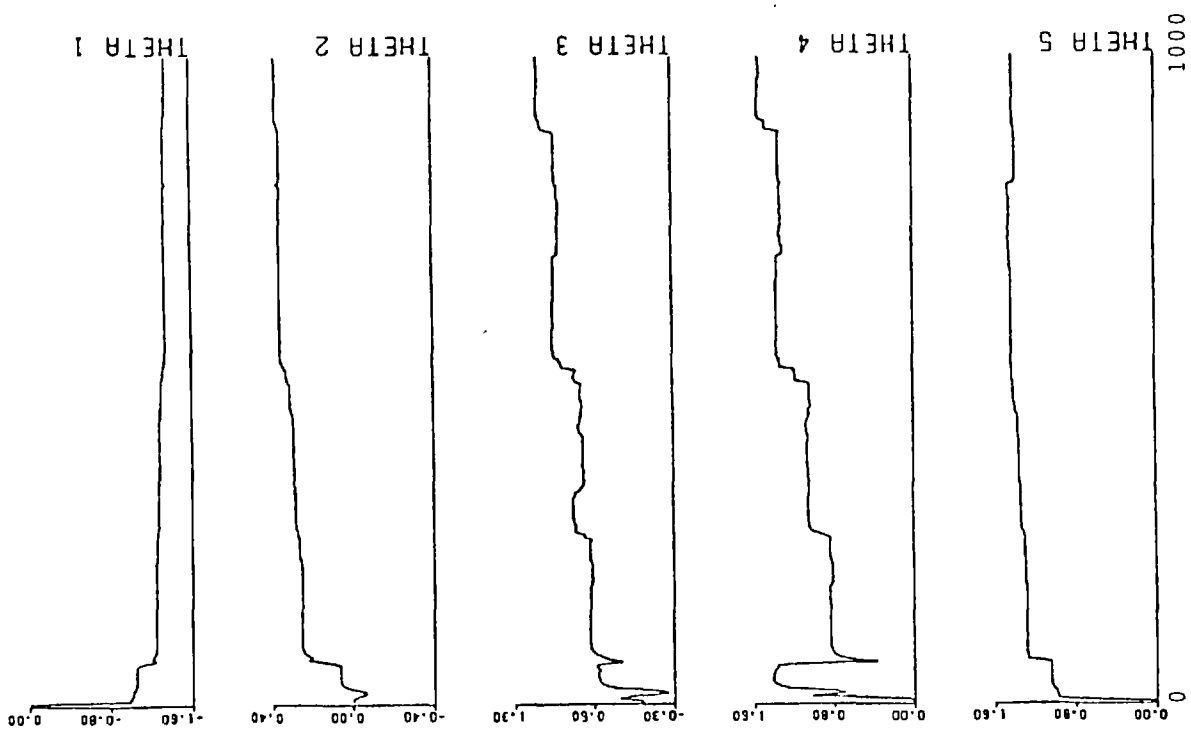


Fig. 5.7 *Recursive parameter estimates showing the effect of exponential weighting. THETA 1, THETA 2 etc. correspond to* $\hat{\theta}(1)$ $\hat{\theta}(2)$ *in fig. 5.6. The predictor* $n = 2$, $m = 2$, $\ell = 1$ *and* $b = 1$ *was used with the same sequence of data as used in fig. 5.5.*

mask the information content of the data from the least squares estimator such that parameter convergence does not occur. This was found to be the case with the Brosna discharge data which has a high baseflow component. To demonstrate this a comparison of $r_{\varepsilon\hat{q}}(\tau)$ and $r_{\varepsilon\varepsilon}(\tau)$ was made using (a) the original discharge data and shown in fig. 5.8 and (b) the same discharge data with the mean level $\bar{q}$ removed and shown in fig. 5.9. When $\bar{q}$ is removed from the output q(t), the predicted variable is the deviation of the output from the mean. The requirement for self-tuning is satisfied when the auto and cross-covariance coefficients as defined by equations 5.21 are zero. The sample size used to compute the sample auto and cross-correlations was N = 1000. The parameters of the predictor were initially calibrated (tuned) with the preceding 1000 data. For this sample size the 95% confidence interval is ±0.06, within which $r_{\varepsilon\varepsilon}(\tau)$ and $r_{\varepsilon\hat{q}}(\tau)$ can be assumed to be zero. Figure 5.8 shows that $r_{\varepsilon\hat{q}}(\tau)$ is highly correlated for small lags, while $r_{\varepsilon\varepsilon}(\tau)$ is non zero for $\tau = 3$ and 4. On the other hand fig. 5.9 shows that by removing the mean level in the output $r_{\varepsilon\hat{q}}(\tau)$ can be assumed to be uncorrelated (the oscillations are the result of correlation between the sample autocorrelation coefficients of the output) while $r_{\varepsilon\varepsilon}(\tau)$ is non zero only for $\tau = 4$. Note however that $r_{\varepsilon\varepsilon}(\tau)$ is zero in both cases for $\tau = 1$ and 2 as required by equations 5.21 ($\ell = 2$ and $k = 1$ for this example).

Comparison of results

A comparative study of 3, 6 and 12 hour forecasts was carried out using the self-tuning predictor. The comparison which was carried out using a six month period of 3-hourly data commencing 9 a.m. 1 October 1960, is divided into two parts:

(i) a 100 day sample was used for obtaining statistical tests of the residuals;

(ii) a 25 day period was used for model structure estimation and visual inspection of observed and forecast hydrographs.

The 25 day period (starting at 9 p.m. 14 January 1961) as shown in

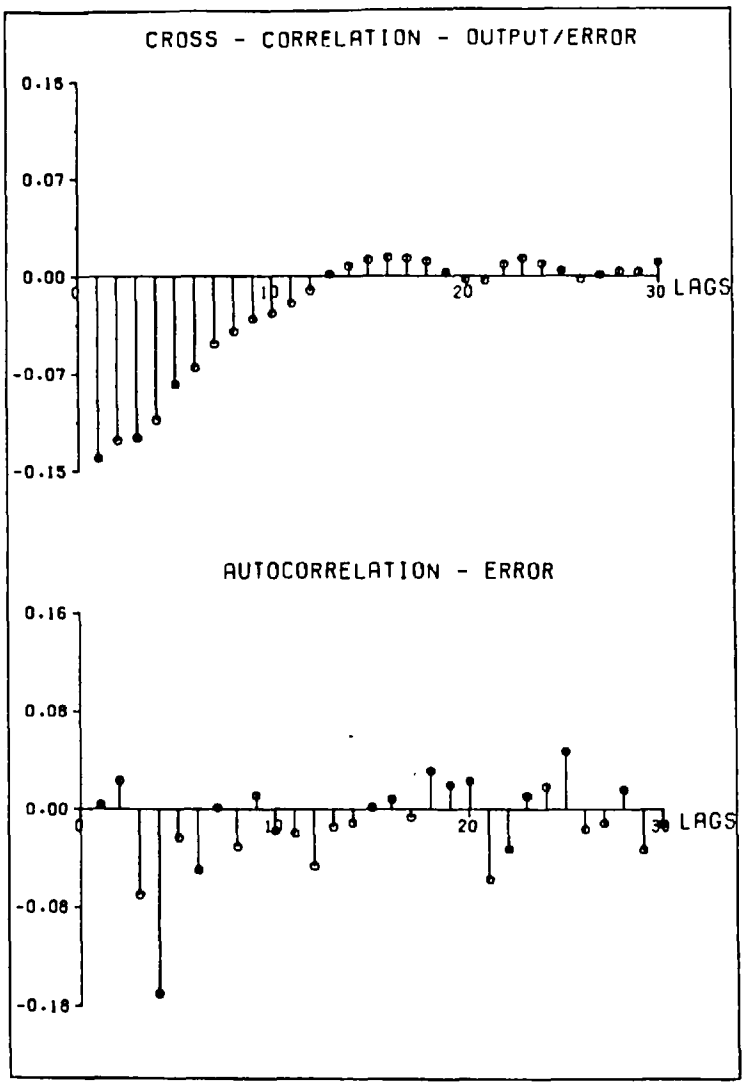


Fig. 5.8 *Sample autocorrelation function $r_{\varepsilon\varepsilon}(\tau)$ and sample cross-correlation function $r_{\varepsilon\hat{q}}(\tau)$ for the Brosna data when the mean level was not removed from the runoff data. The 95% confidence limits are ±0.06.*

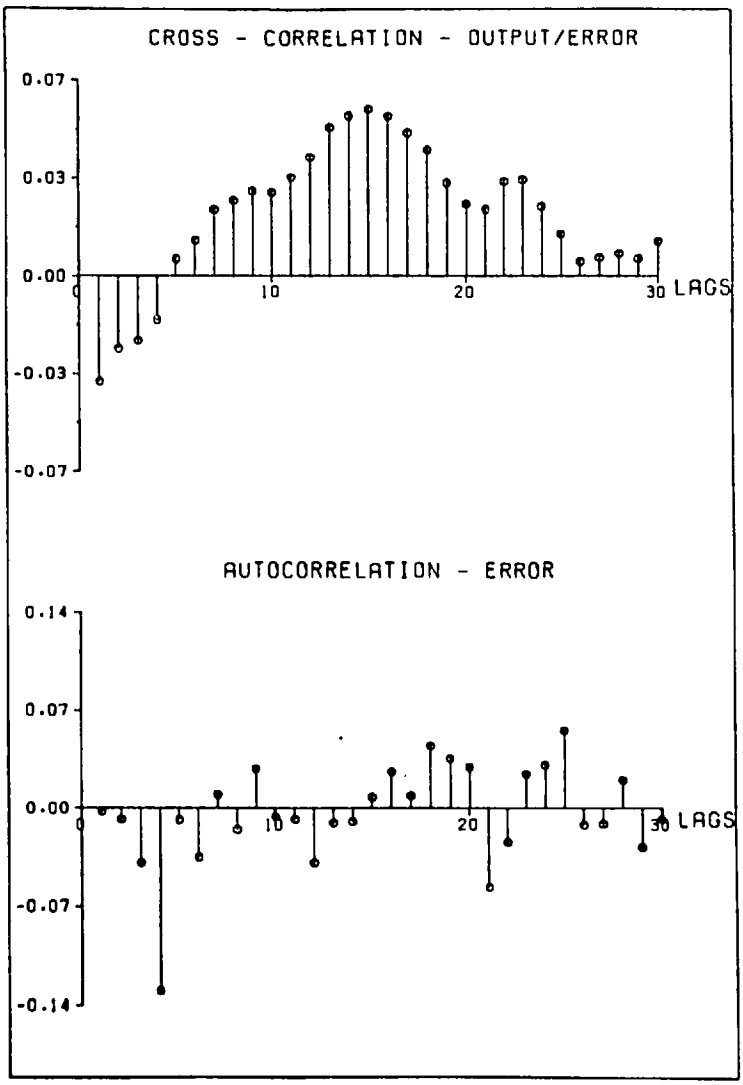


Fig. 5.9 *Sample autocorrelation function $r_{\varepsilon\varepsilon}(\tau)$ and sample cross-correlation function $r_{\varepsilon\hat{q}}(\tau)$ for the Brosna data when the mean level was removed from the runoff data. The 95% confidence limits are ±0.06.*

fig. 5.10 contains a double peaked flood hydrograph which is preceded by a relatively long recession. Time from recession to peak of the first hydrograph rising limb was 27 hours, during which time the flow increased from 29 cumec to 92 cumec.

For each required forecast lead-time the best model structure was selected (the polynomial orders n, m and ℓ and the delay b were estimated) with the criterion of a minimum mean square prediction error over the 25 day period. For forecast lead times of 6 and 12 hours, the sampling interval Δt and the predictor step k were also selected concurrently with the estimation of n, m, ℓ and b (the data were available in multiples of 3 hourly time increments). It was found that best predictors for 3 and 6 hour lead times were obtained with k = 1. The appropriate model was not so apparent with a 12 hour lead time so results for both k = 1 and k = 2 are presented.

The results obtained for the 25 day period of data are shown in figs. 5.10, 5.11, 5.12 and 5.13 (the results were obtained using predictors A B C and D respectively). The starting conditions used for the recursive algorithm were $\alpha_o = 0.99$, $\alpha(0) = 0.95$, $\phi^T(0) = \theta(0) = 0$. The mean level was also removed from the streamflow data. It was assumed that the calibration period available for initially tuning the parameters was from 9 a.m. 1 October 1960 to 9 p.m. 14 January 1961. The number of data actually used for initial tuning in each case is shown in the respective figures. Note that the data are plotted at Δt increments corresponding to the Δt used for the self-tuning predictor, thus in fig. 5.10 the data are plotted at 3-hourly increments and in fig. 5.12 the data are plotted at 12-hourly increments.

Table 5.1 shows a summary of the models selected and statistical analysis of their prediction errors for a 100 day period of data. The data plotted in figs. 5.10, 5.11, 5.12 and 5.13 constitute a subset from this 100 day period. The number of data used for initial model calibration was 1290, 645 and 322 for a Δt of 3, 6

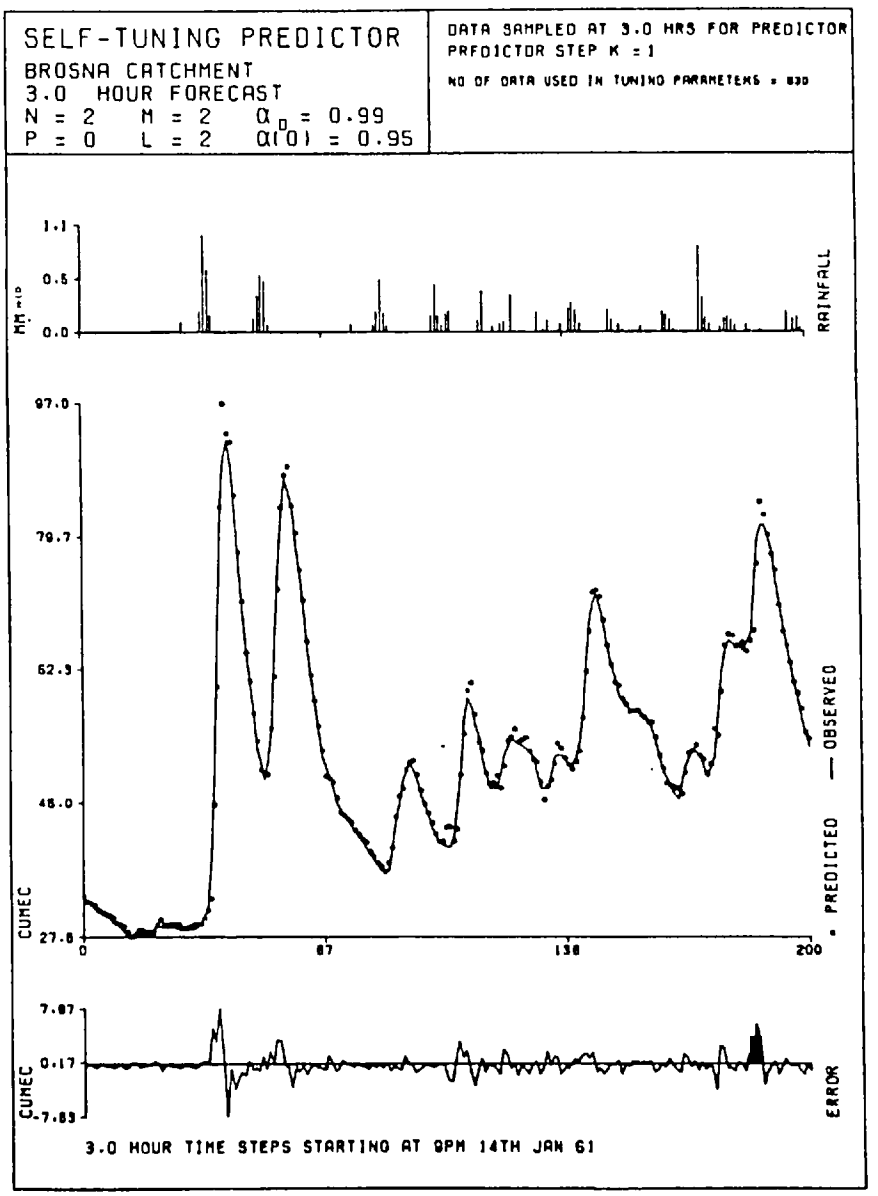


Fig. 5.10 PREDICTOR A: 1-step ahead forecasts with data sampled at 3-hourly increments

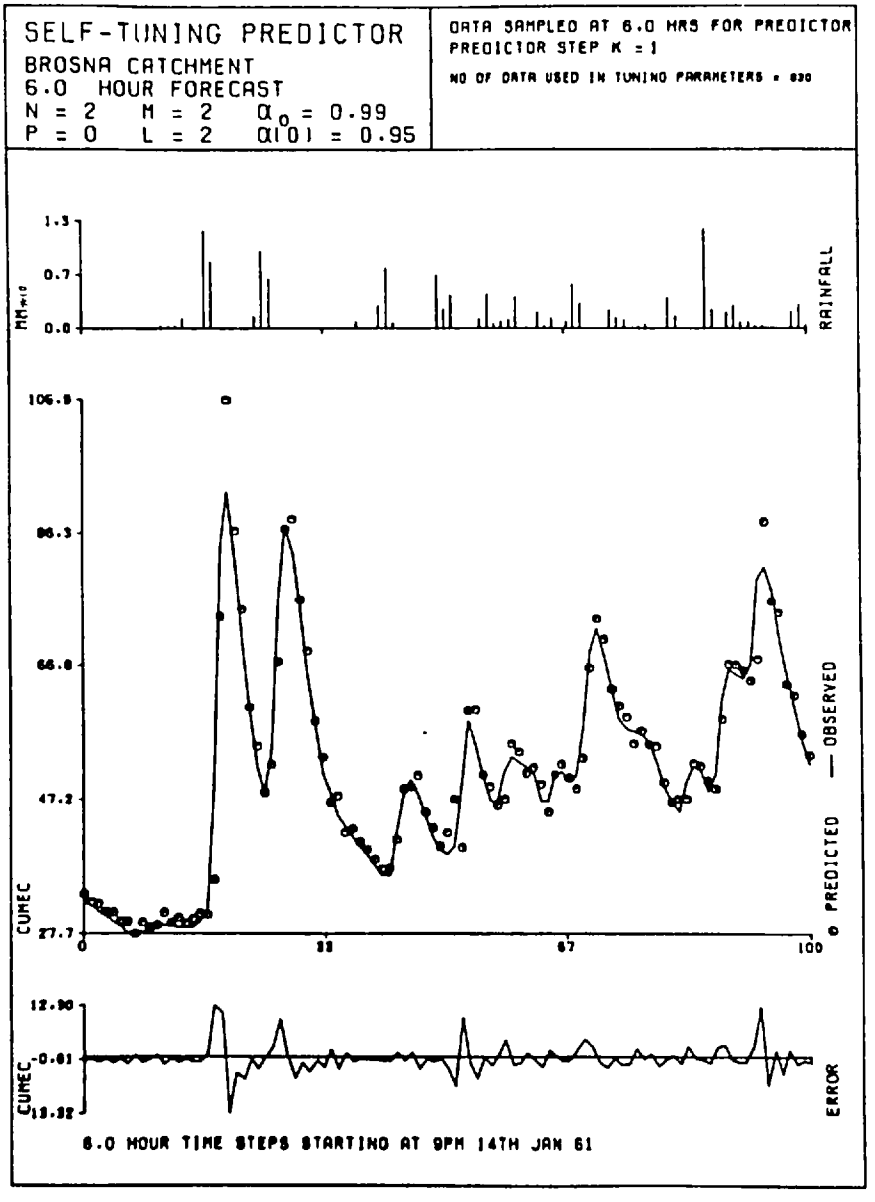


Fig. 5.11 PREDICTOR B: 1-step ahead forecasts with data sampled at 6-hourly increments

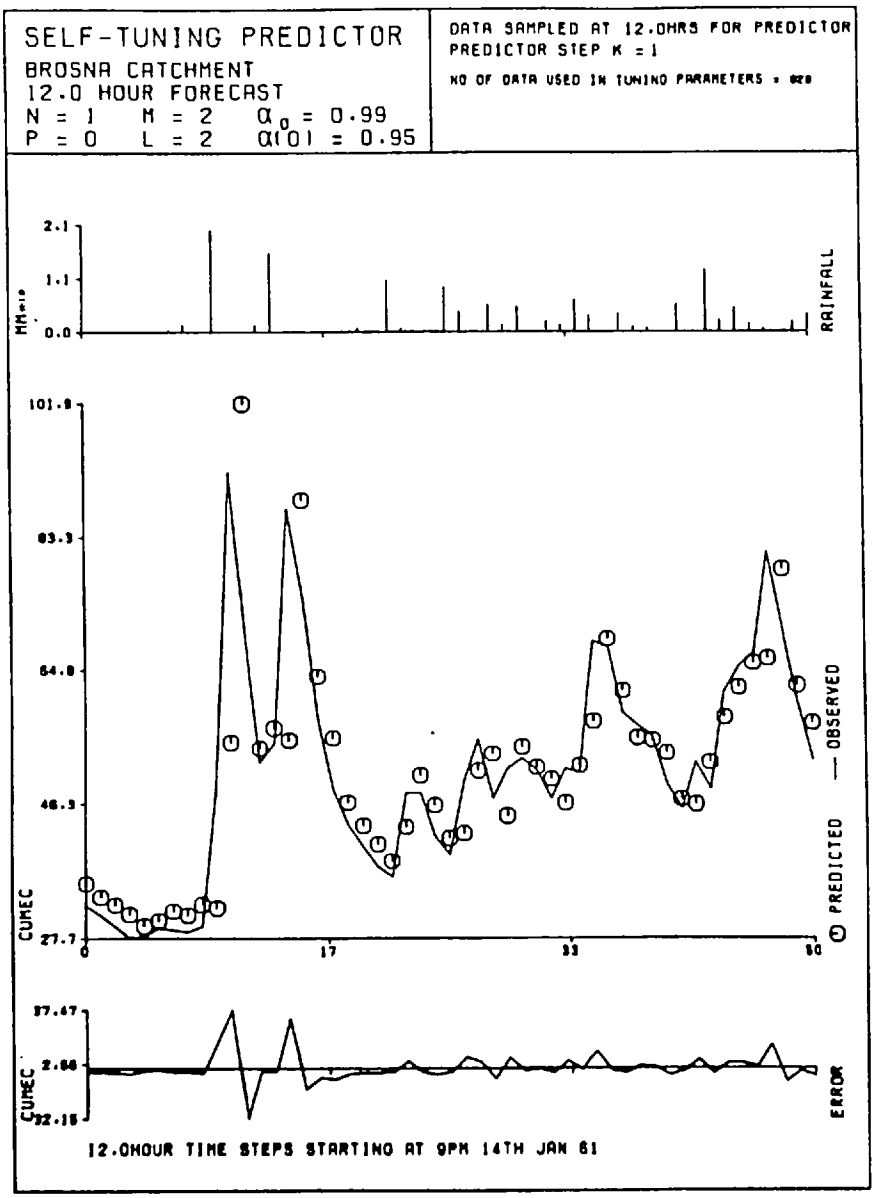


Fig. 5.12 PREDICTOR C: 1-step ahead forecasts with data sampled at 12-hourly increments

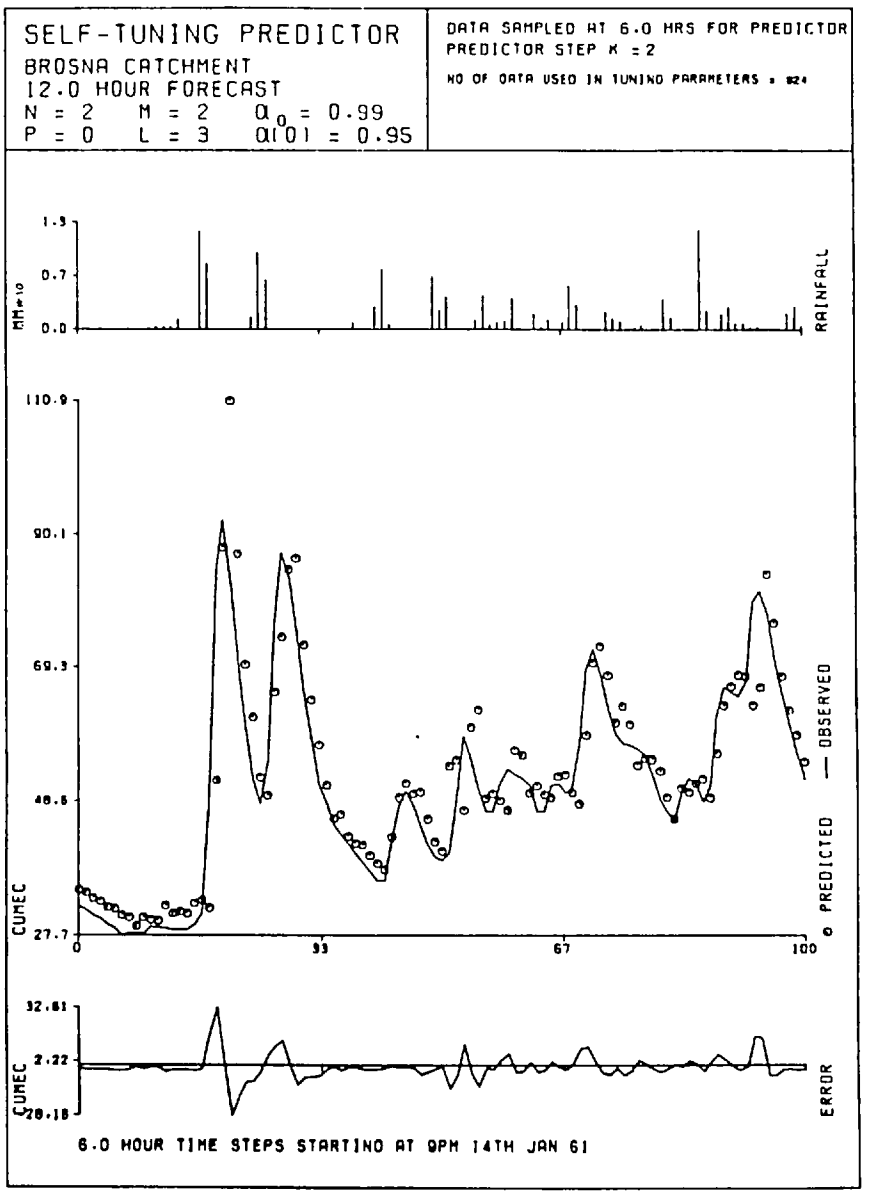


Fig. 5.13 PREDICTOR D: 2-step ahead forecasts with data sampled at 6-hourly increments

and 12 hours respectively (corresponding to the same length of historic record). This ensured that the parameters had converged (or attained their best values) before the statistical properties of the prediction errors were evaluated. Parameter values (coefficients of the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$) for the different predictors corresponding to the same time step are also shown in table 5.1.

PREDICTOR	A	B	C	D
Δt(hrs)	3	6	12	6
k	1	1	1	2
lead time (hrs)	3	6	12	12
n	2	1	1	2
m	2	2	2	2
ℓ	2	2	2	3
b	1	0	0	0
sample size N	800	400	200	400
error variance	2.49	12.32	74.44	43.49
mean error	-0.17	-0.28	-1.93	-1.516
Q-test (*)	53(41.3)	37(41.3)	50(41.3)	136(40.1)
S-test (**)	20(38.9)	23(40.1)	23(40.1)	37(38.9)
mean loss†	2.53	12.37	74.23	45.69
a_1	1.64	0.91	0.80	1.32
a_2	-0.66	-	-	-0.51
b_1	3.27	7.05	36.32	29.44
b_2	3.84	19.64	11.71	26.63
c_1	1.54	1.63	0.99	1.46
c_2	-0.46	0.23	-0.08	-0.64
c_3	-	-	-	0.05

† $\frac{1}{N}\sum_{t=1}^{N}\varepsilon(t)^2$ * χ^2_5 with 30-ℓ degrees of freedom ** χ^2_5 with 30-n-m degrees of freedom

Table 5.1 BROSNA : comparison of 4 self-tuning predictors. The statistical properties of the prediction errors were computed from sample size N

Comparison of the results of predictor A, B and C or D shows that the error variance increases as the lead time becomes greater. Also, corresponding to the increase in lead time are the decrease in magnitude of the a_1 parameters and an increase in the b_1 parameters. This indicates that the importance of the rainfall input to the predictor increases with the forecast lead time.

The Q (for $r_{\varepsilon\varepsilon}(\tau)$) and S (for $r_{\varepsilon\hat{q}}(\tau)$) values have also been shown together with the correponding χ^2 values to give the hypothesis test for model adequacy (see Section 3.3). For all four predictors the S-test indicated that $r_{\varepsilon\hat{q}}(\tau)$ was not significant, while the Q-test indicated that only for predictor B was $r_{\varepsilon\varepsilon}(\tau)$ not significant.

The portmanteau tests were however not used as a check for model adequacy as they have been shown to be unreliable tests, even for moderately large sample sizes (Davis, Triggs and Newbold, 1977; Chatfield and Prothero, 1973; and Ozaki, 1977). The tests are further invalidated as the residual series are not generated from an ARMAX model, but from a self-tuning predictor. Furthermore, the degrees of freedom for χ^2 are based upon the number of parameters in the predictor. It was found though that the tests provided useful information when comparing different predictors with the same Δt and k (in which case exactly the same data are processed for a given length of record).

A difficult situation may arise if predictors with the same lead time but different Δt and k values are compared. This is shown by predictors C and D where the portmanteau tests suggest that predictor C is better than predictor D, while comparison of the error variances suggests the converse. The latter test was assumed as affirmative. Though the variance is a strictly decreasing function of sample size, it seems unlikely that the 58% reduction in the error variances of predictor D over predictor C can be attributed to a 100% increase in the sample size used to compute the error variance of predictor D (the same period of data was used).

In general it was found that for lead times beyond 6 hours the forecast error variance rapidly increased. More clearly shown in table 5.1 is the increase in the magnitude of the mean error. This can be expected as the estimated impulse response shown in fig. 5.2 suggests a rainfall to runoff response time of between 3 and 6 hours.

5.5 FLOW FORECASTING FOR THE HIRNANT CATCHMENT

The 33.9 sq km catchment area for the River Hirnant forms a sub-catchment of the River Dee in North Wales (see fig. 2.2). Thin soil cover over impervious metamorphic rocks provides very little storage and together with steep channel slopes results in fast rainfall to runoff response. There is appreciable afforestation. Soils in the upper part of the catchment are peaty and drain slowly to sustain low flows throughout dry periods. Rainfall is well distributed throughout the year but there is a tendency for greater falls in the winter. Severe storms can occur at any time and in winter snowfall is erratic. A snowpack can develop in exceptional winters, leading to snowmelt runoff. Snowmelt did not occur during any of the periods of data used in this study.

The river is gauged at Plas Rhiwaedog at a natural river section, while rainfall is recorded by a dense network of Plessey tipping bucket raingauges. The data which are available at half hour sampling intervals are of high quality, having been used extensively for studies in conjunction with the Dee Weather Radar Project (see section 2.1). For this study hourly data comprising hourly discharge measurements (rather than averaged ½-hourly data) and 1-hour rainfall totals also were used.

Model structure estimation

The Box and Jenkins correlation approach as applied to the Brosna data in Section 5.4 was used to obtain an initial approximation to the impulse response. Figure 5.14 shows the sample autocorrelation functions of the hourly rainfall and runoff series and the partial autocorrelation function of the hourly rainfall series. A sample size of N = 1400 starting 9 a.m. 1 November 1972 was used. The autocorrelation function of the input (rainfall) series tails off after lag one, while the partial autocorrelation function cuts off after lag one, indicating that the rainfall series can be represented by a first order AR model (the prewhitening filter). The prewhitening filter obtained from the solution of the Yule-Walker equation was

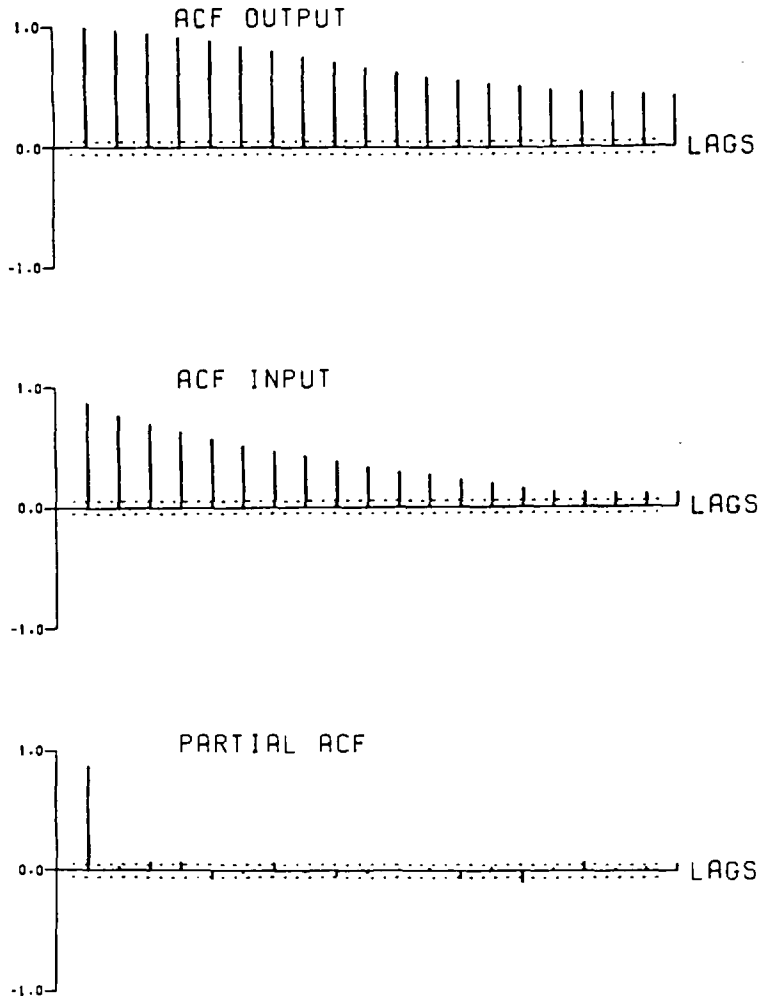


Fig.5.14 *HIRNANT DATA: Sample autocorrelation function of output (flow) and input (rainfall), and partial autocorrelation function for input. Sample size N = 1400 (95% confidence limits ±1.96/√N are shown).*

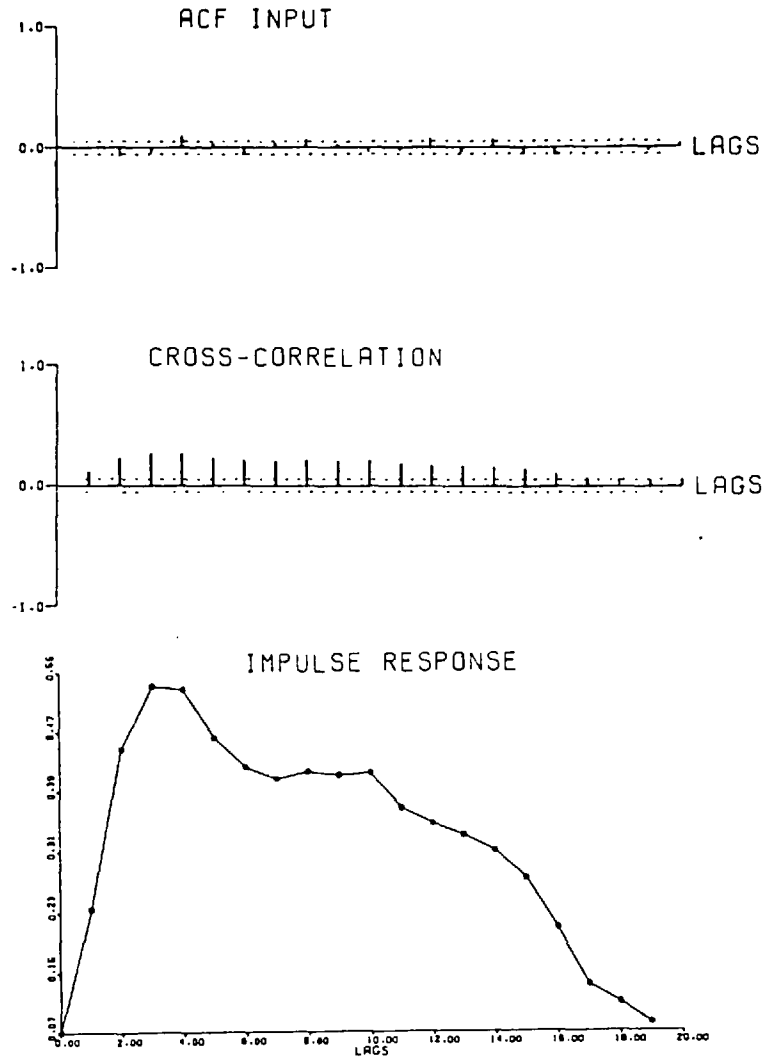


Fig. 5.15 *HIRNANT DATA: Sample autocorrelation function of prewhitened rainfall input, sample cross-correlation function for prewhitened input and runoff output, and approximate impulse response. Sample size N = 1400.*

$$r(t) = 0.88r(t-1) + e_r(t) \qquad (5.33)$$

The sample autocorrelation function of the prewhitened input and the sample cross-correlation function between the prewhitened input and the corresponding filtered output (discharge) are shown in fig. 5.15. Also shown in fig. 5.15 is the estimated impulse response function computed from the sample cross-correlation function.

Following the discussion in Section 5.4, the Box and Jenkins analysis gives an indication of the systems' dynamic characteristics. The estimated impulse response which has a time to peak of 3 hours, suggests that the rainfall to runoff pure time delay is between 0 and 1 hour. As expected, this shows that the Hirnant catchment has a very much faster rainfall to runoff response time than the Brosna catchment (compare fig. 5.2). Again however these results must be viewed with caution, as the estimated cross-correlation function appears to have been influenced to some extent by the non-stationarity present in the output series. No attempt was made to estimate the orders of the polynomials $A(z^{-1})$ $B(z^{-1})$ and $C(z^{-1})$ from the estimated impulse response.

As for the Brosna data in the previous section, model selection was based on the analysis of the prediction error series.

Comparison of results

A 2 month period of ½-hourly data commencing 9 a.m. 1 November 1972 was used in a comparative study of 2 self-tuning predictors (predictors A and B) assuming a 1 hour forecast lead-time. The short lead-time selected reflects the rapid catchment response time. Predictor A is a 1-step ahead predictor while predictor B is a 2-step ahead predictor. In each case, orders of the polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ were selected on the basis of minimum prediction error variance evaluated for the month of December. Initial parameter calibrations were carried out using the preceding month's data, while the starting conditions were $\phi^T(0) = \theta(0) = 0$,

α_o = 0.99, $\alpha(0)$ = 0.95 and $P(0) = I \times 10^4$. The results are shown in table 5.2.

PREDICTOR	A	B
Δt (hrs)	1	0.5
k	1	2
n	1	1
m	3	2
ℓ	2	3
b	0	1
sample size N	724	1448
error variance	0.15	0.06
mean error	0.04	0.06
Q-test	92	293
S-test	77	86

Table 5.2 HIRNANT : Comparison of 2 self-tuning predictors. The statistical properties of the 1 hour lead-time prediction errors were computed from sample size N

The delay parameter b was estimated from the impulse response shown in fig. 5.15. The important results demonstrated in this comparison is that the 1-step ahead predictor (predictor A) performed better than the 2-step ahead predictor (predictor B). This is clearly shown by comparison of the prediction error variances. Furthermore the computed Q values show that the prediction errors of predictor B are more autocorrelated than those of predictor A. No attempt was made to refer Q and S to the percentage points of the χ^2 distribution as the portmanteau tests were suggested as being invalid in Section 5.4. It was also found that removing the mean level in the Hirnant streamflow record did not result in better parameter estimation (hence the mean level was not removed).

Figure 5.16 shows $r_{\varepsilon\varepsilon}(\tau)$ and $r_{\varepsilon\hat{q}}(\tau)$ for predictor A, and computed from a sample size of 724. The parameters were initially tuned over 764 data samples. The 95% confidence limits are ± 0.07 which shows that $\varepsilon(t)$ is autocorrelated, and $\varepsilon(t)$ and $\hat{q}(t|t-k)$ are cross-correlated. The large values for Q and S clearly reflect this.

Figures 5.17 and 5.18 show the results of applying predictors A and B to the event starting at 1 a.m. 4 December 1972. Visual inspection of the two predicted hydrographs indicates that the 1-step ahead predictor gives the best results. Furthermore, while the predictor appears inadequate on the basis of $r_{\varepsilon\varepsilon}(\tau)$ and $r_{\varepsilon\hat{q}}(\tau)$, fig. 5.17 shows that the 1-hour lead-time forecasts are in fact reasonable. The 1-step ahead forecast for the flood peak event that occurred on 6 December 1972 (shown as the largest of the 3 flood peaks in fig. 5.17) has a peak error of 2.2% while the maximum error on the rising limb as a percentage of the observed peak is 11.4%.

The parameter variations for predictor A after the parameters had been initially tuned over 764 data samples are shown in fig. 5.19. The parameter variations are again characteristic of self-tuning algorithms, with large input disturbances causing all the parameters to vary in unison. It is interesting to note that the most stable parameter is $\hat{\theta}(1)$ (which corresponds to α_1), the autoregressive parameter.

The 1-step ahead prediction of the storm event which occurred on 12 December 1972 is shown in fig. 5.20. This example demonstrates the tendency for the predictor to under-estimate the rising limb and over-estimate the initial falling limb of the hydrograph (see the plot of the prediction errors). This is predicated by the dominance of the autoregressive component of the linear model which is being applied to a non-linear system.

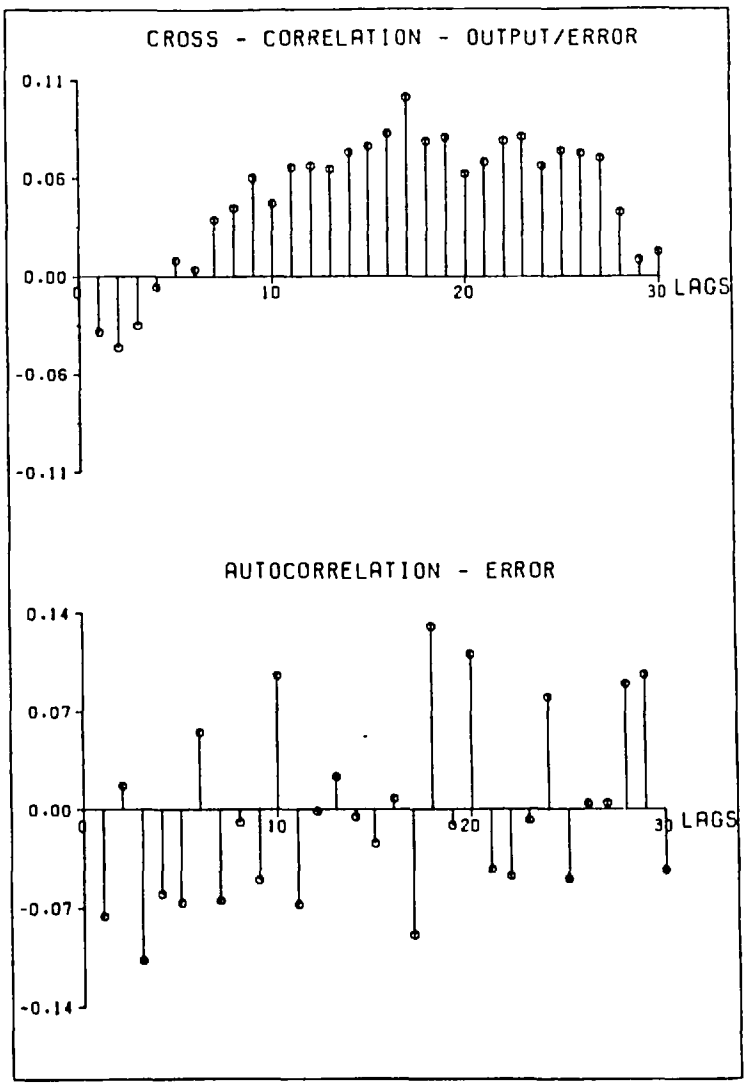


Fig. 5.16 PREDICTOR A: Sample autocorrelation function $r_{\varepsilon\varepsilon}(\tau)$ and sample cross-correlation $r_{\varepsilon\hat{q}}(\tau)$ for the Hirnant data. The 95% confidence limits are ±0.07, and the sample size N = 724.

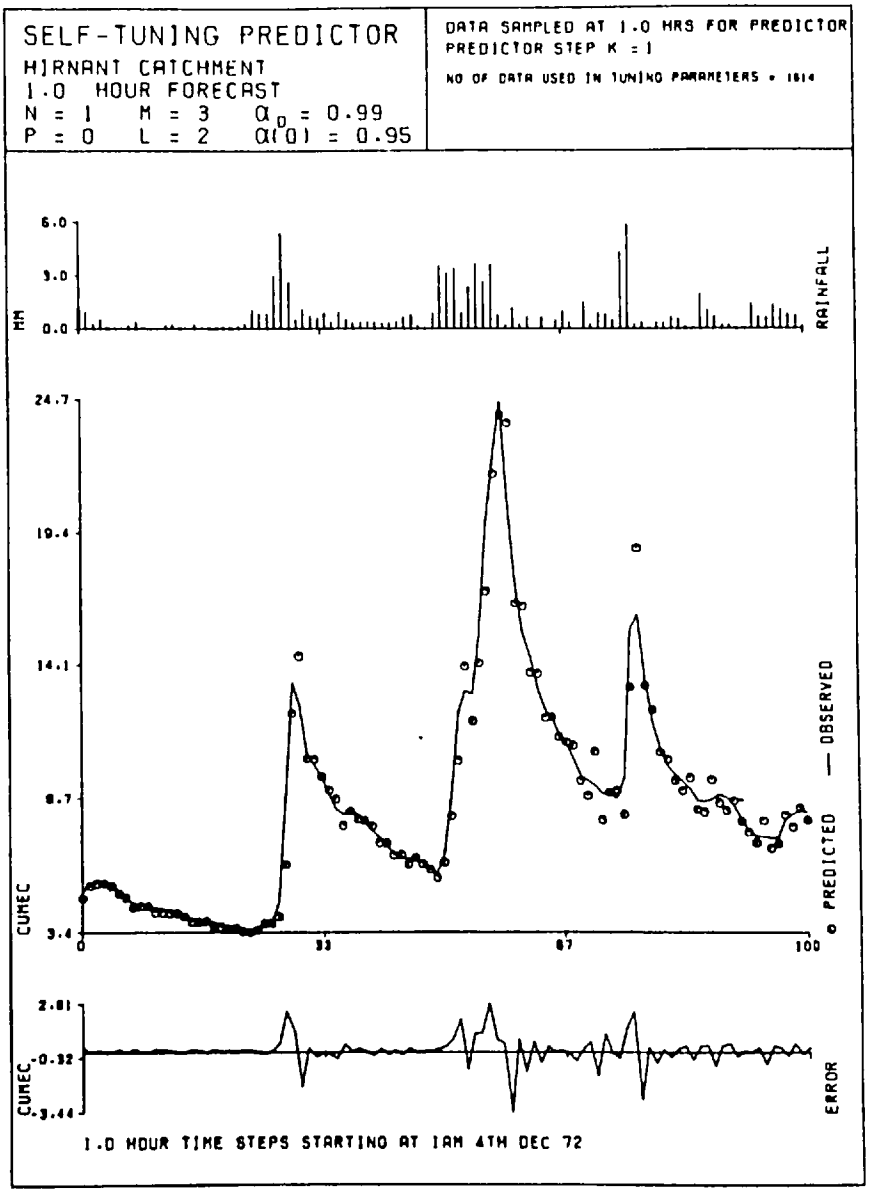


Fig. 5.17 PREDICTOR A: 1-step ahead forecast with data sampled at hourly increments

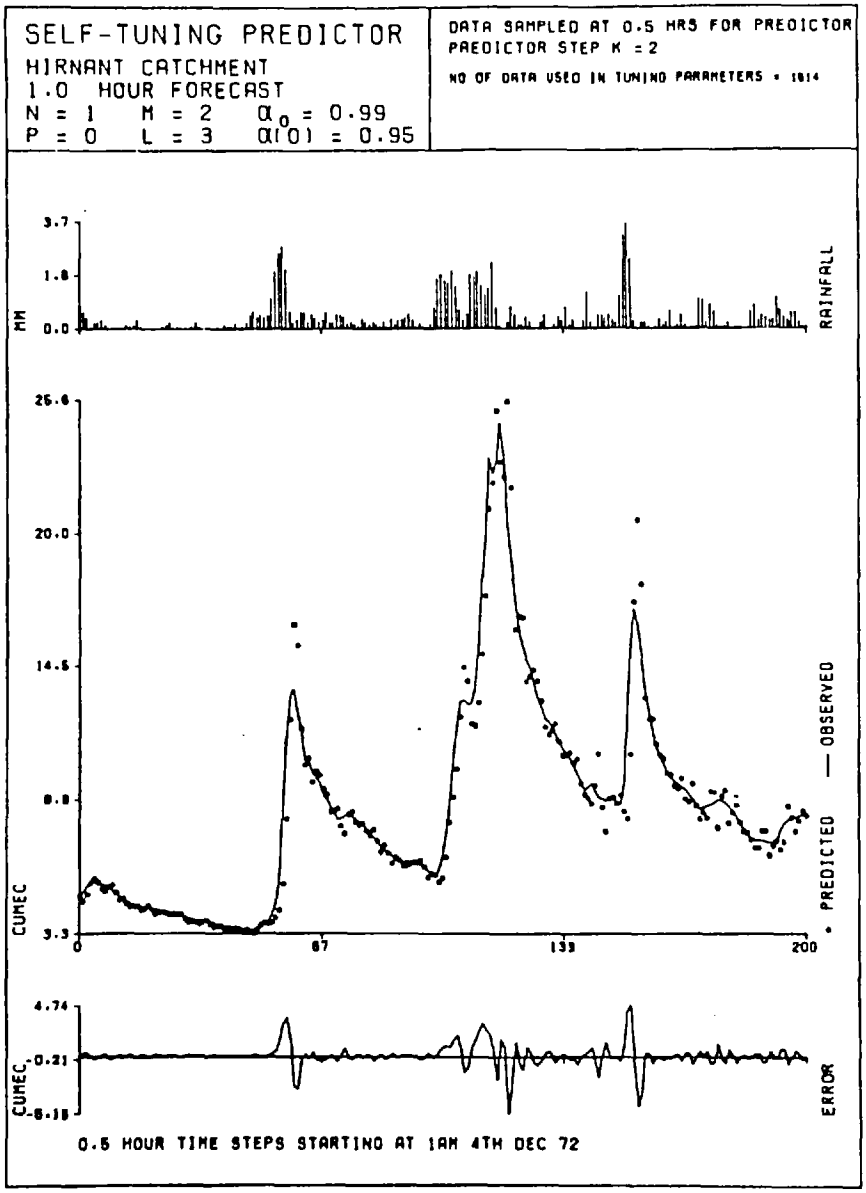


Fig. 5.18 PREDICTOR B: 2-step ahead forecast with data sampled at 1/2 hourly increments

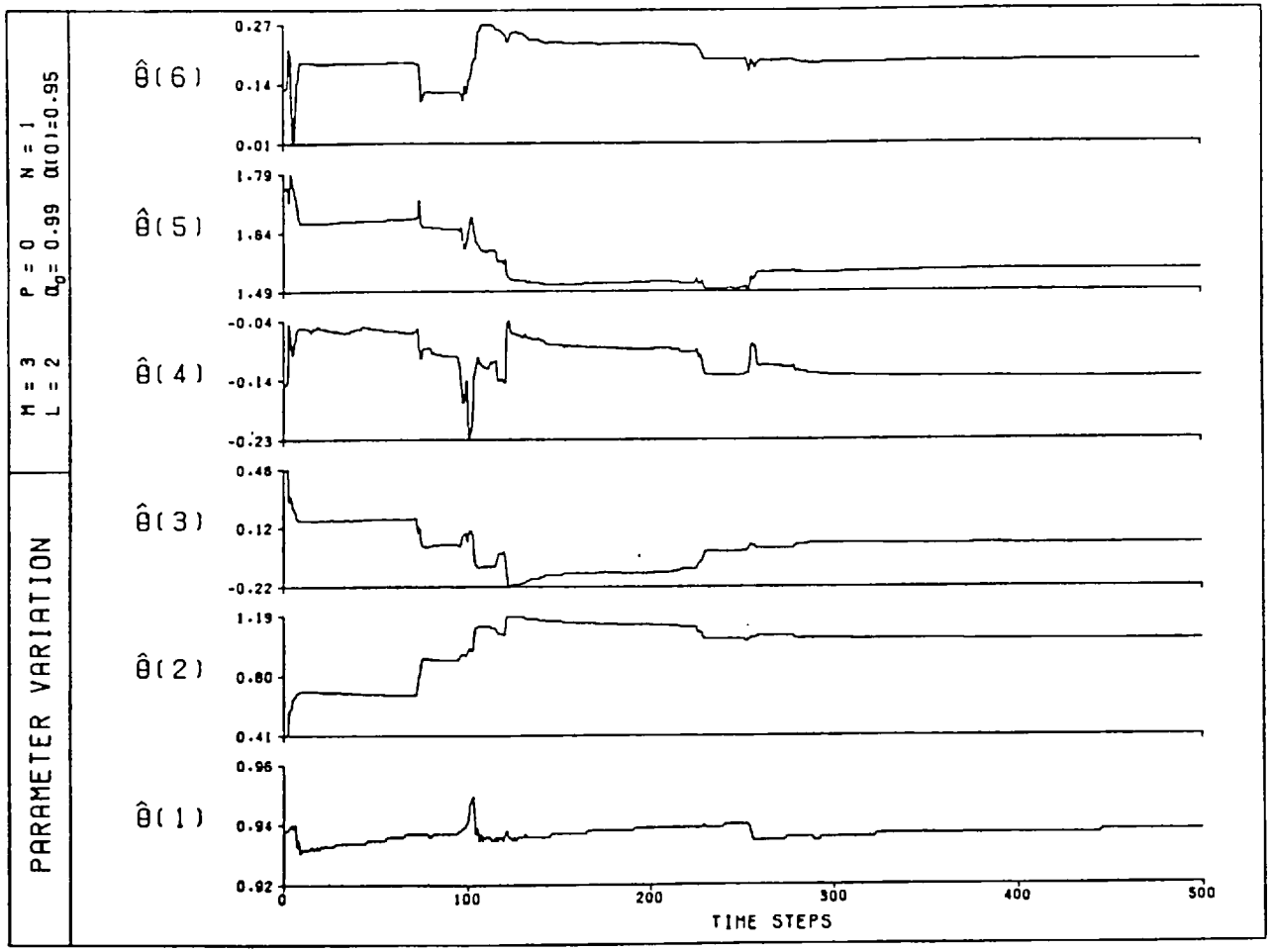


Fig. 5.19 Recursive parameter estimates for PREDICTOR A (see fig. 5.17)

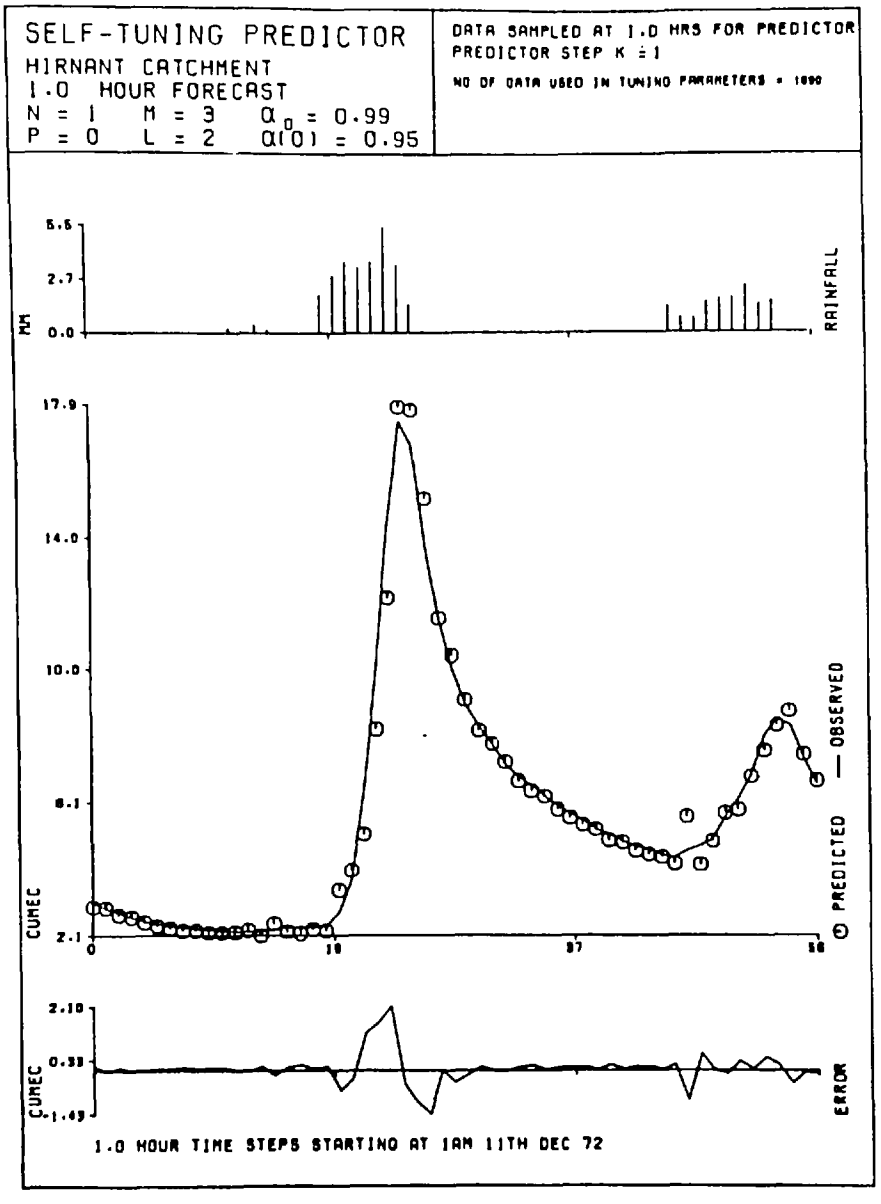


Fig. 5.20 *Flow forecasting for the River Hirnant using the self-tuning predictor. The plotted prediction errors show the tendency for the predictor to underestimate the rising limb and overestimate the initial part of the recession.*

5.6 PERLOG AS AN INPUT TO THE SELF-TUNING PREDICTOR

It is well-known that the runoff response of a catchment to rainfall is highly non-linear, and that the ARMAX model incorrectly assumes that it is a linear process. In common with other forecasting models based on an ARMAX representation of the rainfall-runoff process, the self-tuning predictor tends to under-estimate the hydrograph rising limb and over-estimate the hydrograph peak and the initial stage of the recession. This can to a large extent be attributed to both model inadequacy and poor rainfall data. Often the observed rainfall does not fully represent the true input to the catchment system because of spatial variability, and/or the catchment antecedent conditions are not adequately modelled (particularly by linear models). Then, by virtue of the recursive parameter estimate, flow forecasts will be to a greater extent based on past flows (as an autoregressive process of the output) while little reliance will be placed on the rainfall component of the predictor.

A general black-box method for modelling a non-linear system is provided by the use of the Volterra functional series. There have been several applications of this method to hydrological systems (Amorocho, 1963; Hino, Sukigara and Kikkawa, 1971; Amorocho and Brandstetter, 1971; Bidwell, 1971; Diskin and Boneh, 1972; Diskin and Boneh, 1973; Zand and Harder, 1973; Quimpo, 1975). The problem becomes one of determining the kernal functions of the Volterra series, which usually involves large amounts of data and computations involving high dimensional matrices. Consequently (at present) the method is not suitable for real-time forecasting, though interestingly enough Szöllösi-Nagy (1976) has shown that the Volterra series can be formulated within a state-space framework. This method will not, however, be pursued further here.

An ARMAX model is non-linear if one or more auxiliary inputs which are non-linear functions of the system observations (inputs or outputs) are assumed. In this section the perlog function (Appleby,

1965) will be taken as an auxiliary input to the self-tuning predictor. The perlog $p\ell(t)$, is defined as the differential of the log of the observed runoff,

$$p\ell(t) = \frac{d}{dt}(\log(q(t)) = \frac{1}{q(t)}\frac{d}{dt}q(t) \qquad (5.34)$$

A hydrograph with its associated perlog is shown in fig. 5.21. The perlog can be divided into a forcing phase and a relaxed phase. Clearly when considering the perlog as an input (the runoff being the output) (a) the positive portion of the perlog can be considered as a forcing function associated with the hydrograph rising limb while (b) the negative portion of the perlog representing the relaxed phase occurs when the river is in a state of recession. The perlog characteristically falls to a constant negative value as shown when a long recession occurs. This could be interpreted as a catchment state corresponding to little or no surface runoff and is referred to as the linear reservoir phase (Appleby, 1965).

In general, a particular rainfall event occurring on a wet catchment will result in a larger runoff response (and a larger rate of change of flow) than if the catchment is initially in a dry state. This is due to the additional storage of the dry catchment. Furthermore, when there is no significant surface runoff, the river level may reflect the catchment antecedent conditions (i.e. if the catchment is wet the flow will be greater than had the catchment been dry). The forcing phase of the perlog which is proportional to the rate of change of flow and inversely proportional to the flow itself should thus be highly correlated with the rainfall input. Note that the perlog also reaches a peak when the rate of change of the flow is a maximum and falls to zero when the hydrograph peak is obtained. The perlog can thus be used as an addition to the predictor input containing both information on the catchment antecedent conditions and information on the catchment rainfall input (rainfall observations are usually associated with large measurement errors).

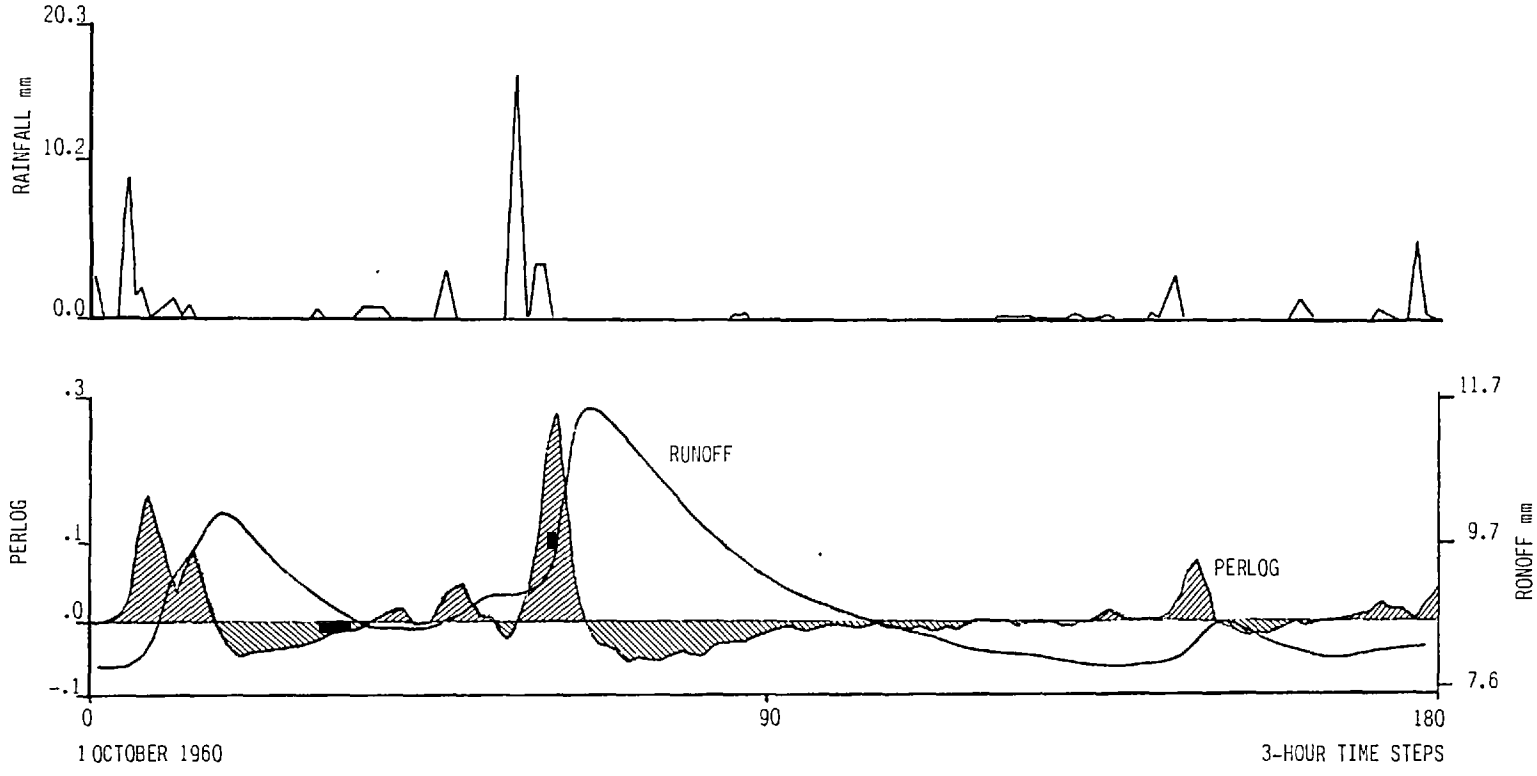


Fig. 5.21 A hydrograph of the River Brosna with the corresponding perlog

The perlog is computed by using a central difference approximation of the differential term in equation 5.34 so that

$$p\ell(t) = \frac{1}{q(t)} \left[\sum_{i=-j}^{j} iq(t+i)/(2 \sum_{i=1}^{j} i^2) \right] \qquad (5.35)$$

The finite difference scheme uses (2j+1) points with j = 1 or 2 giving satisfactory results. The necessary p perlog terms to be included in the predictor are simply added to the data vector $\phi(t)$:

$$\phi(t) = [\hat{q}(t|t-k) \ldots ; p\ell(t-j) \ldots p\ell(t-j-p+1); r(t-b) \ldots ; \varepsilon(t) \ldots] \qquad (5.36)$$

and correspondingly the dimension of the parameter vector is increased. Clearly as q(t+j) is unknown at time t, pℓ must be lagged by j time steps.

Flow forecasting for the River Brosna

One and two step ahead flow forecasts for the River Brosna starting at 9 a.m. 20 November 1960, using the self-tuning predictor with the perlog function as an auxiliary input are shown in figs. 5.23 and 5.25 while the corresponding forecasts using the self-tuning predictor without the perlog input are given in figs. 5.22 and 5.24. Details of the respective predictors A B C and D are given in table 5.3.

The appropriate predictor in each case was selected with the criteria of minimum prediction error variance and the Q and S statistics. It was found that for a 1-step ahead predictor, the best results were obtained using the 5 point perlog approximation (j=2) while for the 2-step ahead predictor the 3 point approximation (j=1) gave the best results. This is because the 2-step ahead predictor giving the best results requires that the perlog delay be less than 2 while the perlog function must be delayed by j time steps. Comparing the 5 point perlog and 3 point perlog functions shown in figs. 5.23 and 5.25 respectively, it is evident that the 5 point finite difference approximation gives a much smoother function than the 3 point approximation.

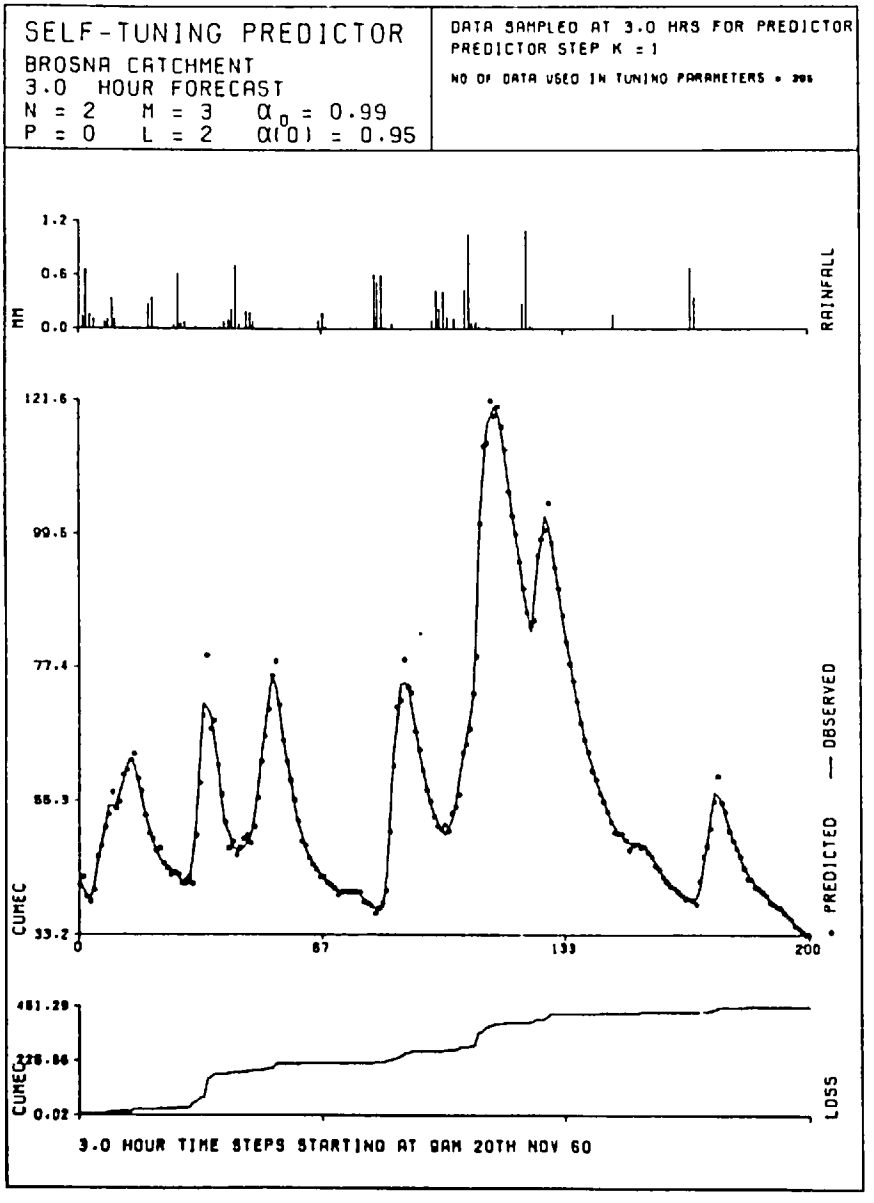


Fig. 5.22 PREDICTOR A: 1-step predictor without perlog input. Cumulative loss $\Sigma\varepsilon(t)^2$ is shown.

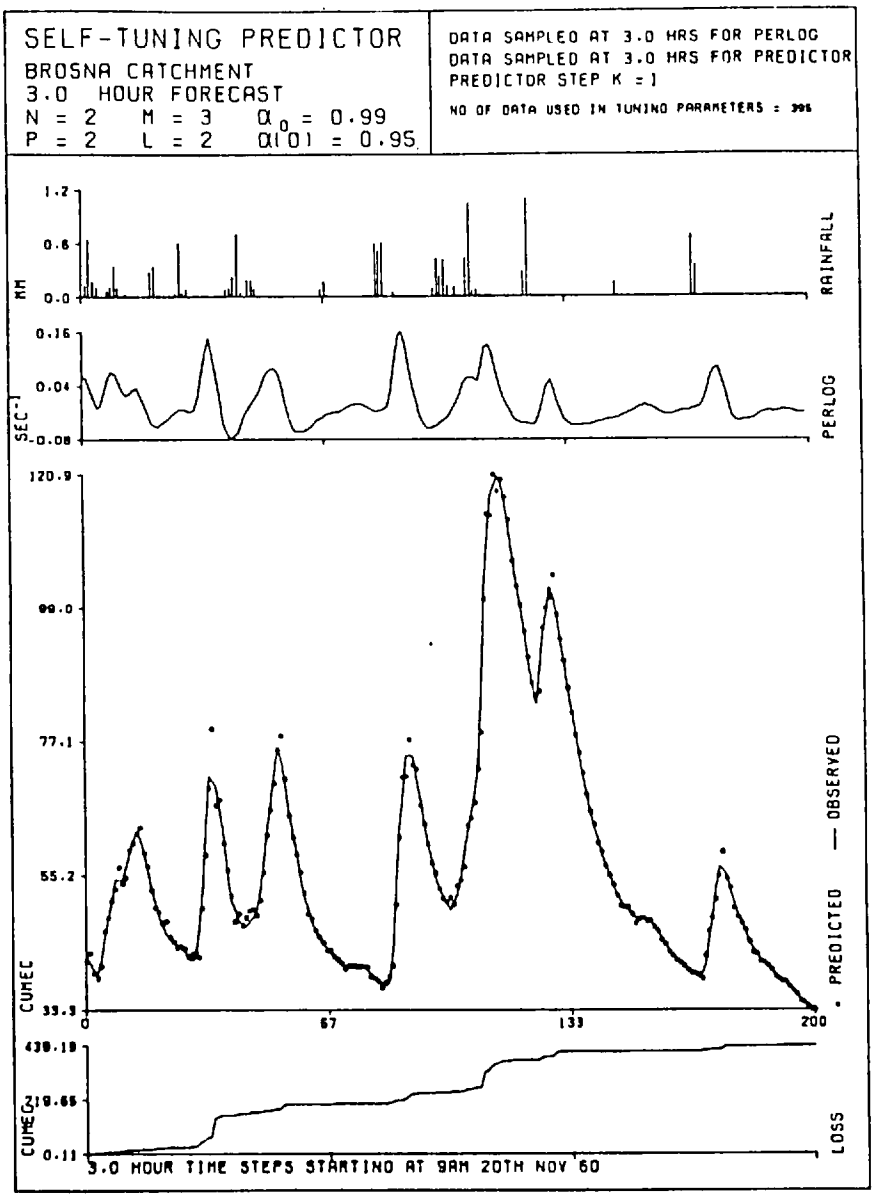


Fig. 5.23 PREDICTOR B: 1-step ahead predictor with perlog input. Cumulative loss $\Sigma\varepsilon(t)^2$ is shown.

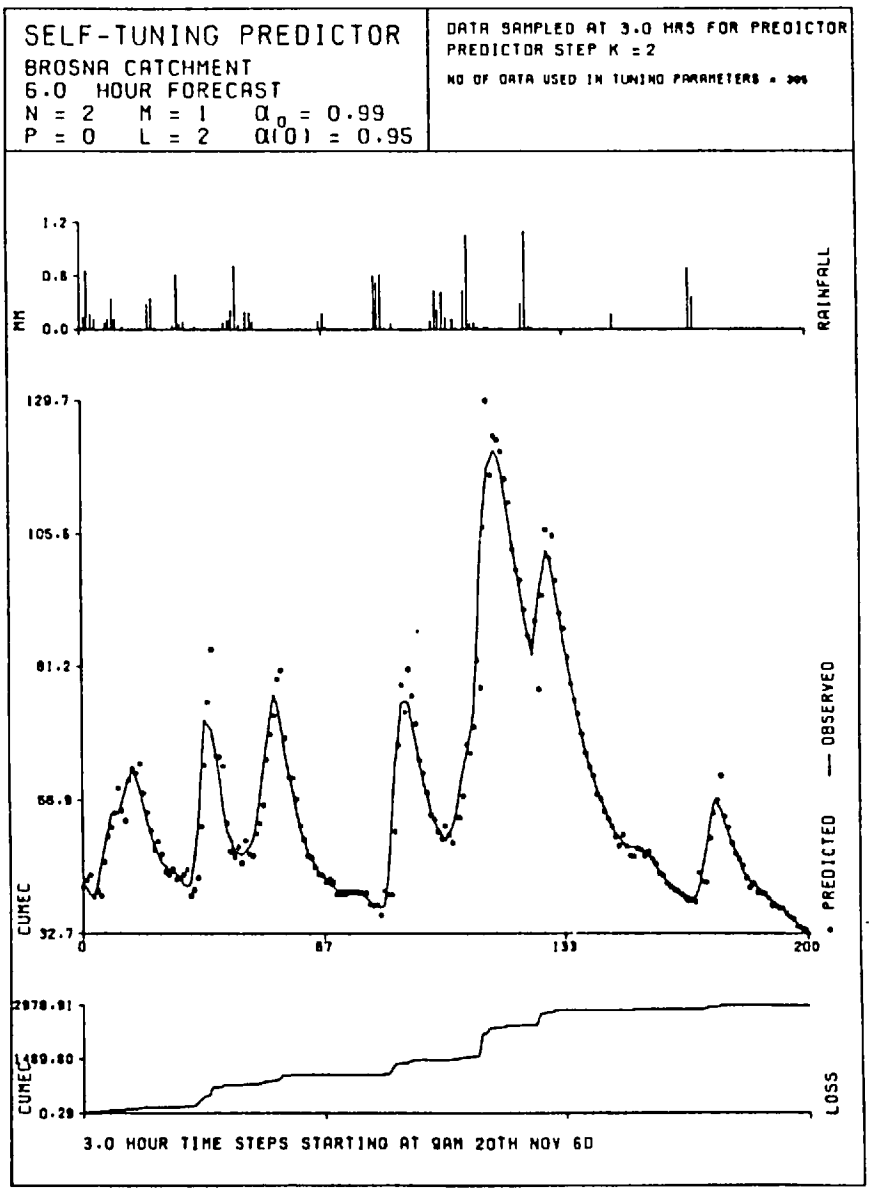


Fig. 5.24 PREDICTOR C: 2-step ahead predictor without perlog input. Cumulative loss $\Sigma\varepsilon(t)^2$ is shown.

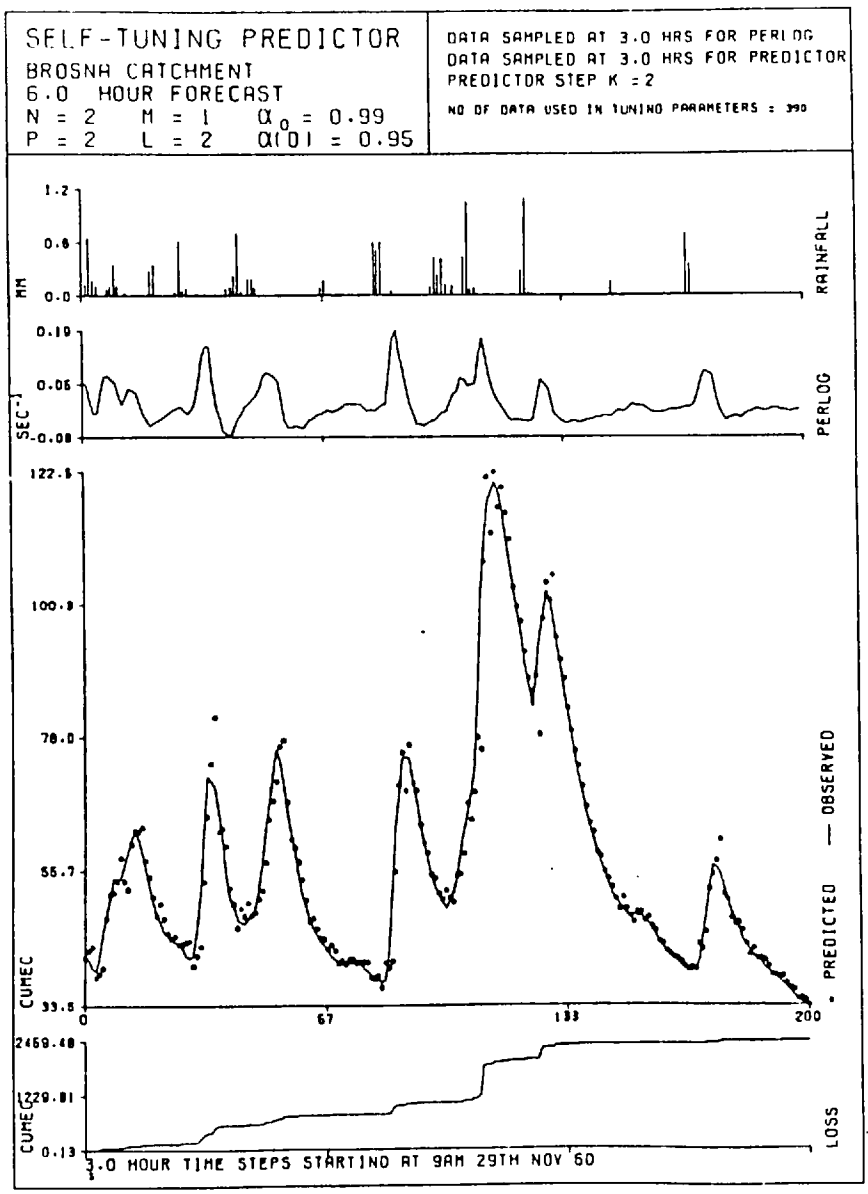


Fig. 5.25 PREDICTOR D: 2-step ahead predictor with perlog input. Cumulative loss $\Sigma\varepsilon(t)^2$ is shown.

PREDICTOR	A	B	C	D
lead time (hrs)	3.0	3.0	6.0	6.0
Perlog input	no	yes(j=2)	no	yes(j=1)
Δt (hrs)	3.0	3.0	3.0	3.0
k	1	1	2	2
n	2	2	2	2
p	0	2	0	2
m	3	3	1	1
ℓ	2	2	2	2
b	0	0	0	0
sample size N	500	500	500	500
error variance	1.81	1.71	11.04	9.01
mean error	-0.19	-0.21	-0.11	-0.19
Q-test	96	81	188	154
S-test	32	8	33	15
peak error %	1.10	1.51	-2.22	-1.55
max error on rising limb %	6.28	-3.52	20.42	21.55

Table 5.3 BROSNA : Comparison of 4 self-tuning predictors. The perlog auxiliary input is applied to predictors B and D

The effect of including the perlog input is to reduce the prediction error variance as shown in table 5.3. This is shown more clearly by comparing the cumulative loss (square error) of the 2-step ahead predictions shown in figs. 5.24 and 5.25. Visual inspection of the respective predicted hydrographs shows that the perlog input reduces the overestimation of the peak flow. Figure 5.26 shows that the parameter estimates for a 2-step ahead predictor are more sensitive to disturbances than a 1-step ahead predictor. The parameters corresponding to the perlog input ($\hat{\theta}(3)$ and $\hat{\theta}(4)$) are also shown.

The sample autocorrelation function $r_{\varepsilon\varepsilon}(\tau)$ and the sample cross correlation function $r_{\varepsilon\hat{q}}(\tau)$ for the same example are given in fig. 5.27. The sample size used was 500, for which the 95% confidence limits are ± 0.09. At lag 2, $r_{\varepsilon\varepsilon}(\tau)$ is significantly non-zero which suggests (by equations 5.21) that the parameter estimates of the 2-step ahead predictor with perlog input do not satisfy the self-tuning requirement. Table 5.3 also shows, by reference to the Q values, that the 2-step ahead predictors in particular do not

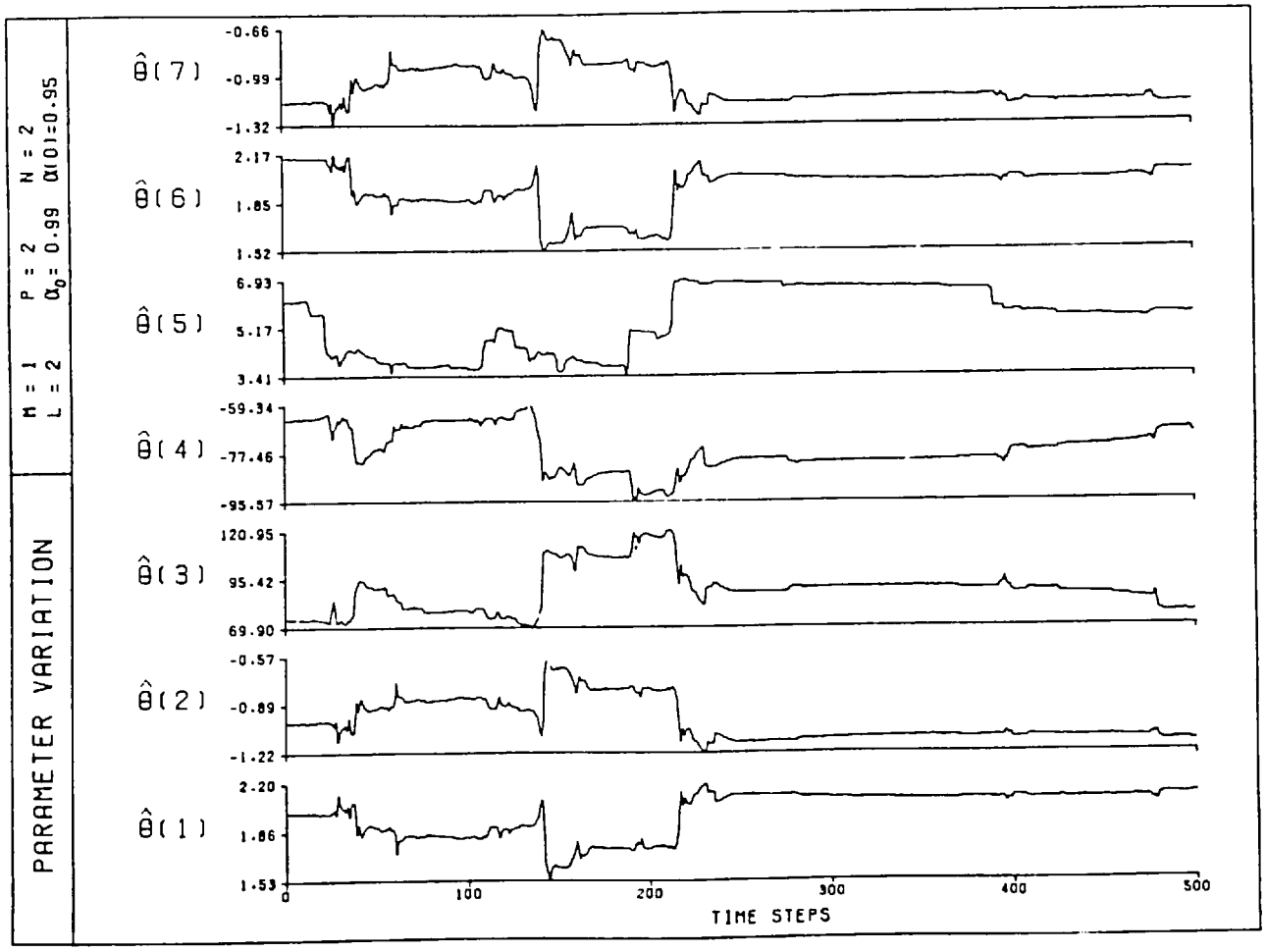


Fig. 5.26 Recursive parameter estimates for PREDICTOR D

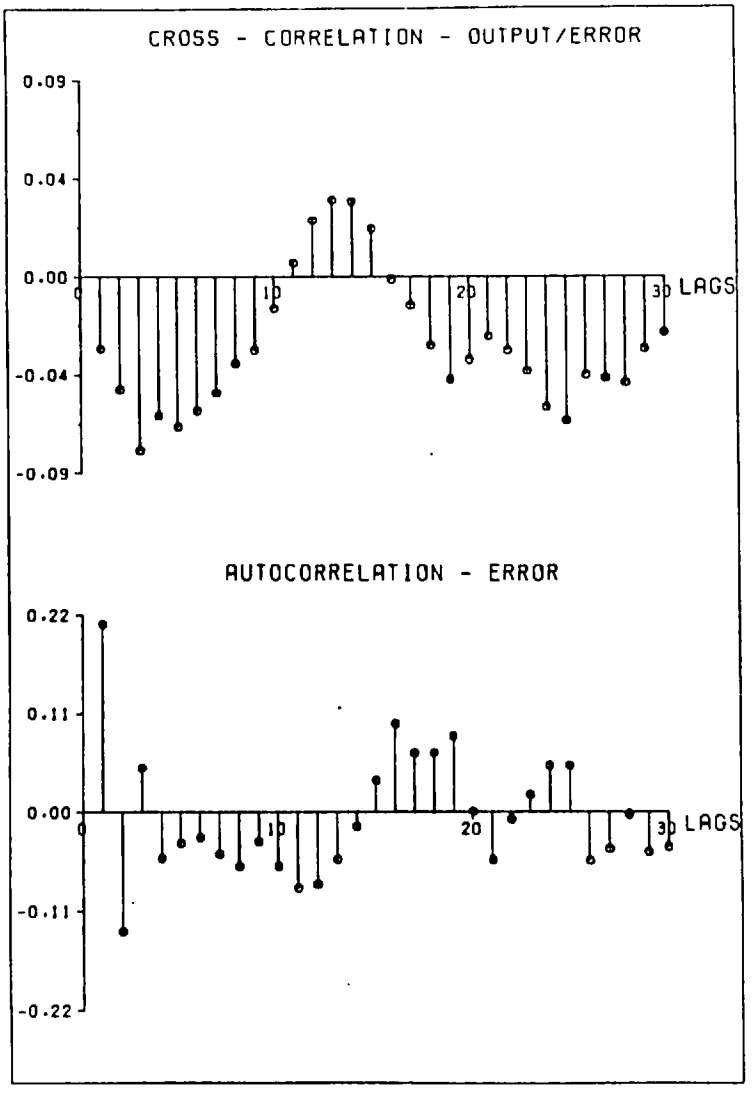


Fig. 5.27 *Sample autocorrelation function $r_{\varepsilon\varepsilon}(\tau)$ and sample cross-correlation function $r_{\varepsilon\hat{q}}(\tau)$ for PREDICTOR D. The 95% confidence limits are ±0.09 for a sample size N = 500.*

satisfy the self-tuning requirement. This supports the argument that a 1-step ahead predictor is generally better than a 2-step ahead predictor (compare the Q and S values for predictor B in table 5.1 to those for predictor C in table 5.3).

5.7 SELECTION OF THE FORECAST VARIABLE

So far it has been assumed that the forecast variable is the discharge $q(t)$, and that the rainfall-runoff process can be represented by an ARMAX model such as equation 5.9, i.e.

$$A(z^{-1})q(t) = B(z^{-1})r(t) + C(z^{-1})e(t) \qquad (5.37)$$

Discharge measurements are usually obtained from the observed stage (river level) at a hydraulic control section using a stage-discharge relationship. This relationship referred to as the rating curve is built up through discharge measurements for a range of stages using current meter gauging methods. Frequent recalibration of the rating curve is often required for an unstable natural control section. The discharge, however, may not be required for flood warning objectives, with the necessary information being given by the stage. Indeed the forecasted discharge is often converted to the corresponding stage in this respect.

Thus, the stage $h(t)$ itself can be assumed as the forecast variable so that the system model, equation 5.37, becomes

$$A(z^{-1})h(t) = B(z^{-1})r(t) + C(z^{-1})e(t) \qquad (5.38)$$

Furthermore, if stage observations at an upstream station are used as an input to the model for forecasting the downstream discharge, the following model is feasible:

$$A(z^{-1})q(t) = B(z^{-1})r(t) + C'(z^{-1})h(t) + C(z^{-1})e(t) \qquad (5.39)$$

where $C'(z^{-1})$ is a polynomial in the backward shift operator for the auxiliary input variable $h(t)$.

The stage-discharge relationship

It is well-known that there cannot be a single unique stage-discharge

relationship, so that discharge is not a function of depth alone (Henderson, 1966). For a single flood event, the stage-discharge path traced will be as shown in fig. 5.28. This is known as the loop-rating curve, and clearly for a given stage, the discharge will be greater for the rising limb than for the falling limb of the hydrograph.

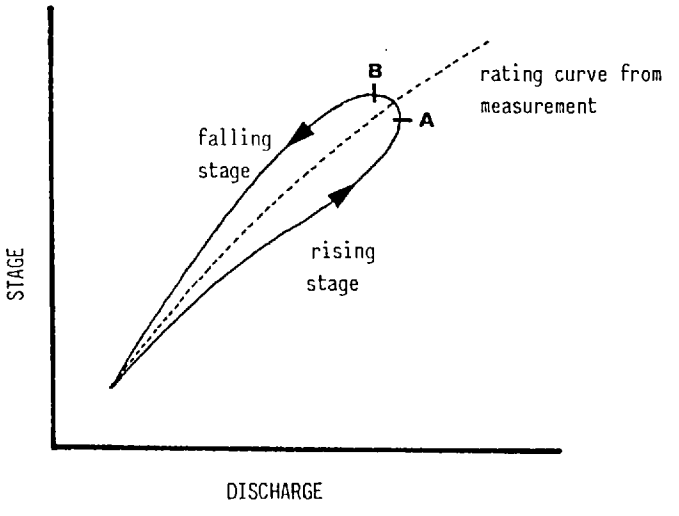


Fig. 5.28 The loop rating curve for a single flood and the estimated rating curve derived from measurement

If discharge is obtained from stage using the approximated rating curve shown in fig. 5.28, the true discharge will be under-estimated for the rising limb of the hydrograph and over-estimated for the falling limb. Thus, if a forecast of discharge is required, for instance as an input to a flow routing model, large errors may result due to the erroneous conversion of stage to discharge. This problem, which will not be discussed further here, is inherent in the estimation of the rating curve and not attributable to the forecasting model used.

When considering the flood peak it is also important to note that the maximum discharge (at point A in fig. 5.28) occurs before the maximum stage (at point B), though this is not shown in the estimated

rating curve. However due to the discrete time system models and fixed interval data sampling used for flow forecasting, the estimated peak may occur in time at neither the time peak discharge nor the maximum discharge. This problem is related to the data sampling increment Δt and the forecast model time-step k.

Forecasting stage

When the flow forecasting model, equation 5.37, is used to forecast the stage, as shown in fig. 5.29, the observed stage (a) is converted to the discharge (b), which is then used in the model to forecast the discharge (c). Note that the discharge obtained with the rating curve is used for model calibration. Using the same rating curve, the forecast is finally converted to the stage forecast (d). The stage forecast error is not due to errors in the

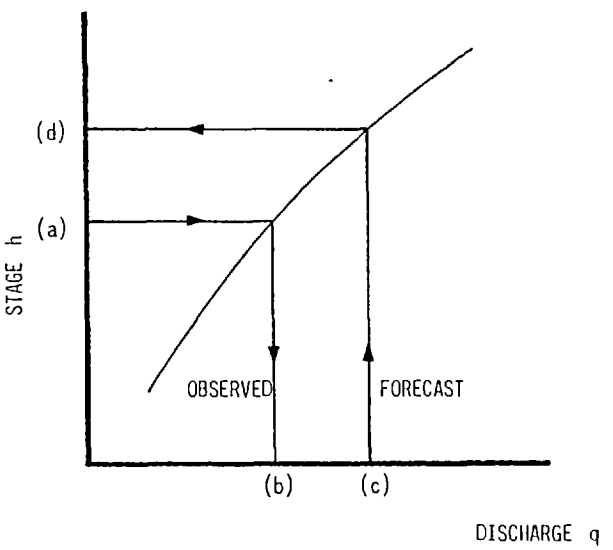


Fig. 5.29 Forecasting of the stage when the forecast variable is the discharge

stage-discharge relationship, even though the discharge forecast error may be related to the stage forecast error through the rating curve.

The discharge is a non-linear function of stage. Thus, while

discharge is a non-linear function of rainfall, it may be heuristically argued that stage is an even more non-linear function of rainfall. This is important when linear models are being considered,and particularly so if the river gauging is carried out at channel sections with highly non-uniform cross-sections as for example shown in fig. 5.30.

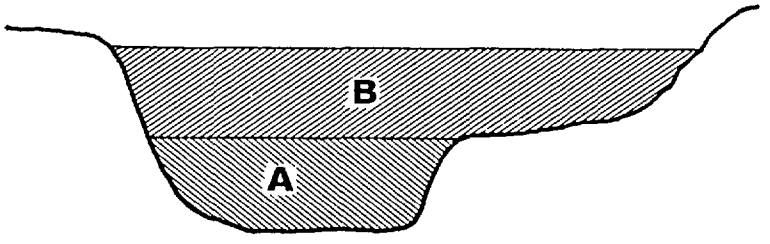


Fig. 5.30 Channel cross-section with non-uniform cross-section (in relation to the depth-surface width ratio)

The stage-discharge relationship for the above example exhibits a threshold effect when the stage transitions from zone A to zone B and clearly a linear model relating the rainfall input to the stage would be highly unsatisfactory. In general, the variation in stage for a given rainfall input is dependent on the channel cross-section geometry.

Forecasting cross-sectional area of flow

Appleby (1977) has suggested that an alternative to using the stage or the discharge as the forecast variable, is to use the cross-sectional area of the river flow x(t). The resulting ARMAX model is given by

$$A(z^{-1})x(t) = B(z^{-1})r(t) + C(z^{-1})e(t) \qquad (5.40)$$

Clearly the response of the area x(t) to rainfall input will be less sensitive to the channel geometry than the response of the stage h(t) to rainfall input. However for rough channel beds, where the local velocity gradients are probably high, it is not obvious that x(t) is a suitable forecast variable when linear models are being considered.

Comparative results of stage forecasts for the Hirnant

The River Hirnant was selected as a case study to demonstrate the forecasting of the stage by the three alternative methods using equations 5.37, 5.38 and 5.40. The characteristics of the Hirnant catchment are described in Section 5.5.

The River Hirnant is a turbulent stream with a steep bed slope, and large boulders distributed alongside the dry-weather flow channel are evidence of the extremely high velocities which may occur. The river gauging station consists of a float well, with connecting pipe to measure the water level upstream of a natural rock control, which has a free overfall at low and medium flows. A straight approach length of about 60 metres exists.

The rating curve (obtained from the Institute of Hydrology, Wallingford) is given as

$$q(t) = 11.89\, h(t)^{2.63} \quad ; \quad 1.5 > h(t) > 0.58 \qquad (5.41)$$

where q(t) is in cumecs and h(t) is in metres. The curve relating the area x(t) in square metres to the stage h(t) was evaluated by discretising points of area and corresponding stage, to which a second order polynomial was fitted using least squares. The area-stage curve obtained was

$$x(t) = -52.6 + 442.3\, h(t) + 229\, h(t)^2 \qquad (5.42)$$

The comparative 1-step ahead forecasts of the stage h(t) were carried out using the self-tuning predictor (without a perlog input)

as follows:

(i) the discharge was predicted using discharge q(t) as the forecast variable, which was then converted to stage h(t) using the rating curve (eqn. 5.41);

(ii) the stage was predicted using stage h(t) as the forecast variable. The original discharge data were converted to stage using the rating curve (eqn. 5.41);

(iii) the cross-sectional area of flow was predicted using the area x(t) as the forecast variable. The original discharge data was converted to area data by using the rating-curve (eqn. 5.41), and the area-stage curve (eqn. 5.42). The forecast area was then converted back to the stage again using the area-stage curve.

Note that the orginal flow data had been obtained from stage records with the rating curve in the first place.

In general, the comparison indicated that the forecasted stages obtained using the three self-tuning predictors resulted in forecast error variances of similar orders of magnitude. Of particular interest here though is the ability of the models to forecast the hydrograph rising limb and peak flow. The results of maximum stage prediction errors on the rising limbs (as percentages of maximum observed stage) and the prediction errors in the estimations of the maximum stage for two peaks of a storm hydrograph are shown in table 5.4. For all three predictors the best results (based again on minimum prediction error variance) were obtained with $n = 1$, $m = 3$ and $\ell = 2$.

The results obtained are plotted in fig. 5.31. The cumulative squared prediction errors for these plotted data are also shown in table 5.4. It was found that the h(t) predictor always gave higher forecasts than the x(t) predictor, while the x(t) predictor gave higher forecasts than the q(t) predictor. From the results

	hydrograph peak A at 6 a m 5 Dec.'72			hydrograph peak B at 11 a.m. 6 Dec.'72		
FORECAST VARIABLE	x(t)	q(t)	h(t)	x(t)	q(t)	h(t)
peak error%	0.05	3.49	-2.02	-0.93	0.83	-2.26
max. rising limb error %	9.97	9.87	9.88	3.45	5.18	-3.51
$\Sigma\varepsilon(t)^2$	0.015	0.017	0.016	0.016	0.020	0.015

Table 5.4 Comparison of results of 1-step ahead stage forecasts using three self-tuning predictors with the forecast variable as the cross-section area of flow, the discharge and the stage respectively

it is difficult to ascertain which is the better predictor. Overall, the x(t) predictor appears to give the best results, while the q(t) predictor gives the worst results.

In conclusion these results do show that either stage, discharge or area of flow for the Hirnant may be used as the forecast variable when forecasting the stage. Of these three, the stage is the most convenient to use. Next, and almost as convenient to use, is the area, which only further requires cross-sectional profiles of the river bed to be measured, and the area-stage curve to be determined. The discharge however must be obtained from the stage-discharge relationship, which for many rivers is difficult to obtain and/or requires frequent re-calibration.

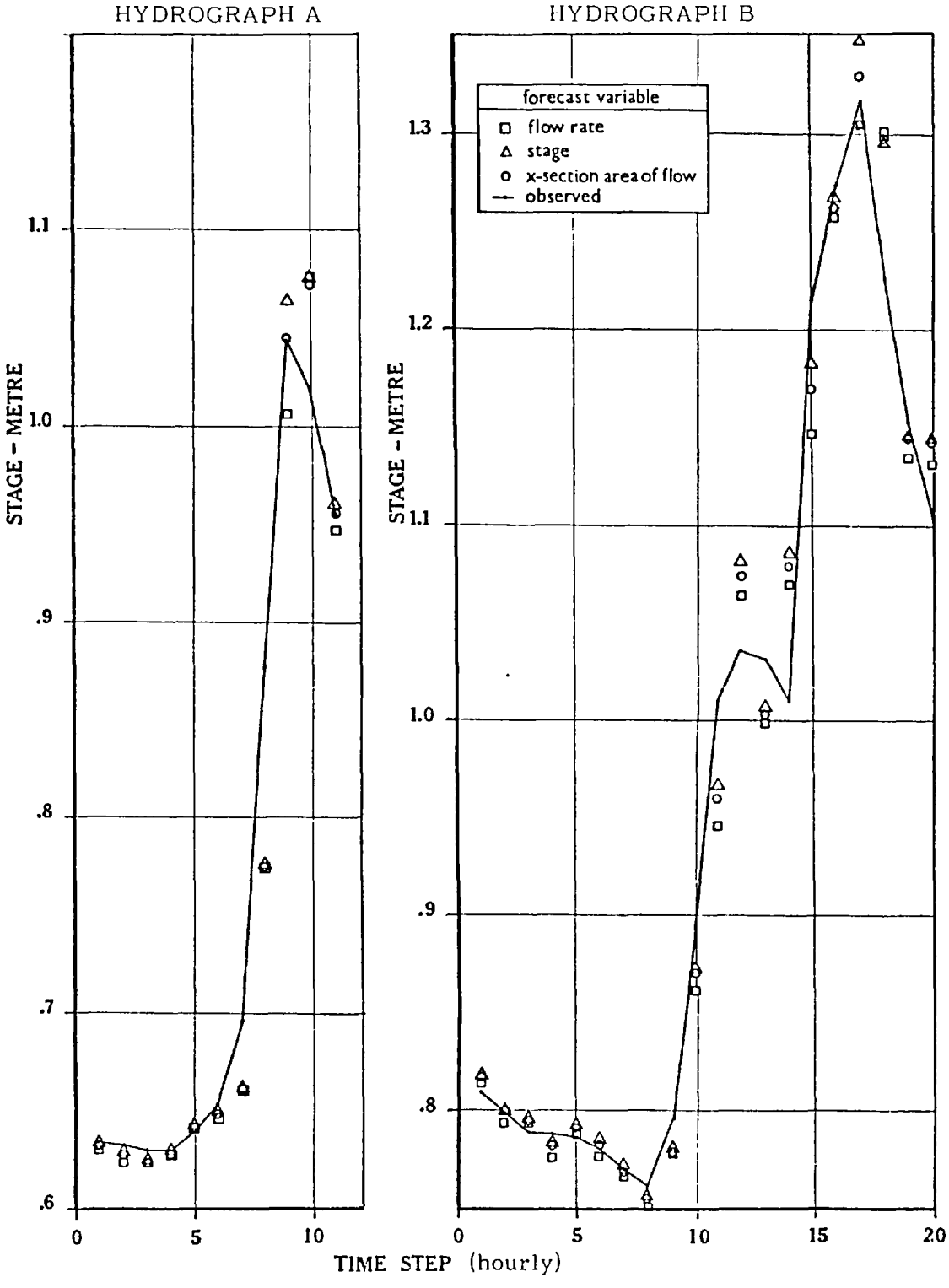


Fig. 5.31 *One-step ahead forecasts of stage for the Hirnant using this flow rate, stage and cross-sectional area of flow as forecast variables*

5.8 SUMMARY

The effective real-time control of a water resource system is dependent on the real-time forecasting of streamflow, which may also be of vital importance in the early activation of flood warning alleviation procedures.

In Section 5.1 the important distinctions between simulation models and real-time models based on recursive estimation techniques have been outlined. The major drawback with most simulation models applied to real-time flow forecasting is that they do not have a correction facility utilising the latest observed streamflow data. Consequently, cumulative forecasting errors are allowed to build up. On the other hand, most recursive real-time models possess an autoregressive component which prevents cumulative errors building up, and also the parameters are estimated on-line (automatically).

The literature survey given in Section 5.2 showed that the number of applications of estimation theory to rainfall-runoff modelling have been limited, and restricted to within the field of research hydrology. The applications so far have been mainly based on Kalman filtering. In this study the self-tuning predictor, a special case of the self-tuning regulator, provided an alternative approach. The algorithm, which is described in Section 5.3, consists of a parameter estimator and the predictor itself. The convergence of the parameter estimates results, in the limit, in the minimum mean square error predictor. This is known as the self-tuning property. The self-tuning predictor requires minimal computer central processing memory.

Analagous to the self-tuning regulator, the self-tuning predictor which has sufficient parameters converges to the optimal predictor when the covariances given in eqn. 5.21 can be assumed as zero. An important additional feature of the self-tuning algorithm (eqns. 5.24-5.28) is that exponential weighting of the past data may be included. This may allow time-varying systems to be modelled and/or to ensure rapid convergence of the parameters.

Flow forecasts (using historic data) were obtained for the Brosna and Hirnant catchments. Some of the main results and conclusions are listed below.

(i) It was found that the Box and Jenkins model structure estimation technique was unreliable and difficult to interpret. Consequently residual diagnostic checking methods were used. The delay parameter b, however, could be estimated from the approximate impulse response (Box and Jenkins technique). It was also found that the Q and S-tests were not good indicators of model adequacy, though the tests were useful in comparing the performance of alternative models. Model selection was finally based on a minimum square prediction error criterion.

(ii) The parameters of the predictor converged very rapidly (at a rate depending on the exponential weighting).

(iii) Removal of the mean output level for the Brosna data resulted in better parameter estimation, which indicated that the mean level was masking some of the information content in the data from the least-squares parameter estimator.

(iv) The values of the parameters, which are estimated recursively, tend to vary in unison. It is suggested that this is due to a finite parameter space for which self-tuning will occur.

(v) The best predictor is obtained with a predictor step of $k = 1$ or $k = 2$. Thus, a better alternative for increasing the lead time for which forecasts are required, is to increase Δt rather than to increase k beyond 2. Results of forecasts for $k > 2$ were very poor and are not presented in this thesis.

(vi) The performance of the predictor is dominated by the AR component. To some extent this reflects the major deficiency of the ARMAX model, i.e. the model is linear and does not take into account antecedent catchment conditions.

(vii) The prediction error variance rapidly increases as the forecast lead-time is extended beyond the rainfall to runoff pure time delay.

An attempt was made to improve the performance of the predictor by including the non-linear perlog function as an auxiliary input. The perlog, described in Section 5.6, can be used as an indicator of catchment antecedent conditions. Comparative results obtained with the Brosna catchment data with and without the perlog input showed that there was a small but consistent reduction in the prediction error variance. This example also showed the multiple input capability of the ARMAX model.

In Section 5.7 forecasts of stage for the River Hirnant were obtained by using as the forecast variable, (a) the stage, (b) the discharge and (c) the cross-sectional area of flow. Forecast discharge and cross-sectional area of flow were both converted to stage using the stage-discharge curve and the area-stage curve respectively. The results obtained showed that there was little difference in which forecast variable was used. Clearly, the important point to note is that stage or cross-sectional area of flow may be all that is required to be forecast in the first place. This may be so if the objective is that of flood warning, or that the output from the forecasting model is the input to another black-box model or a controller to be discussed in Chapter 6.

6. REAL-TIME CONTROL FOR RIVER REGULATION

6.1 A SELF-TUNING CONTROLLER

The concept of reducing a complex water resource system into a number of interconnected subsystems has been introduced in Section 2.2. These subsystems can then be individually controlled in real-time using a satisficing approach. The resulting hierarchical decentralized control precludes global system optimality, but enables a reduction in the complexity of the operational management problem. The subsystems are controlled by specifying the subsystem set-points either at a short-term (real-time) or a medium-term level.

In this chapter a stochastic adaptive controller suggested by Clark and Gawthrop (1975) will be derived and applied to the control of three numerically simulated subsystems. A hypothetical subsystem is shown in fig. 6.1, where a controller is required to determine the reservoir release (system input) to achieve a flow (system output) equal to the set-point at the downstream control gauging station.

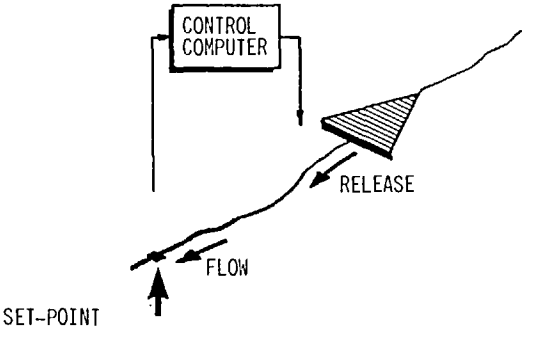


Fig. 6.1 Real-time river regulation example

It is assumed that the processes to be controlled can be represented by an nth order system model

$$A(z^{-1})y(t) = B(z^{-1})u(t-k) + C(z^{-1})e(t) \qquad (6.1)$$

where $y(t)$ is the system output, $u(t)$ the system control input,

e(t) is a Gaussian white noise sequence distributed as $N(o,\sigma^2)$, and k is a pure time delay between input and output.

The polynomials in the backward shift operator with unknown coefficients are given by

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

$$B(z^{-1}) = b_1 + b_2 z^{-1} + \ldots + b_{n+1} z^{-n}$$

$$C(z^{-1}) = 1 + c_1 z^{-1} + \ldots + c_n z^{-n} \qquad (6.2)$$

For the case where the setpoint w(t) is constant at zero, the self-tuning regulator discussed in section 4.3 can be used. For the case where w(t) is non zero and time dependent (this is the servo problem outlined in Section 4.1), Wittenmark (1973) has shown that the following version of the self-tuning regulator can be used. The system model in this case is taken to be

$$A(z^{-1})y_r(t) = B(z^{-1})\nabla u(t-k) + C(z^{-1})e(t) \qquad (6.3)$$

where $y_r(t)$ is the output error to be minimised

$$y_r(t) = y(t) - w(t) \qquad (6.4)$$

and ∇ is the backward difference operator such that

$$\nabla u(t) = u(t) - u(t-1) \qquad (6.5)$$

Thus the system model control variable is selected as the increment of the actual control variable. With $G(z^{-1})$ and $F(z^{-1})$ as defined in equations 4.5, the control law is given by

$$\nabla u(t) = - \frac{G(z^{-1})}{B(z^{-1})F(z^{-1})} y_r(t) \qquad (6.6)$$

A self-tuning algorithm is obtained if the parameters $a_1, \ldots a_m$ and $b_1, \ldots b_\ell$ in the system model

$$y_r(t) + a_1 y_r(t-k) + \ldots + a_m y_r(t-k-m+1) = b_1 \nabla u(t-k)$$

$$+ b_2 \nabla u(t-k-1) + \ldots + b_\ell \nabla u(t-k-\ell+1) + \varepsilon(t) \quad (6.7)$$

are estimated using the method of least squares (by minimising the loss function V given by equation 4.16), while the control law is given by

$$\nabla u(t) = -\frac{1}{b_1}[a_1 y_r(t) + \ldots + a_m y_r(t-m+1) - b_2 u(t-1) - \ldots$$

$$\ldots - b_\ell u(t-\ell+1)] \quad (6.8)$$

Comparing equations 6.7 and 6.8 with equations 4.14 and 4.15, the above algorithm is identical to the basic self-tuning regulator algorithm if y(t) is replaced by $y_r(t)$ and u(t) by $\nabla u(t)$. Because the set-point is not explicitly included in the above controller the model parameters may converge to different values depending on the pattern of set-point changes. If large set-point changes occur there is also the danger that excessive control action (fluctuation in the control input) may result.

The self-tuning controller algorithm

The self-tuning regulator is based on minimising the simple quadratic loss function

$$V_1 = E\{y(t)^2\} \quad (6.9)$$

Clarke and Gawthrop (1975) have shown that self-tuning algorithms can also be derived for the following two important loss functions,

$$V_2 = E\{(y(t+k) - w(t+k))^2 + \lambda^* u(t)^2\} \quad (6.10)$$

$$V_3 = E\{(y(t+k) - w(t+k))^2 + \lambda^* (u(t) - u(t-1))^2\} \quad (6.11)$$

Both these latter loss functions incorporate a set-point term and a weight λ^* which attaches a penalty to the control action. The

effect of using V_2 is to compromise between increased deviations of the system output from the set-point and a reduction in the variance of the control input. At this point it is important to consider the nature of the input and output of the hypothetical control problem in fig. 6.1. The output streamflow is a highly correlated variable with a non-zero mean. The input to the river, i.e. the controlled reservoir release is also a variable with a non-zero mean. Optimality in the sense that V_2 is minimised will clearly result in poor set-point following. The loss function V_3 on the other hand ensures set-point following, though as the rate of change in control input is penalised there is a degradation in controller performance if the set-point is rapidly varying. The self-tuning controller for the case where the loss function is given by V_3 is now derived. This derivation closely follows that of Clark and Gawthrop (1975) for the self-tuning controller with loss function V_2. In order to minimise the loss function V_3, $y(t+k)$, unknown at time t, must be replaced by the prediction $\hat{y}(t+k|t)$. If $\varepsilon(t+k)$ is the prediction error

$$y(t+k) = \hat{y}(t+k|t) + \varepsilon(t+k) \tag{6.12}$$

the function obtained from equation 6.11 is

$$V_3 = E\{((\hat{y}(t+k|t) + \varepsilon(t+k) - w(t+k))^2 + \lambda*(u(t) - u(t-1))^2\} \tag{6.13}$$

It can be shown that the minimum mean square error predictor (see Appendix A) for the system equation 6.1 is given by

$$\hat{y}(t+k|t) = \frac{B(z^{-1})F(z^{-1})}{C(z^{-1})}u(t) + \frac{G(z^{-1})}{C(z^{-1})}y(t) \tag{6.14}$$

with prediction error

$$\varepsilon(t+k) = F(z^{-1})e(t+k) \tag{6.15}$$

where $e(t)$ is a Gaussian white noise sequence distributed as $N(o,\sigma^2)$ and $F(z^{-1})$ and $G(z^{-1})$ are polynomials already defined. Clearly

$\varepsilon(t+k)$ is uncorrelated with $y(t), y(t-1), \ldots$ and $u(t), u(t-1), \ldots$ and hence it is also uncorrelated with $\hat{y}(t+k|t)$ so that equation 6.13 can be written as

$$V_3 = (\hat{y}(t+k|t) - w(t+k))^2 + \lambda^*(u(t) - u(t-1))^2 + E\{\varepsilon(t+k)^2\} \tag{6.16}$$

By setting the derivative of V_3 with respect to $u(t)$ to zero, the minimum loss

$$[\hat{y}(t+k|t) - w(t+k)] \frac{\partial \hat{y}(t+k|t)}{\partial u(t)} + \lambda^*(u(t) - u(t-1)) = 0 \tag{6.17}$$

is obtained. Noting that f_o and c_o by definition equal 1, so from equation 6.14

$$\frac{\partial \hat{y}(t+k|t)}{\partial u(t)} = b_1 \tag{6.18}$$

and the minimum loss control law, if λ^*/b_1 is replaced by λ, becomes

$$\hat{y}(t+k|t) - w(t+k) + \lambda\{u(t) - u(t-1)\} = 0 \tag{6.19}$$

A generalised output function is now defined:

$$\psi(t+k) \triangleq y(t+k) - w(t+k) + \lambda\{u(t) - u(t-1)\} \tag{6.20}$$

The only unknown at time t on the right hand side of equation 6.20 is $y(t+k)$, so the prediction of $\psi(t+k)$, i.e. $\hat{\psi}(t+k|t)$ is obtained by replacing $y(t+k)$ by $\hat{y}(t+k|t)$ so that

$$\hat{\psi}(t+k|t) = \hat{y}(t+k|t) - w(t+k) + \lambda\{u(t) - u(t-1)\} \tag{6.21}$$

The output function prediction error will be equal to the system output prediction error given by

$$\psi(t+k) = \hat{\psi}(t+k|t) + \varepsilon(t+k) \tag{6.22}$$

as

$$y(t+k) = \hat{y}(t+k|t) + \varepsilon(t+k) \tag{6.23}$$

The control objective, the minimisation of V_3 is thus the same as setting the prediction of the output function to zero at each time-step so that the control law (equation 6.19) is satisfied. In the self-tuning scheme, the optimal predictor (equation 6.14) is substituted into equation 6.21 to obtain

$$C(z^{-1})\hat{\psi}(t+k|t) = E(z^{-1})u(t) + G(z^{-1})y(t) - C(z^{-1})w(t+k) \qquad (6.24)$$

The polynomial $E(z^{-1})$ is given by

$$E(z^{-1}) = B(z^{-1})F(z^{-1}) + \lambda\, C(z^{-1})\{1-z^{-1}\} \qquad (6.25)$$

The control law which sets $\hat{\psi}(t+k|t)$ to zero is, from equation 6.24,

$$u(t) = -\frac{1}{E(z^{-1})}\{G(z^{-1})y(t) - C(z^{-1})w(t+k)\} \qquad (6.26)$$

This feedback control law is similar to the minimum variance law of the self-tuning regulator in that it minimises $E\{\psi(t)^2\}$ (instead of $E\{y(t)^2\}$) by setting the prediction of $\psi(t)$ to zero at each time-step. The self-tuning controller however also takes into account a variable set-point which is incorporated directly in the cost function.

In order to implement the feedback control law the coefficients of the polynomials $E(z^{-1})$, $G(z^{-1})$ and $C(z^{-1})$ can be estimated by recursive least squares, using the system model (eqn. 6.1) and the cost function V_3 given by equation 6.13, which have been combined to form the equivalent system model, equation 6.24. To simplify the notation the following data and parameter vectors are introduced.

$$\phi(t) = [y(t), \ldots y(t-m+1), w(t+k) \ldots w(t+k-n+1), u(t) \ldots u(t-\ell+1)]$$

$$\theta^T = [g_o, \ldots g_{m-1}, c_o, \ldots c_{n-1}, -e_o, \ldots -e_{\ell-1}] \qquad (6.27)$$

The parameters ℓ, n and m are simply the number of terms in the

respective polynomials $E(z^{-1})$, $G(z^{-1})$ and $C(z^{-1})$. Equation 6.24 is then written as

$$C(z^{-1})\hat{\psi}(t+k|t) = \phi(t)\theta \tag{6.28}$$

Finally equation 6.28 is combined with equation 6.22 to obtain the following linear regression model

$$\psi(t+k) = \phi(t)\theta - \{C(z^{-1}) - 1\}\hat{\psi}(t+k|t) + \varepsilon(t+k) \tag{6.29}$$

Recursive least-squares is applied to equation 6.29 to estimate the parameter vector θ. The estimation of θ is denoted by $\hat{\theta}$. A least-squares estimator would give an inconsistent estimate of θ when the disturbance term $\{C(z^{-1}) - 1\}\hat{\psi}(t+k|t) + \varepsilon(t+k)$ is correlated with $\phi(t)$. However, if the parameter estimator is cascaded with the control law equation 6.26, which is equivalent to a control law $\phi(t)\hat{\theta} = 0$, and if $\hat{\theta}$ becomes equal to θ, then $\hat{\psi}(t+k|t)$ is set to zero at each step and the disturbance term will be given by $\varepsilon(t+k)$. As $\varepsilon(t+k)$ is generally not correlated with $\phi(t)$ (see eqn. 6.15), $\hat{\theta} = \theta$ is a fixed point of the algorithm.

Clarke, Cope and Gawthrop (1975) have shown that if this fixed point is stable, the algorithm is self-tuning. The stability of fixed points of such algorithms are discussed by Ljung and Wittenmark (1974). The estimation procedure based on the recursive least-squares algorithm (see Appendix D) is outlined below.

At each time-step t the generalised output function (eqn. 6.20) is computed, i.e.

$$\text{(STEP 1)} \quad \psi(t) = y(t) - w(t) + \lambda\{u(t-k) - u(t-k-1)\} \tag{6.30}$$

If $\hat{\psi}(t|t-k)$ is assumed to be zero, from equation 6.29 the prediction error is computed by

$$\text{(STEP 2)} \quad \varepsilon(t) = \psi(t) - \phi(t-k)\hat{\theta}(t-1) + w(t) \tag{6.31}$$

noting that the parameters are time-invariant, i.e. $\theta(t) = \theta(t-1)$.

Even so $\hat{\theta}(t)$ may vary with time if the system changes because the parameters are estimated in the light of new information. The parameter vector θ is estimated using the following equations

$$\text{(STEP 3)} \quad \alpha(t) = \alpha(t-1)\alpha_o + (1-\alpha_o) \tag{6.32}$$

$$\text{(STEP 4)} \quad P(t) = \frac{P(t-1)}{\alpha(t)} - \frac{P(t-1)\phi^T(t-1)\phi(t-1)P(t-1)}{\alpha^2(t) + \alpha(t)\phi(t-1)P(t-1)\phi^T(t-1)} \tag{6.33}$$

$$\text{(STEP 5)} \quad \hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\phi^T(t-1)\varepsilon(t) \tag{6.34}$$

The above equations 6.32 to 6.34 are the recursive least-squares parameter estimator with exponential weighting of past data. This algorithm is also used for the self-tuning predictor (see Section 5.3) and will not be discussed further here.

Finally the control variable is obtained using the control law

$$\text{(STEP 6)} \quad u(t) = -\frac{1}{e_o}\left[\sum_{i=0}^{m-1} g_i y(t-i) + \sum_{i=0}^{n-1} c_i w(t+k-i) + \sum_{i=1}^{\ell-1} e_i u(t-i)\right] \tag{6.35}$$

The only parameters that have to be estimated a priori are ℓ, m, n, λ and k. The delay k should approximate the time delay (transport lag) between the input and output, while ℓ, m, n and λ are best estimated by using a subjective judgement approach based on the mean and variance of $\psi(t)$ and correlation analysis of the $\psi(t)$ and u(t) series. Typical parameter values are indicated by the simulation examples in the following sections.

If the handling of other system inputs is required, the only modification necessary is the extension of the ϕ and $\hat{\theta}$ vectors to include the additional variables and the inclusion of the corresponding additional terms in equation 6.35.

In most cases the control variable will have an upper and lower

bound, u_{max} and u_{min}, corresponding to the maximum and minimum desirable, or physical constraints on the control respectively. If u(t) computed in STEP 6 falls outside these bounds the following adjustments are made

$$\begin{aligned} &\text{set } u(t) = u_{max} \quad \text{for } u(t) > u_{max} \\ &\text{set } u(t) = u_{min} \quad \text{for } u(t) < u_{min} \end{aligned} \tag{6.36}$$

This will not upset the parameter estimation (STEPS 4 and 5) though if u_{min} and u_{max} set constraints such that the set-point cannot be met by control action and this condition persists over several time steps, then STEPS 1-5 can be bypassed in subsequent iterations.

The initial conditions for the algorithm, as in the case of the self-tuning predictor, can be taken as

$$\begin{aligned} P(0) &= I \times 100 \\ \hat{\theta}(0) &= \phi^T(0) = 0 \\ \alpha(0) &= 0.95\text{-}0.98 \\ \alpha_o &= 0.99 - 0.999 \end{aligned} \tag{6.37}$$

As c_o is known

$$\begin{aligned} \hat{\theta}(0)_{m+1} &= c_o = -1 \\ P(0)_{m+1,m+1} &= 0 \end{aligned} \tag{6.38}$$

The zero diagonal element of the P-matrix prevents the corresponding parameter from changing. The subscript m+1 refers to the element of the matrix P(t) corresponding to the element c_o in $\hat{\theta}(t)$.

When starting the self-tuning controller with the above initial conditions, the limits u_{min} and u_{max} play the vital role of constraining the control action before reasonable parameter estimates

have been obtained. So at each sampling interval (time t), STEPS 1-6 are evaluated to obtain the control variable u(t), and only P(t), $\phi(t)$, $\hat{\theta}(t)$ and $\alpha(t)$ are stored for the next sampling interval, a task which could easily be handled by a mini or micro computer (Clarke, Cope and Gawthrop, 1975).

Stability and convergence properties

The least square estimator in cascade with the control law described above has the self-tuning property (see Section 4.3). When $\hat{\psi}(t|t-k)$ is zero $\psi(t|t-k)$ (the output function) is equal to the prediction error $\varepsilon(t)$ as given by equation 6.22. Using the properties derived for the self-tuning regulator, the least squares parameter estimates will give the minimum variance controller when

$$E\{\psi(t)\psi(t+\tau)\} = 0 \quad ; \quad \tau = k, k+1, \ldots$$

$$E\{\psi(t)u(t+\tau)\} = 0 \quad ; \quad \tau = k, k+1, \ldots \tag{6.39}$$

These properties may be used as a statistical test to determine whether the controller has converged to the minimum variance controller (this is the self-tuning property).

While the effect of the weight λ is to attach a penalty on the control action, Clark and Gawthrop (1975) have shown that for some systems, the stability of the controller is dependent on λ. When the output function prediction is set to zero the control law is such that (from equations 6.15, 6.20 and 6.22)

$$y(t) - w(t) + \lambda\{u(t-k) - u(t-k-1)\} = F(z^{-1})e(t) \tag{6.40}$$

The closed loop system model can be obtained directly in terms of the system model parameters by substituting the system equation 6.1 into equation 6.40 giving

$$y(t) - w(t) + \lambda(1-z^{-1})\left[\frac{A(z^{-1})}{B(z^{-1})}\, y(t) - \frac{C(z^{-1})}{B(z^{-1})}\, e(t)\right] = F(z^{-1})e(t) \tag{6.41}$$

The system output is then given by

$$y(t) = \frac{1}{B(z^{-1})+\lambda(1-z^{-1})A(z^{-1})}\{[\lambda(1-z^{-1})C(z^{-1})+B(z^{-1})F(z^{-1})]e(t) + B(z^{-1})w(t)\} \quad (6.42)$$

The stability of the solution of this equation is determined by the roots of the equation

$$B(z^{-1}) + \lambda(1-z^{-1})A(z^{-1}) = 0 \quad (6.43)$$

When λ is small the closed loop behaviour is determined by $B(z^{-1})$, and when λ is large is determined by $(1-z^{-1})A(z^{-1})$.

System models where the roots of the polynomial $B(z^{-1})$ lie outside the unit circle are known as nonminimum phase systems (Åström, 1970). The characteristic of a nonminimum phase system is that it initially gives a negative output response to a positive step input. A minimum variance regulator is obtained if $\lambda = 0$ and it is well known that these regulators are extremely sensitive to variations in the model parameters (Åström, 1970) of nonminimum phase systems. If however λ is large the behaviour of the controller is determined by $(1-z^{-1})A(z^{-1})$. Thus the self-tuning controller is applicable to systems for which the self-tuning regulator would be unstable.

6.2 REAL-TIME CONTROL BY RESERVOIR REGULATION

In this section the self-tuning controller is applied to a hypothetical low flow augmentation problem based on a satisficing approach* Often the object of low flow releases is to augment the total river flow at a downstream control station to a constant or variably prescribed level. For water supply purposes the prescribed level may depend on the rate of abstraction required to meet the demand. If downstream storage is provided, rates of abstraction can vary according to flow; however if there is no storage available, regulation losses must be minimised by correct releases from the regulating reservoir. This may be difficult to effect by trial and error, especially if the travel times are significant and if surface runoff and tributary inflows contribute to the flow. Low flow releases are also often required for navigation, water quality and amenity purposes.

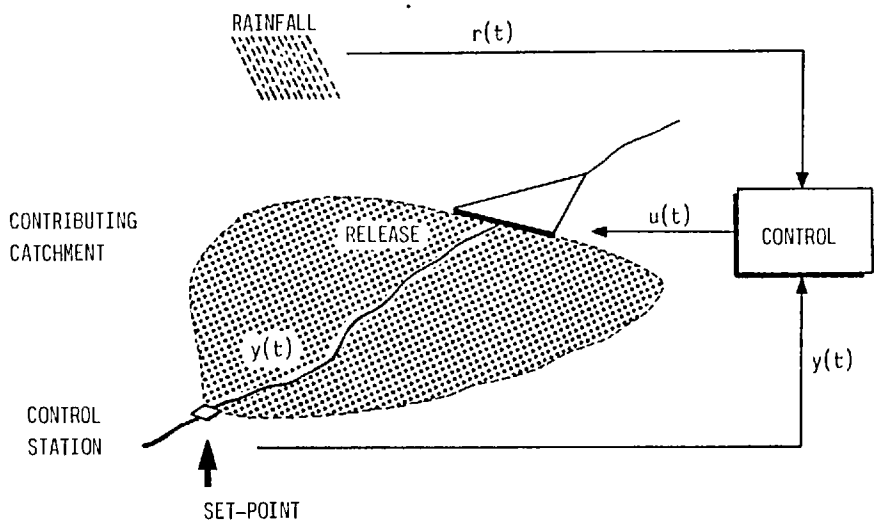


Fig. 6.2 River regulation by computer control

A hypothetical example is shown in fig. 6.2. Releases have to be made from the regulating reservoir so that the resultant flow at the control-gauge is equal to the set-point. In order to perform the release automatically the present control-gauge level, measured

* *The satisficing approach is described in Section 2.2.*

rainfall, and the set-point would be fed on-line into the control computer which, based on a forecast of the future state, computes the required release. The set-point, which can be determined at the medium-term level as described in Section 2.2, represents the subsystem operational management policy. It is assumed that the only constraint the controller must consider is the maximum and minimum release limitations. Clearly the set-point can be used to some extent for managing the reservoir retention level in the sense that for large set-point values large releases will be made and for small set-point values only small releases, if any, will be necessary. For example if a flood were imminent and maximum reservoir storage was required, the set-point would be set to the maximum flow the river could contain safely without flood damage occurring.

The self-tuning controller was applied to a simulated system (analogous to fig. 6.2) consisting of a river with an upstream input being regulated by reservoir releases for low flow augmentation, and lateral inflow due mainly to surface runoff contributing to the total river flow measured downstream at the control-gauging point. The system was simulated by two components which added together represent total river flow. A river flow model comprising two Muskingum reaches (Chow, 1959), was used to route the reservoir releases down to the control-gauge. The impulse and step responses for the model are shown in fig. 6.3. The rainfall-runoff component of the simulated system was represented by a set of historic data. Two-hourly values of rainfall and runoff from the Hirnant catchment described in section 5.5 were used. The simulated system is shown in fig. 6.4. Comparing figs. 6.2 and 6.4, r(t) the observed rainfall input is given by the Hirnant data. The release u(t) (control variable) is routed using the linear river flow model down to the control station. This routed flow is combined with the non-linear and stochastic runoff component (the Hirnant flow data) to give the control gauge reading y(t). The time scale of the elements of the simulated system are as shown below. The rainfall to runoff response time refers to the Hirnant catchment discussed in section 5.5.

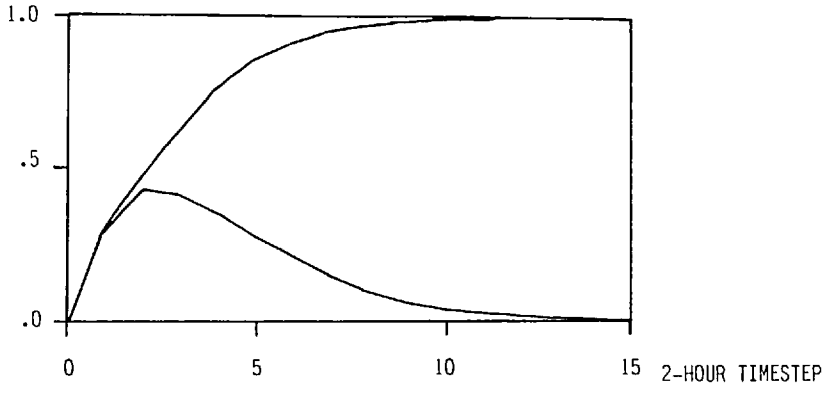


Fig. 6.3 Impulse and step response of flow routing model

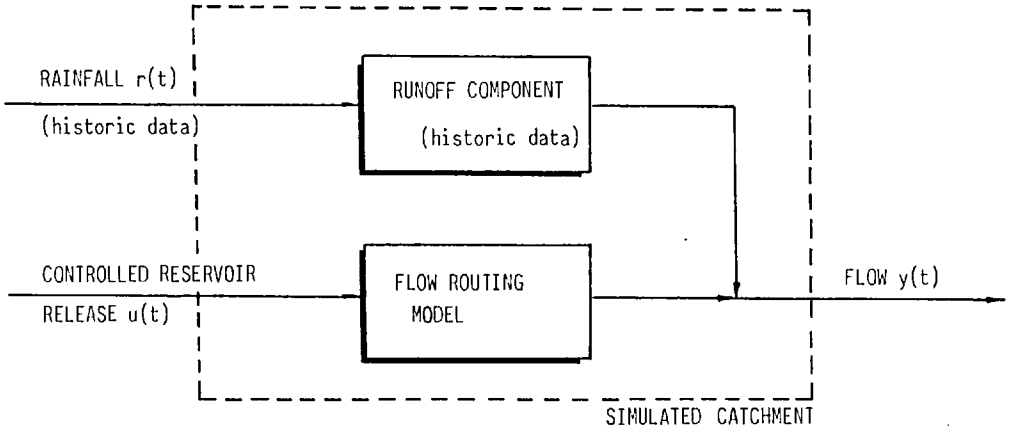


Fig. 6.4 Simulated subsystem

controller time-step/data sampling	2 hours
river flow travel time	2 hours
rainfall to runoff response time (pure time delay)	½ to 1 hour

The results of the study are shown in fig. 6.5. A sine wave set-point was chosen to provide a difficult situation for the controller to handle. The regulated flow is the total flow observed at the control gauge while the flow due to catchment runoff is a component of this flow. The objective of the controller is to minimise the mean square error between the regulated flow and the set-point. The flow demanded at the control gauging is often less than the flow due to runoff, in which case the control releases will be set to zero as shown, and the total flow becomes equal to the runoff. However as the rainfall to runoff response time is much less than the travel time of the reservoir releases down to the control-gauge, the releases prior to the flood event are often over-estimated (e.g. near time-step 40). The sensitivity of the controller to rainfall (feedforward) input is clearly illustrated at time-step 120. The best controller was achieved with

$$\ell = 3 \qquad n = 1 \qquad k = 1$$
$$m = 1 \qquad p = 2 \qquad \lambda = 1.0 \tag{6.44}$$

where p is the number of rainfall (r(t)) terms with corresponding parameters $h_o, \ldots h_{p-1}$. The vectors $\phi(t)$ and θ are given by

$$\phi(t) = [y(t), r(t), r(t-1), w(t+1), u(t), u(t-1), u(t-2)]$$
$$\theta^T = [g_o, h_o, h_1, -1, e_o, e_1, e_2] \tag{6.45}$$

The procedure adopted for determining ℓ, n, m, p, k and λ for the self-tuning controller was similar to the procedure finally adopted for the self-tuning predictor and described in section 5.3. The procedure is as follows:

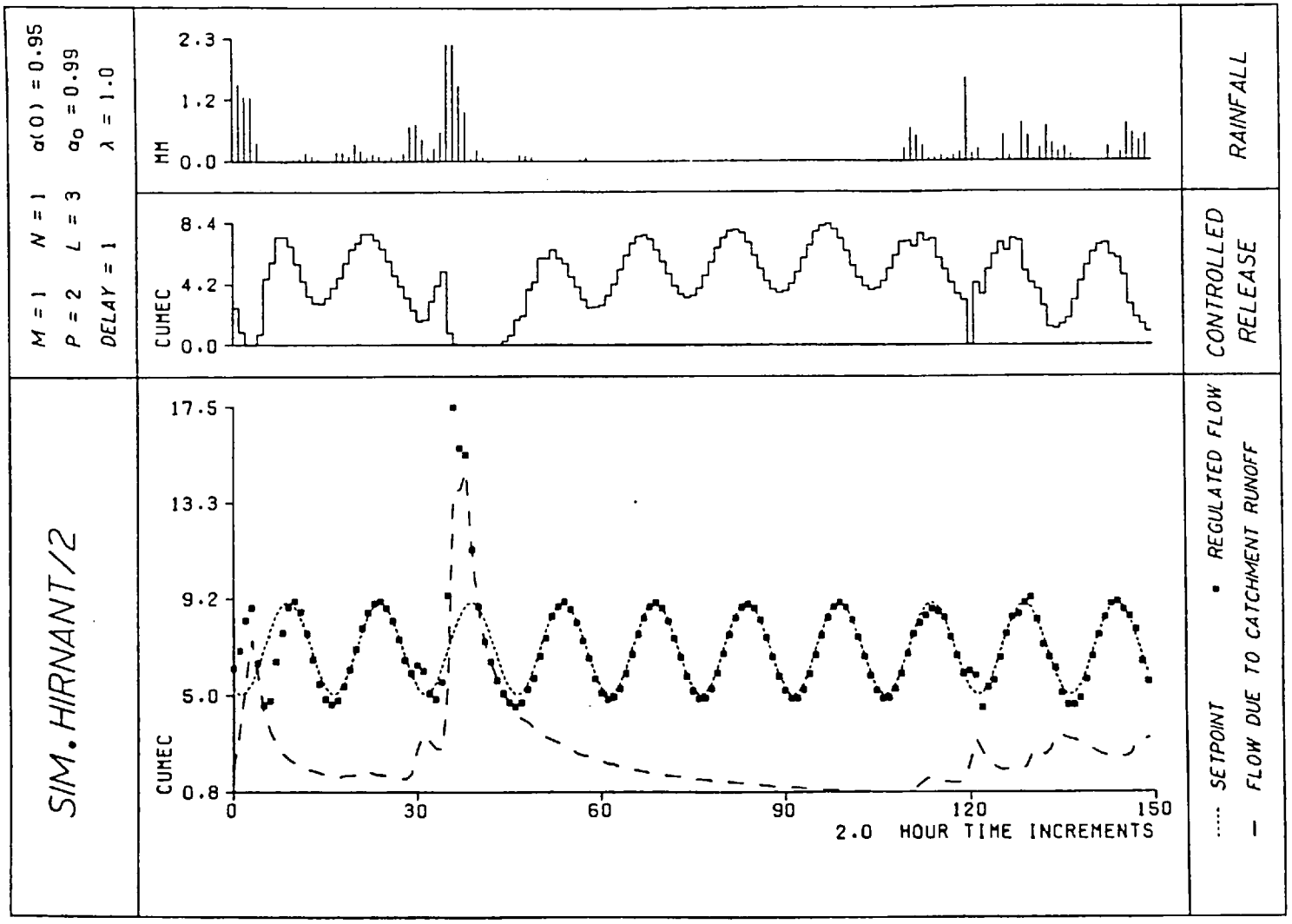


Fig. 6.5 River regulation by reservoir releases using the self-tuning controller.

(i) a number of competing self-tuning controllers are selected (with different values of ℓ, n, m, p, k and λ) and corresponding sequences of $\psi(t)$ generated;

(ii) the controller giving the smallest mean and variance of $\psi(t)$ was selected as the best controller;

(iii) in the event of two or more controllers giving similar mean and variances for $\psi(t)$, the statistical properties given by equations 6.39 were used as a test criterion to aid in the selection of the appropriate controller.

The variations in parameters estimated recursively and starting with the initial conditions given by equations 6.35 and 6.37 are illustrated in fig. 6.6. The parameters shown are correspondingly

$$\hat{\theta}(1) = g_o \qquad \hat{\theta}(5) = e_o$$

$$\hat{\theta}(2) = h_o \qquad \hat{\theta}(6) = e_1$$

$$\hat{\theta}(3) = h_1 \qquad \hat{\theta}(7) = e_2$$

Good controller performance is achieved generally after about 20 steps, even though parameter estimates seem uncertain. As for the self-tuning predictor described in Chapter 5, the parameters tend to wander in unison. This is the behaviour typical of self-tuning algorithms and is due to deviations in the parameter space to which the control law is insensitive. An initial forgetting factor of $\alpha(0) = 0.95$ with $\alpha_o = 0.99$ was used to ensure rapid parameter convergence.

The autocorrelation function $r_{\psi\psi}(\tau)$ and sample cross-correlation function $r_{\psi u}(\tau)$ are shown in fig. 6.7. The sample size used was 500 for which the 95 percent confidence interval is ± 0.088. Except for some spurious values, the sample auto and cross-correlation functions indicate a reasonable agreement with the self-tuning properties given by equations 6.3 . The computed mean and variance of the output function $\psi(t)$ were -0.09445 and 0.7299 respectively.

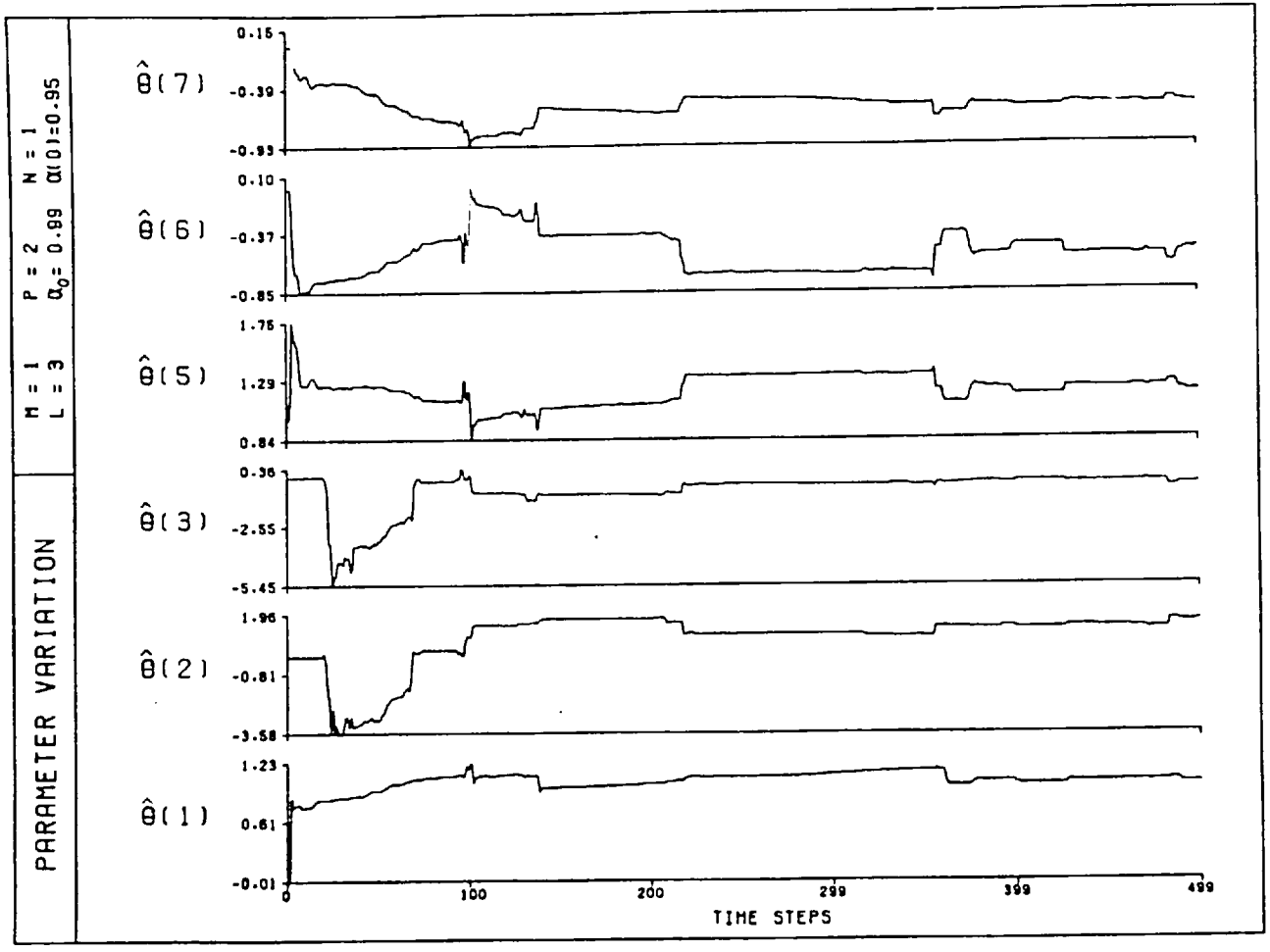


Fig. 6.6 Recursive parameter estimates for the self-tuning controller.

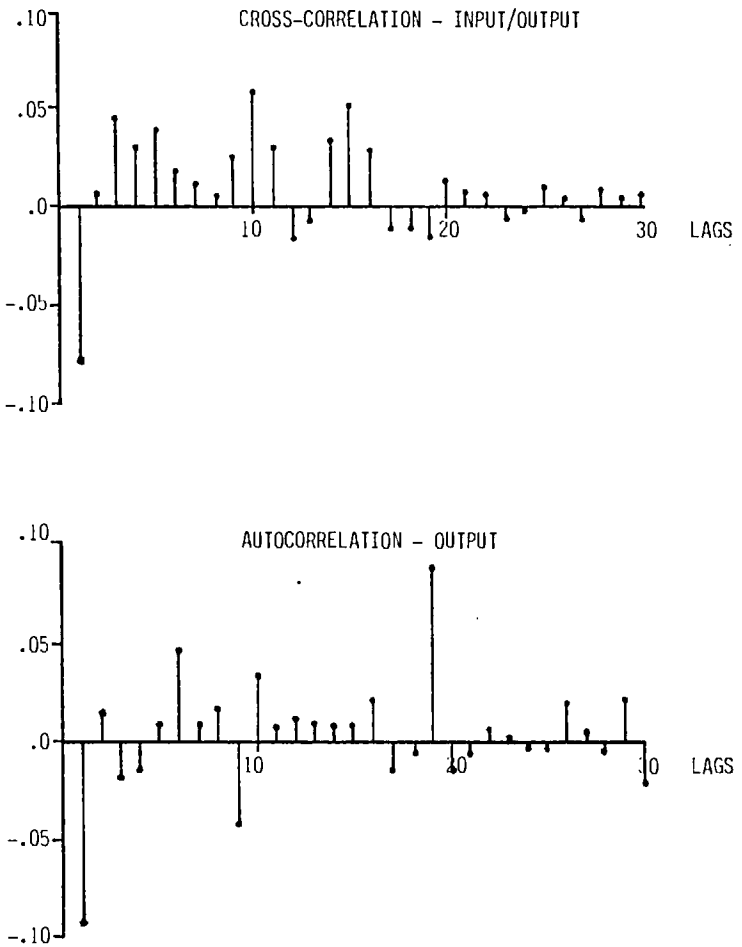


Fig. 6.7 Sample autocorrelation function $r_{\psi\psi}(\tau)$ and sample cross-correlation function $r_{\psi u}(\tau)$ for sample size N = 500. (95% confidence intervals is ±0.088).

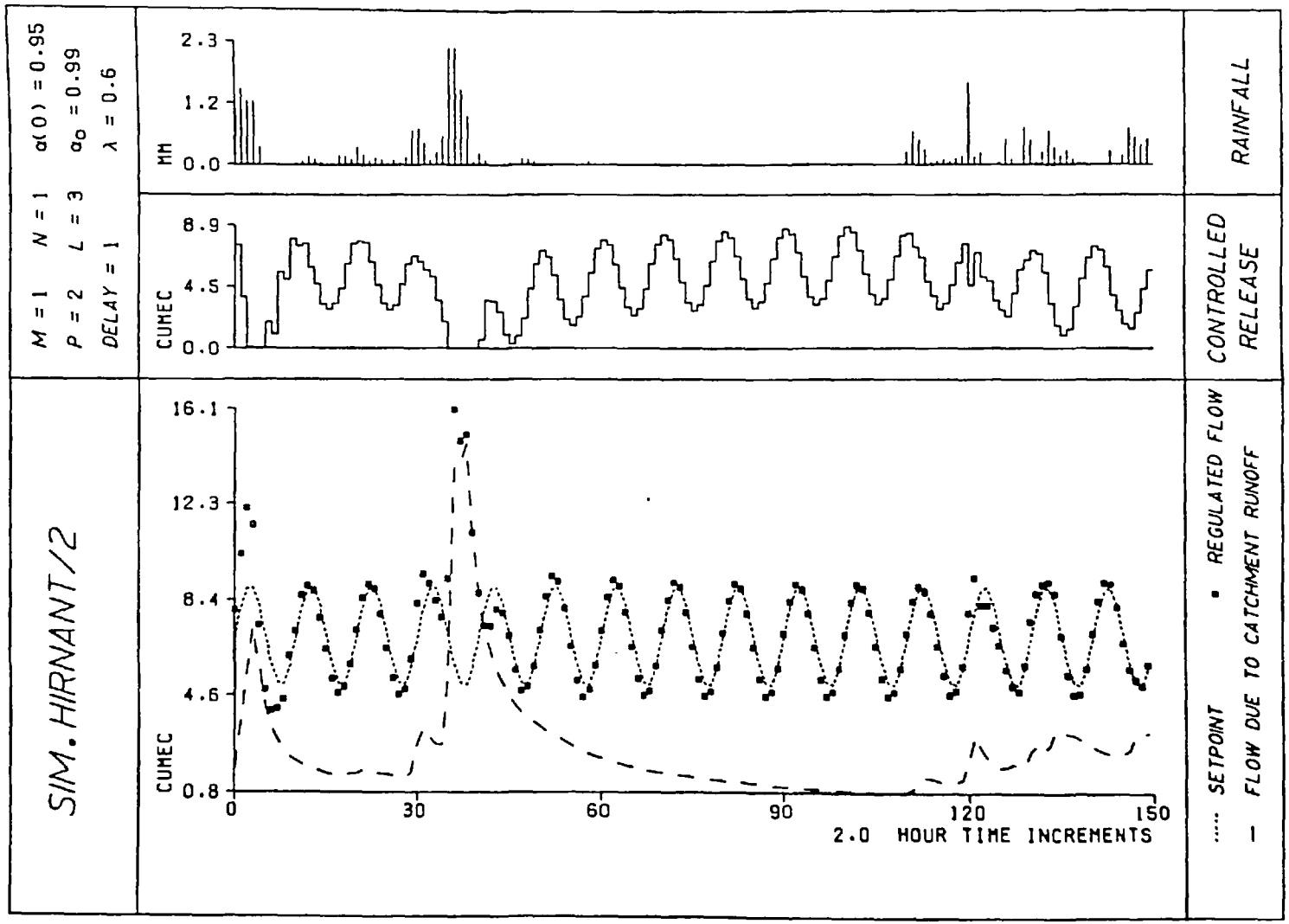


Fig. 6.8 The self-tuning controller with a small value for λ.

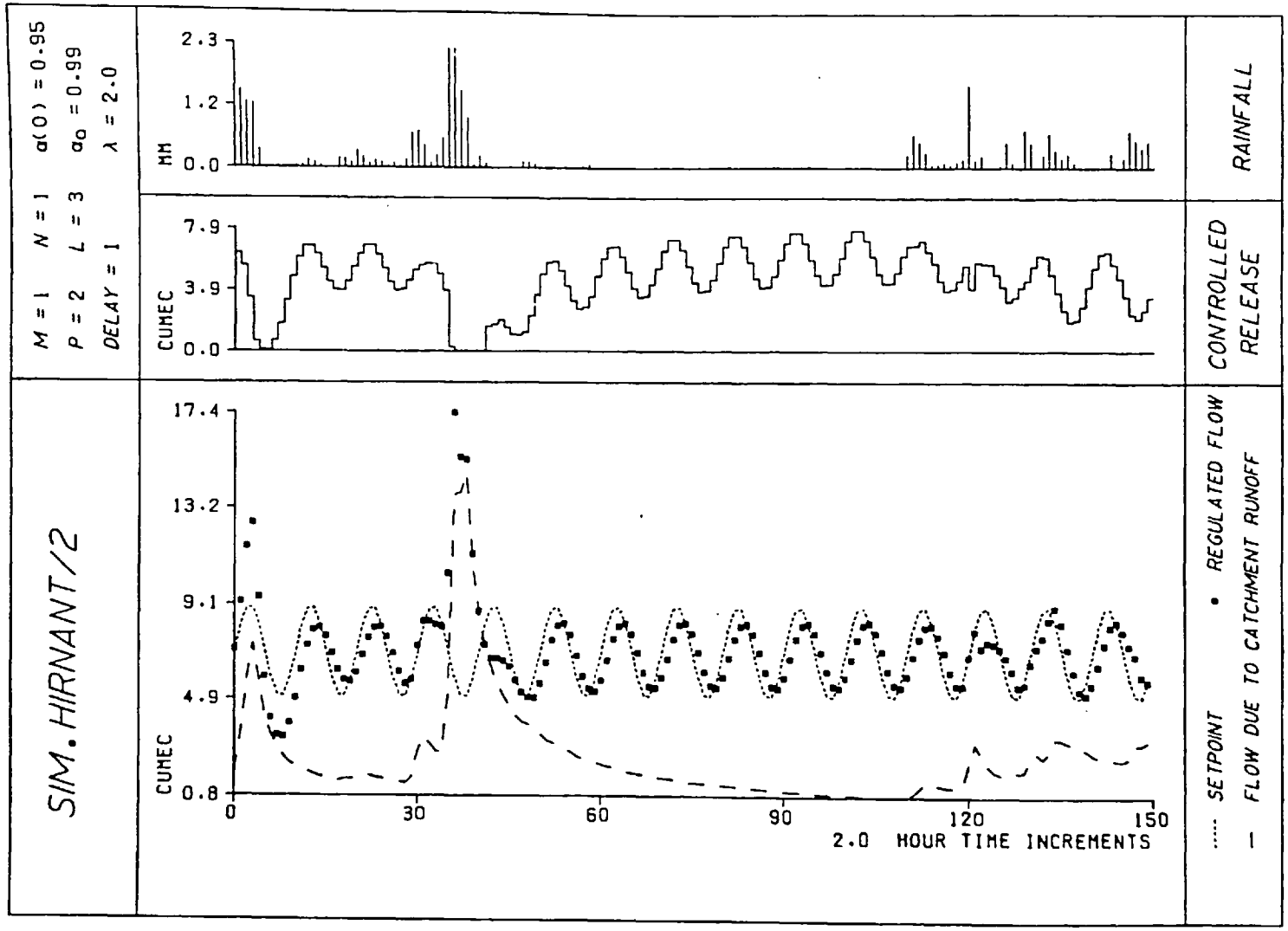


Fig. 6.9 The self-tuning controller with a large value for λ.

The effect on the control of varying λ is shown in figs. 6.8 and 6.9. In these examples a set-point with a smaller period was used. Note that at time step 35 in fig. 6.8 the controller is already expecting a decrease in the control input, and consequently the loss due to over-release is less than in fig. 6.5 where the controller is expecting an increase in the control input. When λ is too small as in fig. 6.8 the control action is excessive, as shown by the consistent over and under-fitting of the set-point peaks and troughs, while when λ is too large as in fig. 6.9 the control action is clearly over-damped. Fortunately, sensitivity to λ is not so great as to make the trial and error selection procedure described above tedious.

The above simulation was carried out with u(t), the control variable, as the actual discharge value in m^3/s. The same problem was then considered with u(t) as the spillway gate lift in meters. The following relationships define the reservoir illustrated in fig. 6.10.

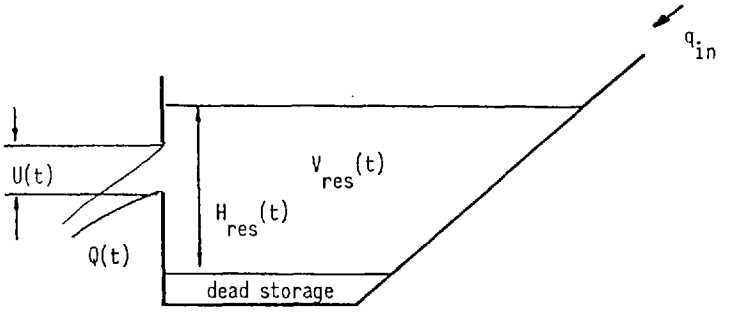


Fig. 6.10 Simulated reservoir

$$H_{res}(t) = [\sqrt{(0.1024.10^{12} + 12800 \times V_{res}(t) - 320000}]/6400 \quad (6.46)$$

$$Q(t) = 1.5H_{res}(t)\sqrt{(20u(t))} \quad (6.47)$$

$$V_{res}(t+1) = V_{res}(t) - (Q(t) - q_{in}) \times \Delta t \quad (6.48)$$

$$H_{res}(0) = 25 \text{ m}$$

$$V_{res}(0) = 10^7 \text{ m}^3$$

$$q_{in} = 1 \text{ m}^3/\text{s constant}$$

$$\Delta t = 7200 \text{ s} \qquad (6.49)$$

The discharge from the reservoir is not only a non-linear function of the gate lift,but a function of the reservoir level (assumed unknown to the controller) as well. It was assumed that the river upstream of the reservoir is regulated. Because the reservoir inflow was not included as a feedforward term in the controller the inflow was taken as constant. The set-point at the control station downstream of the reservoir was assumed to be 7 m^3/s, thereby ensuring a rapid reservoir drawdown. The system to be controlled is in this case both non-linear and time-variant. The result of the control plotted after 200 2-hour time steps is shown in fig. 6.11. A constant set-point was used so that the effect of the reservoir level on the required gate opening is clearly shown. A delay k = 2.0 was used in this case to obtain good control while with a unit delay it was found that λ had to be set to $\lambda = 10.0$ to achieve similar results. With $\lambda < 10.0$ it was found that the parameters would not converge. This shows that if either λ or k (or both) are increased, the stability of the controller increases (see Section 6.1).

To take into account changes in the system dynamics, a constant forgetting factor $\alpha(t) = 0.95$ was used. The control could further be improved by including observations of the reservoir level, and forecasts of the reservoir inflows (if they are not constant as assumed in the above example) as feedforward terms in the controller.

The significance of the above example is that

(i) the self-tuning controller can be used for non-linear and

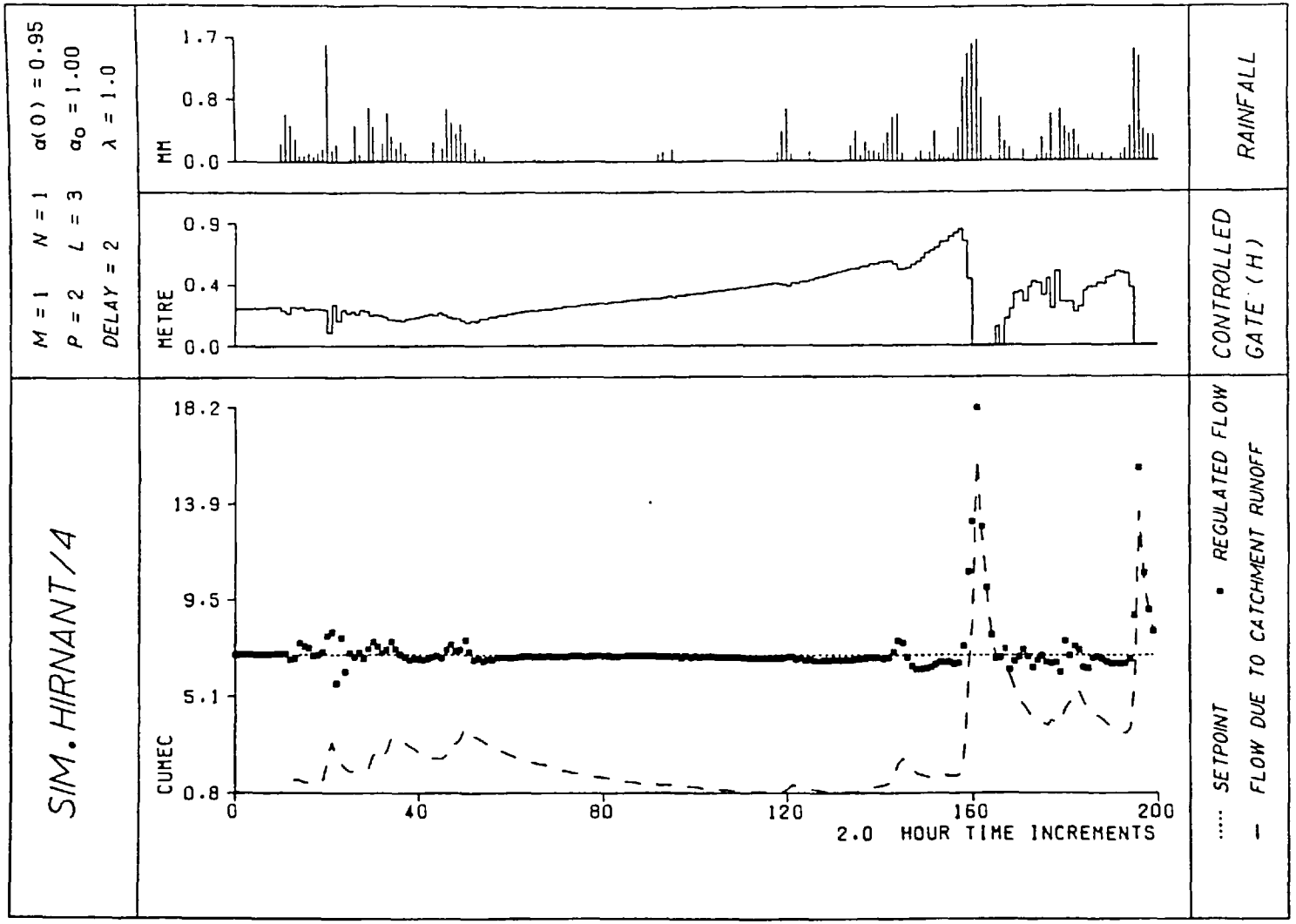


Fig. 6.11 *River regulation using the self-tuning controller. The control variable u(t) is the gate lift H(t).*

time-variant systems, even though the control law is linear and time-invariant.

(ii) the system is represented by a black-box model which allows considerable freedom in the selection of the inputs (control and feedforward) and output variables.

From the very simple examples shown in this section it is evident that the self-tuning controller may be applied to simple control problems associated with the real-time operation of a subsystem. In the next two sections the self-tuning controller is applied to hypothetical real-time control problems which are based on the simulation of real physical systems.

6.3 RIVER REGULATION BY CONTROLLED GROUNDWATER ABSTRACTION

Groundwater abstraction discharge to a river for augmentation purposes is frequently employed where sites for impounding reservoirs are not available. It is important that groundwater pumping is carefully scheduled so that pumping (electricity) costs are minimised and unnecessary depletion of the aquifer is avoided. The basis for a case study of a moderately complex control problem is provided by the Lambourn Valley chalk aquifer. This site was selected by the former Thames Conservancy for a pilot scheme to examine the possibility of using the groundwater storage in the chalk aquifer for regulation of the River Thames. The study carried out here represents hypothetical use of the Lambourn Valley aquifer for river regulation, and does not relate to the actual objectives of the Lambourn Scheme.

The purpose of the study presented herein is to investigate the feasibility of applying the self-tuning controller to the optimal selection of pumping rates from a given distribution of well sites.

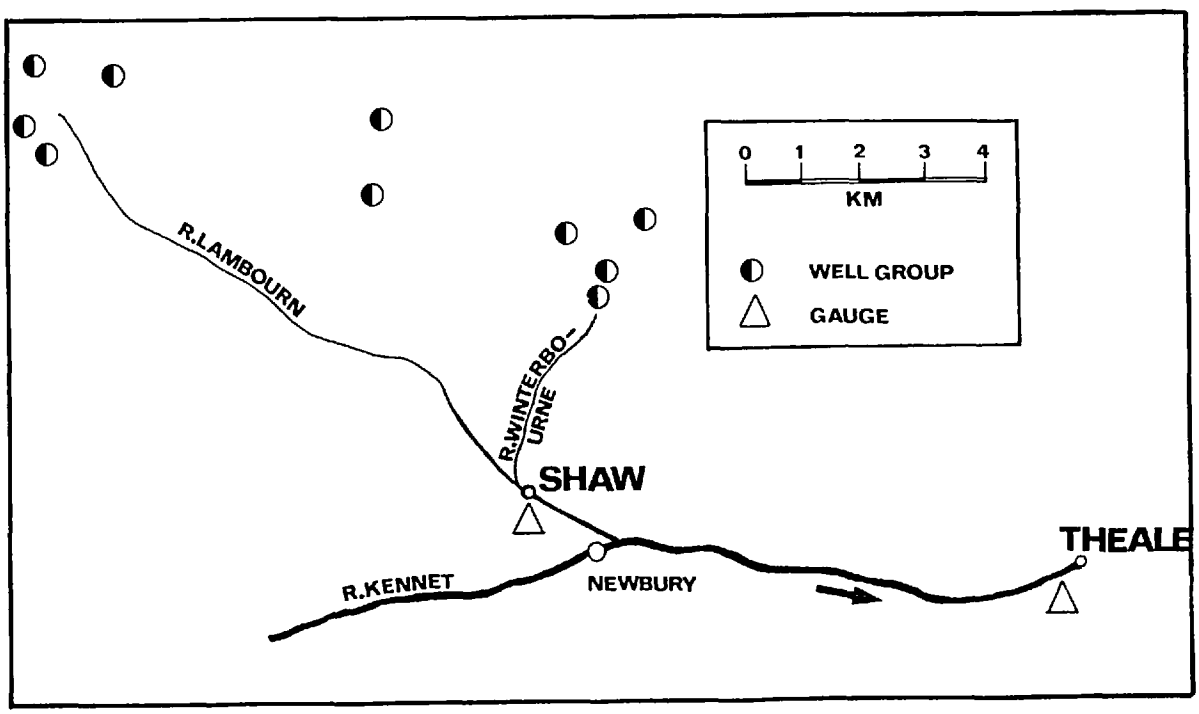


Fig. 6.12 Groundwater abstraction from the Lambourn Aquifer

Groundwater from the Lambourn aquifer is pumped directly into the River Lambourn to augment the flow of the River Kennet. The objective of the control as shown in fig. 6.12 is to select pumping rates such that a minimum statutory flow requirement is not violated. The River Kennet is a tributary of the Thames.

The Lambourn groundwater model

The Lambourn Valley groundwater system consists of Upper and Middle chalk and is located on the dip slope forming the north-west limb of the London Syncline. The aquifer is unconfined with a saturated thickness ranging from 100 m at the outfall of the catchment to about 30 m at the head. The discharge of the Lambourn and its tributary the Winterbourne is almost exclusively base flow (from the aquifer). Apart from exceptional cases flow only occurs when the river bed intersects the groundwater table, resulting in seepage through the river bed. If the water table is below the river bed, the flow in the river will seep back into the aquifer. This hydraulic connection between the river and the aquifer causes a problem when water from the aquifer is discharged into the river. If no recharge occurs, pumping will lower the water table so that eventually water discharged into the river will be lost back to the aquifer. A pilot scheme pumping test (Thames Conservancy, 1971) has shown that the net gain, defined as the ratio of increment of river flow due to pumping to the total pumping rate, could drop from almost a hundred percent to zero percent within some weeks. As pumping costs are high the location of wells and the pumping policy should be such that the net gain is maximised for a given river flow requirement.

The design of a real-time control system for the Lambourn Valley pumping scheme was carried out with a simulation model incorporating a number of simplifications as outlined below. In fact the control simulation problem is reduced to one that is similar to the reservoir regulation problem described in section 6.2. The system in this case is clearly time-variant due to the fluctuating water

table. The Lambourn groundwater system was simulated by a digital model developed by the Water Resources Board, Reading (Oakes and Pontin, 1976). Some comments on this model follow.

Flow through a porous medium can be mathematically represented by combining Darcy's Law with the equations of mass conservation. If the vertical flow component is negligible the following partial differential equation (Remson, Hornberger and Moltz, 1971) is obtained

$$\frac{\partial}{\partial x}\left(T\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial h}{\partial y}\right) = S\frac{\partial h}{\partial t} + Q \qquad (6.50)$$

where
x and y are horizontal space coordinates
t is the time
h is the mean potential of the saturated thickness and corresponds to the water table elevation
T is the transmissivity
S is the specific yield
Q is the linear flow rate from the aquifer and may include a negative infiltration component

The model which is fully described in Oakes and Pontin (1976) uses an explicit method for solving the finite difference representation of equation 6.50. The aquifer is modelled using 204 squares with a 1-km grid size as shown in fig. 6.13. This permits (for stability reasons) a time step of up to 2.1 days. Concurrently with the calculation of groundwater levels, the variation in base flow for the Lambourn and the Winterbourne are computed and incorporated in the calculation of potentials at the river grid points. The river grid points have been selected to follow as closely as possible the actual tracts of the Lambourn and the Winterbourne, while the model boundary corresponds to the natural water divide (assumed constant). It is implicitly assumed in the model that each river is in hydraulic continuity with the aquifer and fully penetrates it. The baseflow calculation at each gridpoint includes the flow contribution from

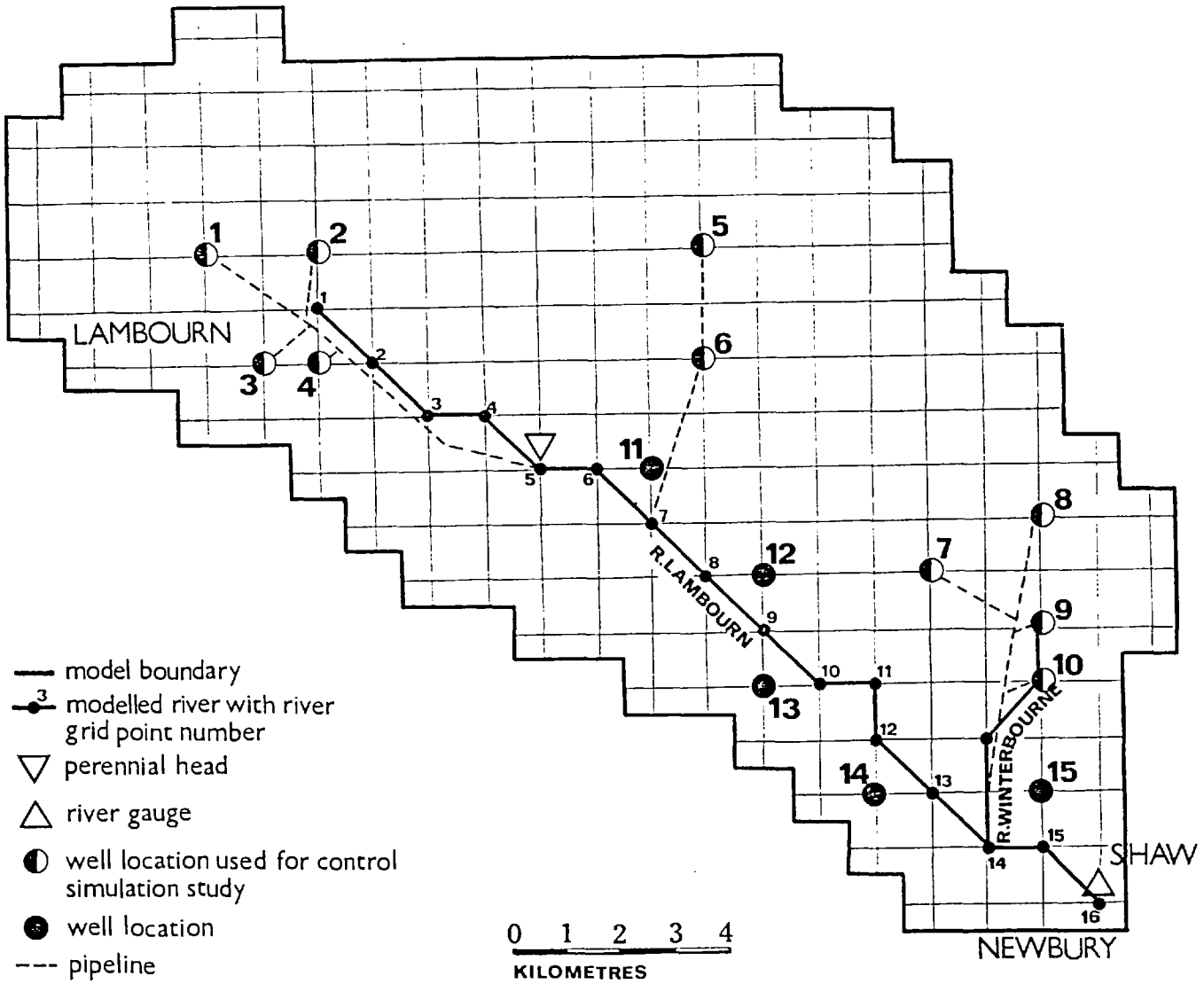


Fig. 6.13 Discretised representation of the Lambourn valley groundwater system.

the upstream gridpoint as well as water piped into the river from abstraction wells.

Simulated Control System

For this study ten well site locations (wells 1 to 10 in fig. 6.13) were chosen with the aim of restricting the drawdown along the river (to maximise the net gain) due to pumping. The well site locations which were based on the findings of the 1968-1969 pilot scheme and the consequent recommendations for groundwater abstraction (Thames Conservancy, 1971), and their respective discharge points along the River Lambourn are shown to the nearest grid point in fig. 6.13. The pumping rate from each well site was assumed to be the same at any one time, thus reducing the number of control variables down to one (i.e. the aggregate aquifer pumping rate u(t)). Furthermore it was assumed that

(i) flow forecasts for the River Kennet directly upstream of the confluence of the River Kennet and Lambourn were available and

(ii) the routed flow from Shaw to the confluence and the combined routed flow from the confluence down to Theale can be approximated by transport lags (pure time delays).

The required flow forecasts directly upstream of the Kennet-Lambourn confluence were not available in this study, but were assumed as follows. The flow component of the Kennet at Theale excluding the contribution of the Lambourn can be estimated simply as the diff-rence in the suitably lagged flows at Theale and Shaw. This flow component can then be considered as the lagged flow upstream.

$$\left(\begin{matrix}\text{lagged flow of Kennet} \\ \text{above confluence}\end{matrix}\right) = (\text{Flow at Theale} - \text{Lagged flow at Shaw})$$

For this study the flows of the Kennet as derived above were assumed to be the required forecasted flows.

The simulation was carried out using a 2 day timestep. As the lags

in this case must be taken as integer multiples of the timestep (also the sampling interval) and the river reach lengths are relatively short, the best approximation of the River Kennet flow component at Theale was in fact obtained with zero lags. In practice, if a less than 2 day timestep is being used, forecasted flows for the River Kennet above the Kennet-Lambourn confluence would be routed down to Theale (this would be the actual flow at Theale, minus the flow contribution from the Lambourn). Thus, the problem reduces to defining the control strategy for regulating the Lambourn at Shaw to achieve the desired flow at Theale.

The simulation procedure is outlined below.

(1) Flow records were available for the River Lambourn (Shaw) and the River Kennet (Theale) covering the period April 1975 to June 1976, during which time there was no pumping, and recharge was negligible. The hydrographs are shown in fig. 6.14.

(2) The uncontrollable portion of the flow of the River Kennet at Theale which is the total flow minus the baseflow contribution of the River Lambourn was obtained by subtracting the flow at Shaw from the total flow at Theale (the lag was zero).

(3) The desired compensating flow from the Lambourn can then be obtained by subtracting the 'forecast' uncontrollable flow component at Theale from the demand at Theale. This is the flow demand at Shaw which must be met by pumping. Negative flow demands are set to zero.

A schematic representation of the control system is given in fig. 6.15. The set point $w(t)$ for this control problem is the estimated demand flow at Shaw which has to be met by the aggregate pumping rate $u(t)$. The only observable quantity for this control loop is the flow at Shaw $q(t)$.

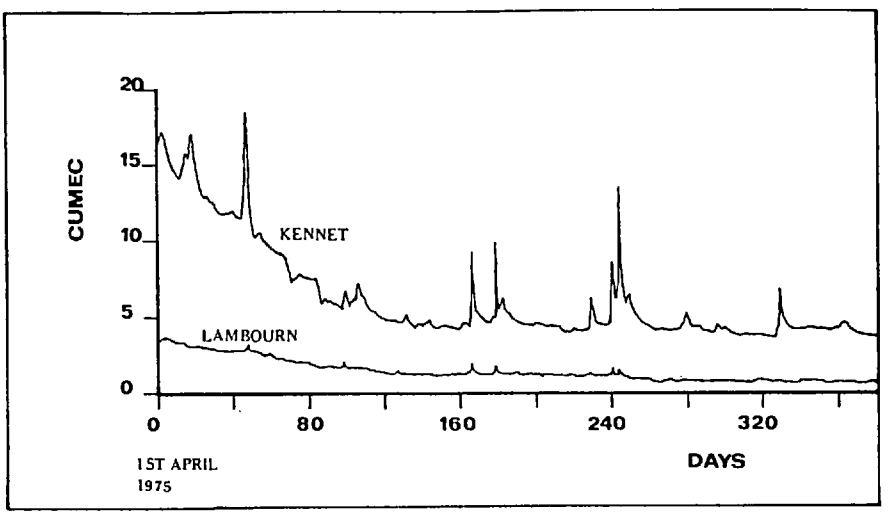


Fig. 6.14 Flow records for the River Kennet at Theale and the River Lambourn at Shaw

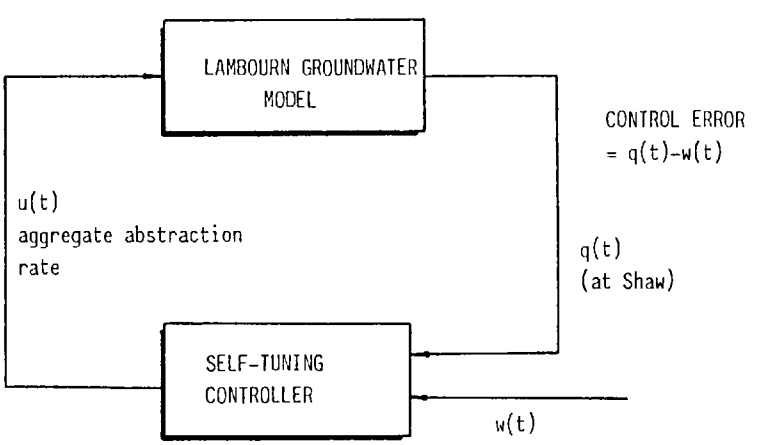
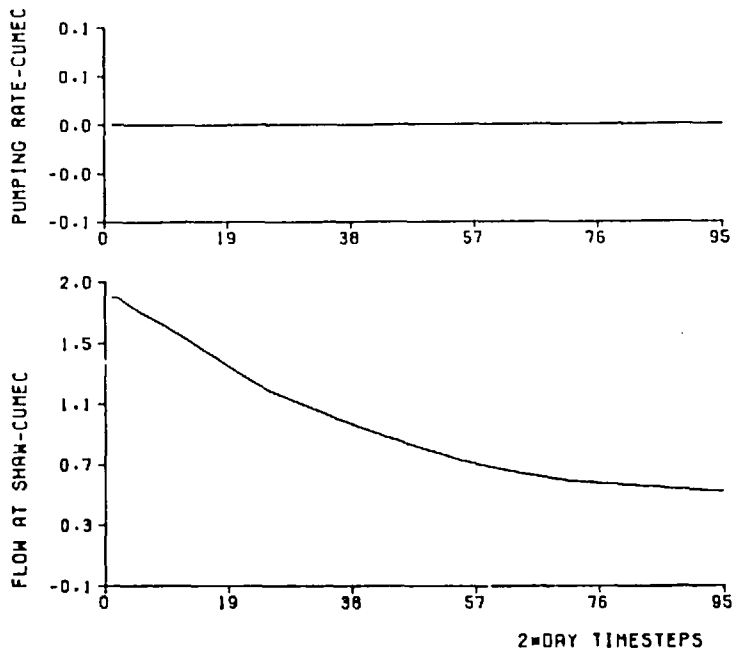


Fig. 6.15 Schematic representation of hypothetical Lambourn control system

(a)



(b)

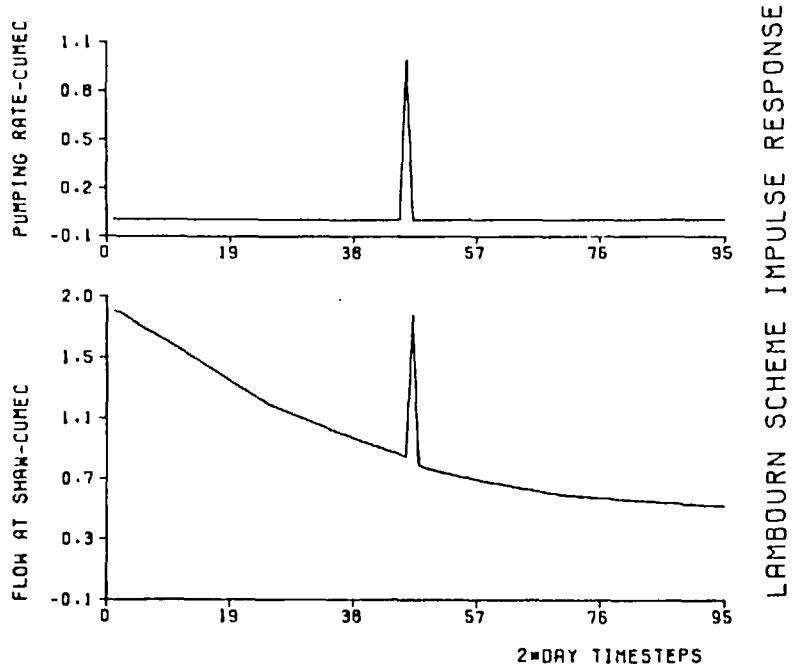


Fig. 6.16 Results from the Lambourn groundwater simulation model showing (a) a baseflow recession (no recharge) and (b) the pulse response (response obtained from pumping 1 m^3/s over 1 2-day timestep.

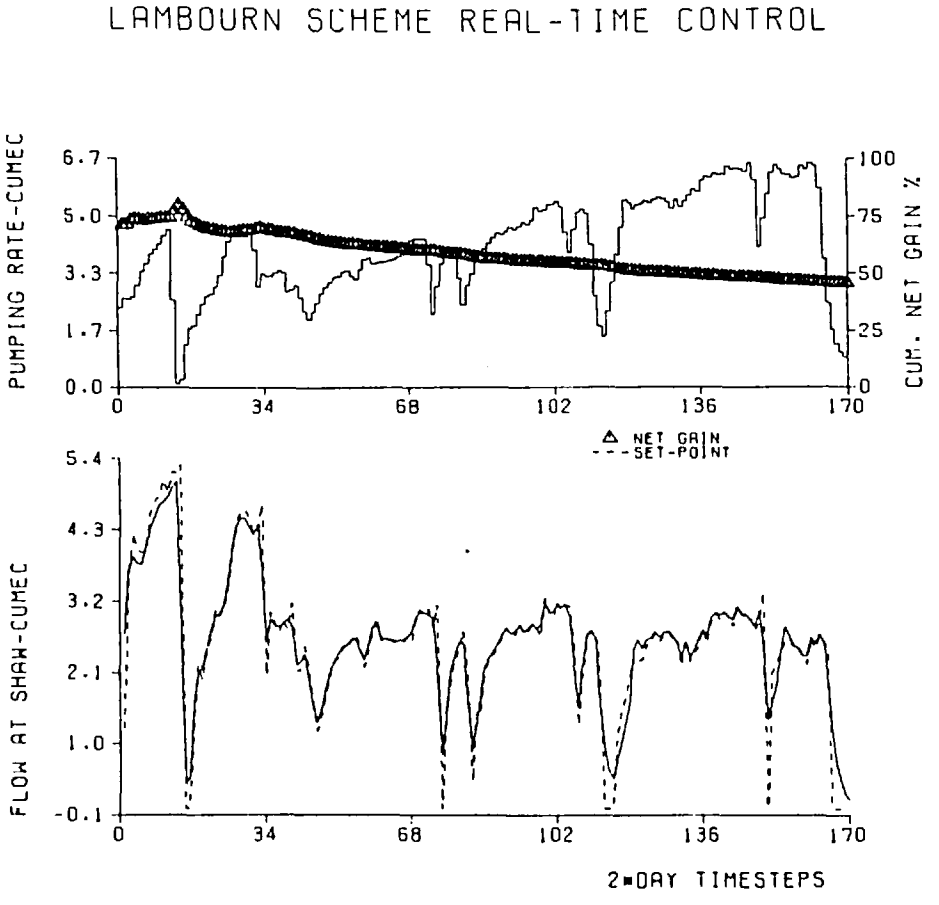


Fig. 6.17 The self-tuning controller used for real-time control of pumping rates for low-flow augmentation.

The variations in the parameters starting with the initial values $\phi(0)$, $\hat{\theta}(0)$, $P(0)$ as given by equations 6.37 are shown in fig. 6.18, again demonstrating the characteristic behaviour of the recursive parameter estimates of a self-tuning algorithm.

The autocorrelation and cross-correlation functions $r_{\psi\psi}(\tau)$ and $r_{\psi u}(\tau)$ are shown in fig. 6.19. The sample size used was 160 (limited by data availability) for which the 95% confidence interval is ± 0.15. This indicates, from equation 6.39, that the self-tuning properties are satisfied except for $r_{\psi\psi}(1)$ which is significant. The mean and variance of the output function $\psi(t)$ were 0.0077 and 0.111 respectively.

A simulation over the same period of data, but using a different set-point is shown in fig. 6.20. In this case the set-point is less than the base flow for about 70 days after the starting point in time. During this period, in the absence of system response, the model parameters are not estimated. After 70 days the parameters however tune very rapidly and good control is obtained after about 10 timesteps (20 days).

With the commencement of pumping (fig. 6.20) it is interesting to note that the cumulative net gain drops sharply and then recovers. This is due to the water pumped into the river initially seeping back into the ground at a high rate. As a result, a rise in the water table occurs underneath the river bed, which consequently reduces the loss through the river bed after some time. The required increase in the pumping rate is clearly shown corresponding to the decreasing cumulative net gain.

The above simulation shows that the self-tuning controller is able to handle the control problem associated with the above groundwater abstraction and river regulation problem. The simplifications and assumptions made on the simulated system would require some extensions and clarification if the control system were to be applied in reality.

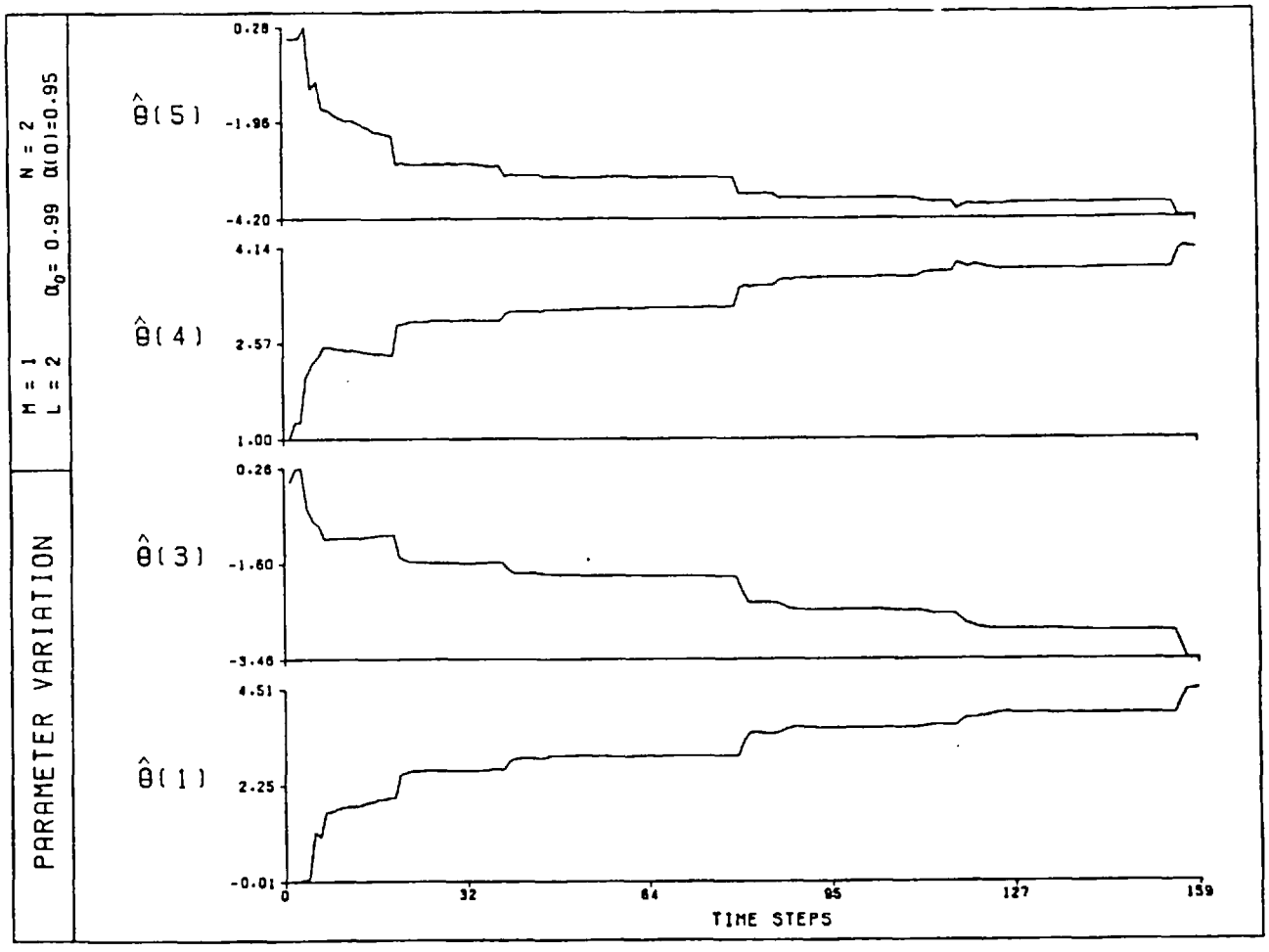


Fig. 6.18 Recursive parameter estimates of the self-tuning controller. Here $\hat{\theta}(5)$ corresponds to e_o, $\hat{\theta}(4)$ to e_1, $\hat{\theta}(3)$ to c_1 and $\hat{\theta}(1)$ to g_o in equation 6.52.

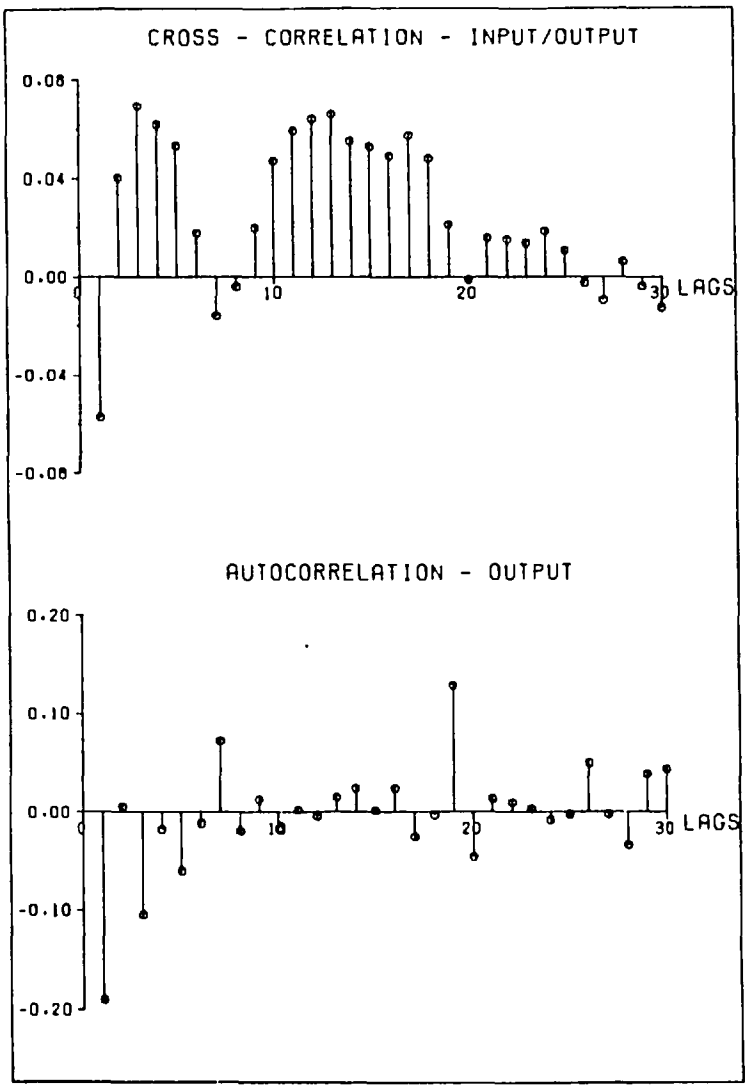


Fig. 6.19 *Autocorrelation $r_{\psi\psi}(\tau)$ and cross-correlation $r_{\psi u}(\tau)$ corresponding to fig. 6.17. The 95% confidence interval is ±0.15.*

Simulation results

The pulse response of the groundwater model and the simulated base-flow recession are shown in fig. 6.16. For all the simulations carried out, the starting groundwater table used was characteristic of the end of an average wet winter, and it was assumed that no recharge occurred. The apparent rapid response of the River Lambourn (model) is due to the large 2 day timestep used. The variable demand at Theale was chosen such that the required compensation flow (set-point) due to groundwater pumping was suitably complex. The purpose of the exercise was to show that a controller could be applied to the real-time control of the Lambourn pumping scheme and hence the set-point does not necessarily represent a true-life situation. On the other hand it may present a control problem which is of equal complexity.

Results of the simulation plotted after the 10th time-step and given the starting parameter values of equations 6.37 are shown in fig. 6.17. The deep troughs in the set-point correspond to the flood waves of the River Kennet and a set-point of zero indicates that zero compensation flow from the Lambourn is required to satisfy the flow demand at Theale. The controller parameters were selected (using the procedure already described) as follows

$$\begin{aligned} \ell &= 2 \qquad n = 2 \qquad \lambda = 0.001 \\ m &= 1 \qquad k = 1 \end{aligned} \tag{6.51}$$

so the vectors $\phi(t)$ and θ are given by

$$\begin{aligned} \phi(t) &= [q(t), w(t+1), w(t), u(t), u(t-1)] \\ \theta^T &= [g_o, -1, c_1, e_o, e_1] \end{aligned} \tag{6.52}$$

The cumulative net gain G_t defined as

$$G_t = \sum_{i=0}^{t} [q(i) - b(i)] / \sum_{i=0}^{t} u(i) \tag{6.53}$$

where $b(t)$ is the natural base flow corresponding to $q(t)$ in the absence of any pumping, is also shown in fig. 6.17.

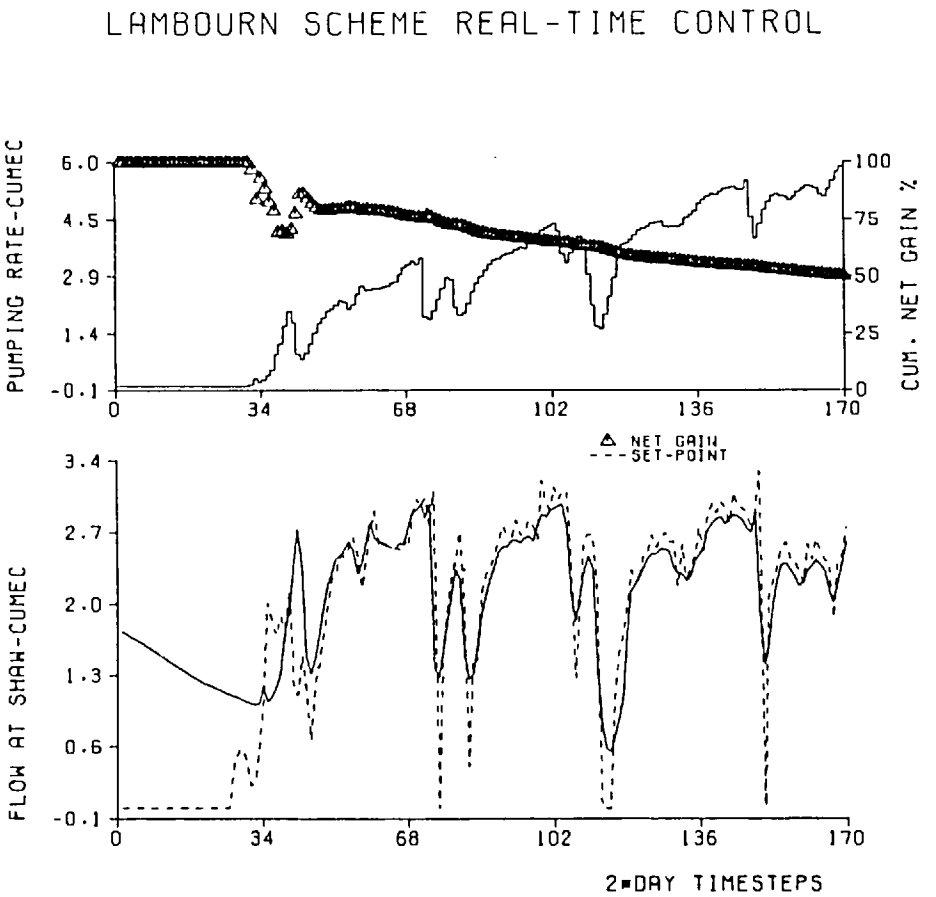


Fig. 6.20 *The self-tuning controller as applied in fig. 6.17, but with a different set-point.*

The cumulative net gain after 170 days was raised from 45% to 47% (the same data and set point were used). The relationships of equations 6.54 were arbitrarily chosen and in practice relationships would have to be derived which would satisfy

(i) the efficient lowering of the water table in the sense of maximising the net gain;

(ii) the consideration of different pump capacities and the scheduling of pump operations for multi-well abstraction sites (well location);

(iii) the handling of the loss of one or more pumps due to failure or to maintenance operations.

It was assumed that the pumps were of the continuously variable speed type. If this is not so (with item (ii) above) pumps would have to be scheduled with aggregate pumping rates varying in steps so as to approximate as closely as possible the control variable. Obviously if the incremental change in aggregate pumping is too coarse parameter estimates will be very poor and an unstable controller may result.

If cheap rate electricity is available during certain times (usually night and weekends), a further consideration to be made is the minimisation of pumping costs by running the pumps only during low tariff periods, or alternatively reducing the pumping rates outside these periods. The control system can be applied to such a situation provided the pumping periods are consistently scheduled for every sampling interval, and careful consideration is made of the sampling interval itself. For example if the pumps were allowed to run for a fixed 12 hour period every day a 9 hour time-step could not be used. A 24 hour time-step could however be used as the period during which the pumps are switched on is consistent relative to the time-step. Similarly a 3 hour time-step could be used if the controller parameters are not updated and the control variable is set to zero during the period of no pumping.

Well location

To minimise loss due to recirculation it is desirable to restrict the lowering of the watertable along the river. An obvious solution is to position wells well away from the river, although constraints on this exercise include the cost of transmission and water quality and yield criteria. Abstraction from the river wells in the 1968 pilot scheme Lambourn Group (see well locations 11 to 15 as shown in fig. 6.13) has been simulated (Oakes and Pontin, 1976) with an aggregate pumping rate of 0.5 cumec. From data given by Oakes and Pontin the cumulative net gain for a 20 day period starting with a baseflow at Shaw of 1.7 cumec was computed and found to be 28%. On the other hand fig. 6.17 shows that the cumulative net gain after 170 days when pumping from the chosen 10 sites was 45%. The figures are not directly comparable though they do give an indication of the effect of the drawdown near the river. Oakes and Pontin also found that when well location no. 11 (see fig. 6.13) was included in a group pumping test with well locations 1 to 10 the River Lambourn around river gridpoint 8 would dry up, while base flow still continued to occur further upstream. The loss of base flow over this section is due to the influence of the drawdown zone of well location 11.

Pump scheduling

The self-tuning controller can only handle one control variable, e.g. the aggregate pumping rate u(t). It was assumed that the abstraction rates from all 10 well sites were the same. To investigate the effect of differential abstraction rates, the well sites furthest from the river were assumed to be pumping at a greater rate than those nearer the river. The following relationships were used

$$P_a = 0.1667\, u(t)$$

$$P_b = 0.0714\, u(t) \qquad (6.54)$$

where

P_a = pumping rate for each well location 1, 5 and 8

P_b = pumping rate for each well location 1, 3, 4, 6, 7, 9 and 10.

6.4 REAL-TIME CONTROL FOR THE RIVER DEE SYSTEM

The River Dee basin (see fig. 2.2) and problems associated with its operational management have already been described in section 2.1. The availability of flow data (held at the Water Research Centre, Medmenham and the Institute of Hydrology, Wallingford) together with a mathematical flow routing model of the River Dee developed by the Hydraulics Research Station, Wallingford (Price, 1975b) makes this system a convenient case study for a simulated real-time control problem. The system was considered from Bala down to Manley Hall. Half-hourly flow data for the River Dee at Bala and Manley Hall, and for the Alwen, Hirnant and Ceiriog tributaries were available. The control problem considered was again that of low flow augmentation, with releases being made at Bala, so that a minimum flow requirement was not violated at Manley Hall. The main problem in this case is the travel time from Bala to Manley Hall (64 kilometres) which may vary from 19 hours at low flows to 8 hours during flood events.

An indication that it would be possible to apply the self-tuning controller to the regulation of the River Dee was given by Wood (1979) who showed that it was possible to get good 2 hour ahead forecasts of the flow at Manley Hall by using a linear ARMAX model representation of the system.

River Dee simulation model

For this simulation study, ½ hourly flow data for Bala, Manley Hall and the Alwen, Hirnant and Ceiriog tributaries for the period 2 November to 27 November were used. Simulation of the River Dee between Bala and Manley Hall was carried out using a model based on the Variable Parameter Diffusion (VPD) method (Price, 1975a). An important feature of this method is that it includes implicitly the variation of travel time and attenuation parameters with discharge for simulating both low and high flows. The model has been verified on several British rivers (Price, 1973) and in particular has been found to give acceptable results for the River Dee (Price, 1975b). The basic equation describing flow along a river reach for

the VPD method is

$$\frac{\partial Q}{\partial t} + \bar{c}(Q)\frac{\partial Q}{\partial x} = Q\bar{a}(Q)\frac{\partial^2 Q}{\partial x^2} + \bar{c}(Q)Q_i \qquad (6.55)$$

where

Q is the discharge;

x is the distance from the upstream section;

$\bar{c}(Q)$ is the average wave speed along the reach;

$\bar{a}$ is a diffusion parameter averaged along the reach and describing the attenuation of peak discharge;

Q_i is the lateral inflow per unit length along the reach.

Equation 6.55 can be formulated into a set of implicit linear finite difference equations which are solved using Gaussian elimination.

The main channel of the River Dee from Bala to Manley Hall is modelled as two distinct reaches; Bala to Corwen and Corwen to Manley Hall. The first reach has glacial deposits along its length which creates an aquifer although this does not considerably affect the flows. There is also an irregular flood plane which is subject to some inundation but does little to reduce the flood peaks. The reach between Corwen and Manley Hall runs between two ranges of hills and is comparatively steep. The inundated area of the flood plane can be significant under high flows. The subcatchment flows from the three gauged tributaries, the Hirnant, the Alwen and the Ceiriog which represent in area 40% of the Dee catchment, typically contribute to 60% of the flow measured at Manley Hall. The lateral inflows from the ungauged catchment are included in Q_i. The routing is done from tributary to tributary, so that in the upper reach there are four sub-reaches: Bala to the Hirnant, the Hirnant to the Ceidiog, the Ceidiog to the Alwen, and the Alwen to Corwen. The lower reach is divided into two sub-reaches: Corwen to the Ceiriog and the Ceiriog to Manley Hall. The flows of the Ceidiog for which data were not available are assumed to be proportional to the flows of the Hirnant. The computer program and a more detailed description of the model as applied to the River Dee is given by Price (1975b).

Control simulation

The model was tested using wavespeed and attenuation parameters obtained from the Institute of Hydrology and gave a goodness of fit (explained variance) of 81.1% for a sample size of 500 data. A schematic representation of the simulated control system is given in fig. 6.21.

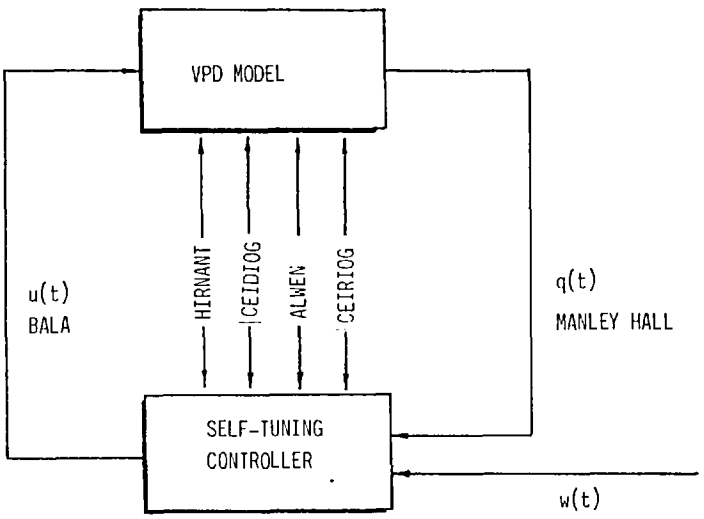


Fig. 6.21 Control system for the River Dee

The control variable u(t) is the release from the Bala sluice gates and the system output q(t) (to be controlled) is the discharge at Manley Hall. Observations of the Ceiriog qc(t), Hirnant qh(t) and Alwen qa(t) tributary flows were assumed available and treated as feedforward terms. The digital simulation procedure consisted of the VPD model of the Dee, between Bala and Manley Hall, the self-tuning controller algorithm and the data for the tributary flows. The Bala releases obtained with the self-tuning controller were then routed down to Manley Hall using the VPD model. The data and parameter vectors are given by

$$\phi(t) = [q(t),..;\ qa(t),..;\ qc(t),..;\ qh(t),..;\ w(t+k),..;\ u(t),..]$$

$$\theta^T = [g_o,g_1,..;\ ga_o,ga_1,..;\ gc_o,gc_1,..;\ gh_o,gh,..;\ -1,c_1,..;\ e_o,e_1,..] \quad (6.56)$$

Using the procedure outlined in section 6.2, the number of terms in the ϕ and θ vectors were selected as follows.

n = 1 ; no of w(t) terms
m = 2 ; no of q(t) terms
ℓ = 2 ; no of u(t) terms
pa = 1 ; no of qa(t) terms
pc = 1 ; no of qc(t) terms
ph = 1 ; no of qh(t) terms

The controller time-step (sampling interval), delay and input weight which gave the best results were 3 hours, k = 2 and λ = 1.4 respectively. Results for the simulation are shown in figs. 6.22 and 6.23. A constant set-point (at Manley Hall) of 40 cumec was assumed. The data are plotted in $\frac{1}{2}$ hourly increments as this was the time-step of the VPD model. By visual inspection of the hydrographs it was determined that the travel time from Bala to Manley Hall was between 8 hours (a in fig. 6.22) and 9 hours (b in fig. 6.23). Travel times for the various gauged inputs down to Manley Hall obtained by Wood (1978) using the CLS model (Natale and Todini, 1976) were

Bala 8 hours
Hirnant 7 hours
Alwen 4 hours
Ceiriog 0 hours

It should be noted however that these values may be influenced by the effect of ungauged lateral inflows. The travel time from Bala to Manley Hall assumed for the controller was 6 hours ($\Delta t = 3$ hrs and k = 2). The reason for this is as follows:

(i) To take into account a particular tributary flow, the controller delay should be less than or equal to the lag between the tributary confluence and Manley Hall.

(ii) While it is not obvious that the output q(t+2) is dependent on the control u(t) if a travel time of 8 to 9 hours is assumed (the delay, k = 2), it can be assumed that q(t+2) is dependent on u(t-1). Thus,as u(t) is correlated with u(t-1), q(t+2) can also be assumed to be related to u(t).

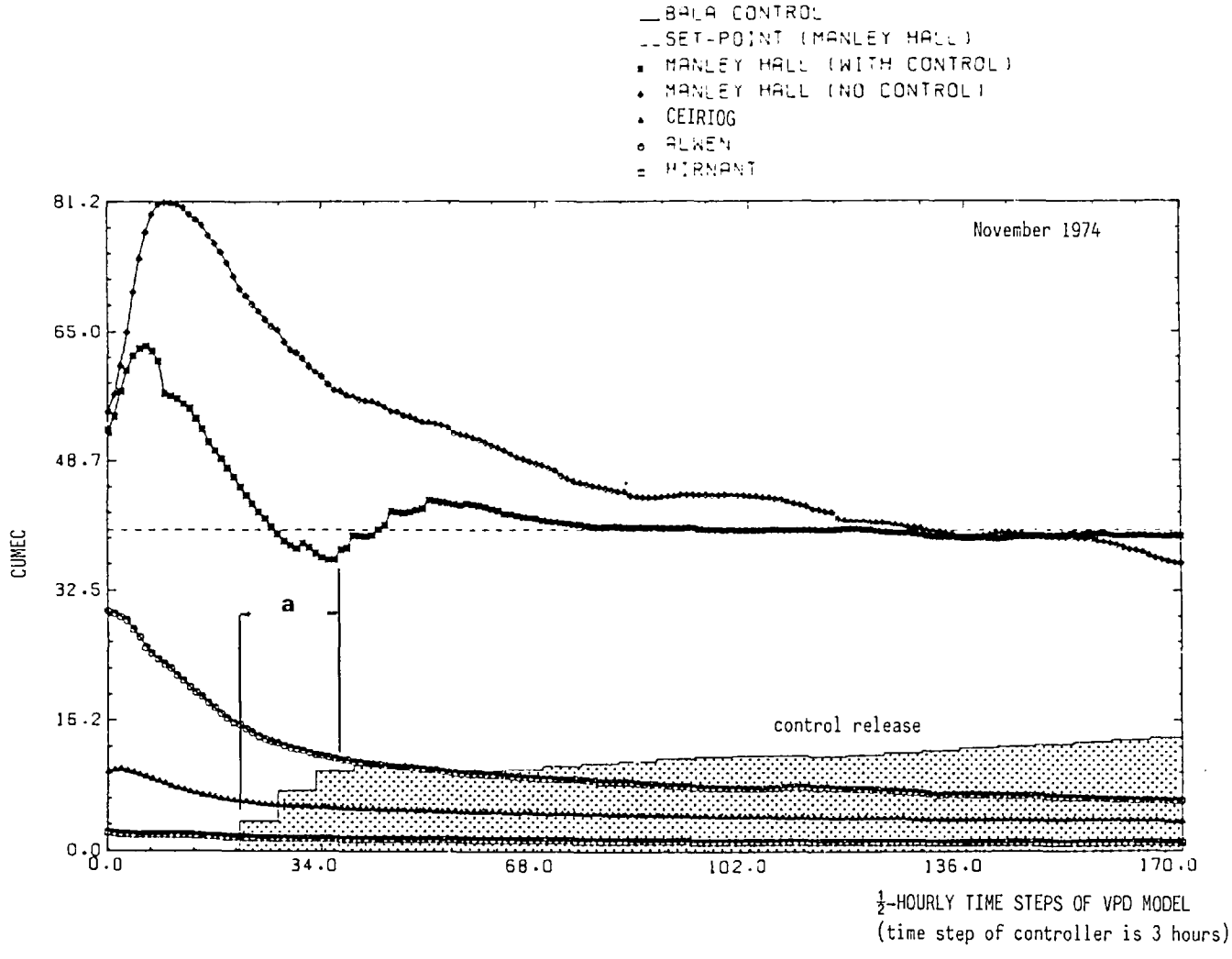


Fig. 6.22 The self-tuning controller applied to the real-time control of the simulated River Dee system.

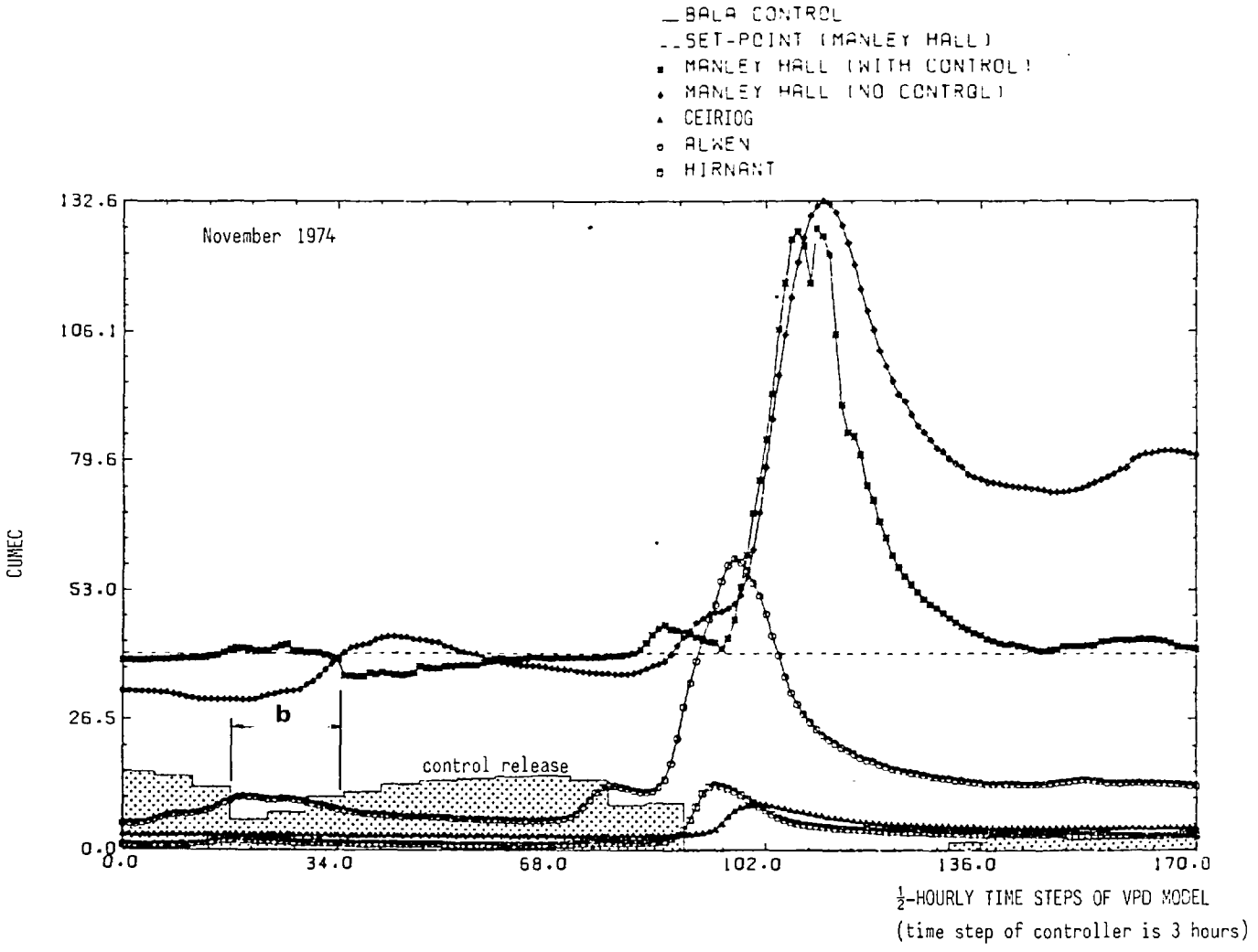


Fig. 6.23 The self-tuning controller applied to the real-time control of the simulated River Dee system.

The values of k and Δt selected therefore achieve a compromise between (a) using the correct controller delay corresponding to the Bala to Manley Hall travel-time and (b) taking into account the tributary flows (as feedforward variables). It should be noted that the travel-time for the River Dee between Bala and Manley Hall can be considerably greater than 9 hours, however this occurs at very low flows when the variations in control inputs are not likely to be great. Relatively speaking, under such conditions reasonable control may be obtained with poorly estimated controller parameters.

A compromise of the flows at Manley Hall (figs. 6.22 and 6.23) (a) with controlled releases and (b) without controlled releases show that the use of the self-tuning controller results in a reduction in the excess releases when low flow augmentation is the control objective. The flow at Manley Hall shown corresponding to uncontrolled releases from Bala is in fact the historic flow sequence corresponding to the flow record for the Ceiriog, Alwen and Hirnant tributaries, and the actual release from the Bala sluice gates. The recorded releases from Bala were almost constant during the periods in November 1974 that are shown in figs. 6.22 and 6.23. The results also show that reductions in the excess releases are complemented by reductions in the flood peak.

The above example shows a relatively complex subsystem,where from the control aspects, the long travel time presents a difficult problem. The results show that the controller is able to handle this control problem without difficulty. The controller also allows multiple inputs (feedforward terms), and clearly in this example these could have been forecast variables. These forecast variables, which would be outputs from forecasting models, would almost certainly ensure a reduction in the control error. In Section 5.5 forecasts for the River Hirnant (one of the River Dee tributaries) were in fact obtained using the self-tuning predictor.

So far, subsystems with only one control loop have been considered. If the Alwen reservoir (fig. 2.2) was also to be controlled in real-time (this is a hypothetical situation), two control loops would

have to be used, as the self-tuning controller can only handle one control input. In effect, there would be two separate controllers with independently estimated parameters employed, though only one control loop could be operating at any one time. Following the satisficing approach outlined in Section 2.2, the two control loops would form a primary and secondary control objective. The primary objective could be to satisfy the downstream set-point by releasing from Bala until there is insufficient storage left (defined by a rule curve), while the secondary objective could be to satisfy the set-point by releasing from Alwen when there is insufficient storage to release from Bala, but storage available in Alwen. When controlled releases are being made from Bala, releases from Alwen would be held constant or set to zero, while when controlled releases are being made from Alwen, the releases from Bala would be held constant or zero.

A more interesting problem however is that of flood alleviation (again a hypothetical situation), when both the reservoirs could be controlled simultaneously to allow for maximum releases from the storages without causing downstream flood damage.

The flood control problem has previously been studied using dynamic programming techniques by Schultz and Plate (1976) (see also Meyer and Zurwelle, 1973). They studied the problem of developing optimal release strategies for a single storm event for two reservoirs arranged both in series and in parallel. Their method assumes however that a forecast of the whole flood event (rather then the one step ahead forecast) is available. Based on the forecast hydrograph, strategies which minimise the magnitude of the downstream flood peak while preserving the flood volume (inflow and outflow hydrographs) during the optimization period are developed. It is assumed that the releases can be routed using simple lags, while the constraints applied are that the reservoirs are empty at the beginning of the flood and full at the end of the operation.

It would be possible to run two controllers (the Bala to Manley Hall loop, and the Alwen to Manley Hall loop) simultaneously provided that the sampling times for the controllers (and the computation

of the control inputs) were separated by half a sampling interval. Thus the releases from the Alwen could be made at times t, t+1, t+2, and the releases from Bala at times $t+\frac{1}{2}$, $t+1\frac{1}{2}$, $t+2\frac{1}{2}$, ... Data would be actually sampled at time-steps t, $t+\frac{1}{2}$, t+1, $t+1\frac{1}{2}$, ... In other words the recursive equations would be 'out of phase' by half a sampling interval. The control input from the Alwen controller would be a feedforward input to the Bala controller, while the control input from the Bala controller would be a feedforward input to the Alwen controller. The objective of both controllers would be to minimise the control error in the control station flow (using the same set-point). The weights λ of the respective controllers serve as important factor in assigning relative priority levels to making releases or conserving water from the corresponding reservoirs.

In practice, the two control algorithms would actually be the same computer program in the same computer. The two controllers would be specified by the respective sets of variables $P(T), \phi(t), \theta(t)$, etc., stored in the computer.

Clearly, the scope for solving different problems using control modules and forecasting modules is extensive, though the complexity of the problem that can be solved will be largely dependent on the ingenuity of the system designer. The resulting controls will generally be sub-optimal, though feasible in a practical sense to implement.

6.5 SUMMARY

In Section 6.1 a self-tuning controller, which can be applied to the control of a water resource subsystem, was developed. The self-tuning controller is similar to the self-tuning regulator described in Section 4.3 except that a set-point is directly included in the quadratic loss function V_3 (equation 6.13), to be minimized. The self-tuning controller consists of a predictor of the output function $\psi(t)$ given by the equivalent model, equation 6.24. This model is obtained by combining the original system model equation 6.1 and the loss function V_3. The estimated parameters of the predictor (using recursive least-squares with exponential weighting) give the parameters of the control law.

The loss function V_3 also includes a weight λ which attaches a penalty to the rate of change of the control variable. The important results that this leads to are:

(i) excessive control action can be reduced by increasing λ;

(ii) the stability of the controller increases as λ is increased and systems (especially non-minimum phase systems) that are difficult to regulate using the self-tuning regulator may be controlled.

As with the self-tuning regulator, the convergence of the controller (in the sense of minimum mean square error between output and set-point) can be verified by checking whether conditions given by equations 6.39 hold.

The self-tuning controller assumes the system can be represented by an ARMAX model. This allows considerable flexibility when considering more complex subsystems, and in particular, when linking together forecasting models and controllers.

In Section 6.2, 6.3 and 6.4, the self-tuning controller was applied to river regulation problems using three simulated subsystems. These were as follows:

(i) regulation by reservoir releases of a hypothetical river with lateral inflow (surface runoff);

(ii) regulation of the River Lambourn by groundwater abstraction which is piped directly into the river;

(iii) regulation of the River Dee at Manley Hall by controlled releases from the Bala sluice gates.

In each case the control objective was to minimise the variances of the flow about a set-point at the downstream control station.

The implementation of the self-tuning controller requires that $P(0)$, $\phi(0)$, $\theta(0)$, k, λ, $\alpha(0)$, α_o and the number of variables in the data vector $\phi(t)$ and the number of parameters in the parameter vector θ,be known a priori. These were estimated using a subjective selection procedure based on a minimum $\psi(t)$ variance criterion. The selection procedure is analogous to that finally adopted for the self-tuning predictor in Chapter 5.

Examples (ii) and (iii) above represented hypothetical subsystem control problems based on the simulation of real physical systems and associated control problems. It appears that the self-tuning controller could form the basis, in the form of a standard control module, of a relatively complex subsystem control problem. Here control modules could be linked in a hierarchy of control loops, aided by the output from forecasting modules. The system operation would be dependent on a satisficing approach.

7. CONCLUSIONS

7.1 GENERAL SUMMARY AND CONCLUSIONS

The development of optimal strategies for the real-time (or short-term) management of a large water resource system is usually a very complex procedure because of the many variables to be considered, the many uncertainties affecting the dynamic behaviour of the system inputs and outputs, and the many alternative strategies which have to be evaluated. Thus while real-time models have been developed with some success for hydrological forecasting to aid decision-making, there has been little research in the field of real-time (and automatic) control of a water resource system. This may be due to engineers tending to assume that the real-time control problems associated with even a small system are extremely difficult to solve.

Large multi-functional water resource systems however, can generally be reduced to a number of simpler subsystems which can be individually controlled in real-time. In Section 2.2 a simple strategy was formulated for the real-time control of a water resource system. The strategy also takes into account the medium-term system operation so that while the subsystems are to be operated in real-time, the actual objectives of subsystem control are based on medium-term strategies. Geographical separation of the subsystems favours a distributed computer configuration comprising a central mainframe computer and a mini/micro-computer for each subsystem. This reduces the unnecessary unreliability that may be associated with on-line data acquisition and real-time processing at one central site.

The computations for the real-time operation of the subsystem can be based on a forecasting algorithm, and a stochastic feedback control algorithm. The controller objective can consist of the minimization of the control error (difference between the set-point and the output variable). Each subsystem may have several control loops which can be arranged in a hierarchical order of a primary control objective and several secondary control objectives. Only one control objective is considered at one time, and the selection of this control objective may be based on a simple pre-determined set of rules or a decision table. The set-point for the current

control loop may be determined at a medium-term level. This satisficing approach is clearly sub-optimal (in terms of the least-cost solution to the real-time operation), though the attractiveness of the method is that it appears both technically and economically feasible. Upgrading of the real-time control strategy would be done mainly be altering procedures for selecting the set-points; for instance if the controller time-step was one hour, the set-point could be updated in real-time at six hourly intervals rather than at a medium-term level. Furthermore a switch between any two of the hierarchy of primary objective and secondary objectives could be performed, not from a set of pre-determined rules, but instead by some real-time decision making procedure based on minimising a complex cost function.

For the economical implementation of the above proposed strategy, an important aspect to be considered is the standardization of hardware and software. It is envisaged that a mini/micro-computer based system be used for the subsystem control and data handling. While the cost of hardware is falling rapidly, the cost of software development is increasing. Standardization of black-box forecasting and control packages could substantially reduce software development costs, and enable micro-processor based computers to be developed specifically for these tasks.

In this study, the application of the self-tuning predictor to flow forecasting and the application of the self-tuning controller to river regulation have been investigated. The principle conclusions are outlined below.

The self-tuning predictor

The self-tuning predictor described in Section 5.3 is a minimum mean square error predictor (assuming an ARMAX model), the parameters of which are estimated using recursive least-squares. The self-tuning predictor is essentially an adaptation of the self-tuning regulator, discussed in Section 4.3, to the prediction problem. An important issue concerning the self-tuning predictor is that it is basically a linear black-box model, rather than a conceptual

one. Many hydrologists believe that a linear black-box model is unable to model non-linear catchment behaviour because specific catchment processes such as infiltration, evaporation etc., are not explicitly described by the model (and also that the model is linear). The points to note are however:-

(i) the predictor is being implemented in real-time, and thus forecasts are updated using the latest on-line observations. Linear ARMAX models have autoregressive components which prevent cumulative errors building up;

(ii) the parameters of the predictor are recursiviely estimated (using recursive least-squares), and by use of a 'forgetting factor' the parameters can be made time-variant. The parameter estimator also takes into account system noise and modelling errors.

The above features usually do not apply to conceptual models. Furthermore, the self-tuning predictor requires a minimum of computer memory and processing time. Conceptual models on the other hand often require to re-process the complete set of historic data at each time-step and therefore require large computer memories.

Flow forecasts for the River Brosna and the River Hirnant using the self-tuning predictor as described in Sections 5.4 and 5.5 showed that good results were obtained, providing the lead-time did not exceed the pure time delay of the catchment.

Referring to some more specific points relating to the self-tuning controller; it was found that the model structure selection, as described in Section 3.3, was best carried out on a subjective judgement basis, guided by both statistical tests of the residuals and a minimum square prediction error criterion. The results obtained using the Box and Jenkins method of prewhitening and cross-correlating the input and output were found to be difficult to interpret and hence unreliable. The convergence of the parameters was also found to be very rapid, assuming the general initial conditions given in Section 5.3.

An attempt was also made to improve the performance of the self-tuning predictor by introducing an auxiliary variable, the perlog. Thus the multiple input capability of a black-box ARMAX model was demonstrated. The results, given in Section 5.6 showed that there was a marginal but consistent reduction in the prediction error when using the perlog input. This is to be expected as the non-linear perlog function is related to the catchment rainfall input and antecedent conditions.

In Section 5.7 one step ahead forecasts of stage were obtained using three self-tuning predictors with the forecast variable as stage, flow rate and cross-sectional area of flow respectively. From the results obtained it was not possible to conclude which was the best predictor (see table 5.4), though a close inspection of figure 5.31 indicates that the results are different while the pattern of these differences is surprisingly consistent. General conclusions relating to the forecast variables are difficult to make as these are dependent on the cross-sectional geometry of the channel section in question. The important conclusions that can be made here however, are that with a black-box model, any forecast variable describing the magnitude of the flow may be used, and that errors in the stage discharge curve are not necessarily related to errors in forecast stage.

The self-tuning controller

The self-tuning controller differs from other algorithms previously used in river regulation studies (such as dynamic programming), in that it is a feedback controller and hence is dependent on real-time applications. The method does not handle complex objective functions and constraints, but careful formulation of the operational management procedure (for instance as proposed in Section 2.2) can lead to a sub-optimal, but satisfactory solution. The self-tuning controller comprises a predictor (assuming an ARMAX system model) and a feedback control law, whose parameters are determined from the parameters of the predictor. The parameters of the predictor are estimated using recursive least-squares.

The self-tuning controller is similar to the self-tuning regulator, from which the self-tuning predictor is also derived, except that the cost function to be minimised incorporates the set-point directly, and that a weight is selected to limit the rate of change of the control variable (eq. 6.11). The advantages this self-tuning controller has over the self-tuning regulator are as follows:-

(i) the set-point which may be variable is included directly in the loss-function;

(ii) a weight λ is assigned to limit the rate of change of the control variable, preventing undesirable excessive fluctuations in this control input;

(iii) the stability of the controller can be increased by increasing the magnitude of λ;

(iv) the regulator requires that one parameter be estimated a priori and fixed during recursive parameter estimation, while for the self-tuning controller, the parameter c_o is known ($c_o = -1$) and so this is assumed to be the fixed parameter.

Three real-time river regulation studies were made using the self-tuning controller (see Sections 6.2, 6.3 and 6.4) each with a simulated subsystem. The simulated subsystems were as follows:-

(i) A reservoir from which controlled releases were made and routed through two Muskingum reaches down to the control station. A lateral inflow component and corresponding rainfall observations were included as part of the simulated subsystem.

(ii) An unconfined aquifer, from which controlled ground-water abstraction was piped directly into a river to maintain set-point flows at the downstream control station. A digital model of the Lambourn Valley chalk aquifer (and the River Lambourn) was used in this study, with control being made with respect to a set-point at Shaw as shown in the figure 6.12. The particular problem associated with this example is that, as the ground water table is depleted the loss of flow in the river through the river bed is increased. This system is non-linear and time-variant.

(iii) A reservoir from which controlled releases were made and routed down the control-station using the VPD method. The River Dee between Bala and Manley Hall was the subject of this simulation. The River has a variable travel-time ranging from 9 to 16 hours, making this system very difficult to control. Observations from three tributaries of the Dee were used as feedforward terms in the self-tuning controller.

The results from the above examples were, from theoretical and practical viewpoints, highly acceptable. Statistical tests carried out with the residual (control error) series obtained from examples (i) and (ii) showed that the minimum variance controller was often not being obtained. This may be attributed to the loss function of the self-tuning controller which is only zero when the rate of change in the control variable is zero. The results obtained from these sub-optimal controllers were still good. Example (iii) showed that the self-tuning controller can be used for the control of a more complex subsystem.

The major advantages of the feedback controller approach over the classical approach (e.g. dynamic programming) for aiding operational management decisions, are that the computing capacity required is far less, and that the parameters of the controller are estimated recursively using on-line data. The self-tuning controller assumes an unknown system, and could form the basis of a modular control package which can be applied to a wide range of subsystems. Furthermore, the outputs of forecasting models can be used as inputs (feedforward terms) to the control module. The self-tuning predictor can itself form the basis of a modular forecasting package.

The overall design of the water resource management system as proposed in this study would require a careful design of the decentralized control system configuration, with special attention being paid to the hierarchical control structure for each individual subsystem. The primary design tasks associated with the setting up of an on-line control procedure for an existing water resource system would be as follows:-

(i) The decomposition of the water resource system into appropriate subsystems.

(ii) The design of the hierarchical real-time control system for each subsystem based on the linking of forecasting and control modules. The medium-term control of the subsystem would be dependent upon the set-points.

(iii) The design of the medium-term control strategy to be implemented at the central computer centre. The medium-term strategy would have to assure the necessary level of interaction between the subsystems. The control strategy would be obtained using some optimization procedure, probably based on dynamic programming. This problem would form the major design effort for the implementation of the operational procedure.

(iv) The design of the control system hardware, including the distributed mini/micro-computer based computing system, and all necessary equipment interfacing.

One of the major constraints in the efficient real-time operation of the water resource system will be that of equipment failure. However, the improvement of equipment reliability, well organized maintenance programmes and measures taken to increase the robustness of installations should within some years result in highly reliable rainfall and streamflow measurement equipment and even more reliable on-line control devices.

The justification for a real-time automatic control strategy for the operation of the water resource system is in a reduction in the cost of control hardware and the increase is the cost of employing staff, together with an increase in the standard of service provided. Standardisation of black-box forecasting and control algorithms would also reduce software development expenditure.

Main conclusions

The main conclusions that can be synthesised from the foregoing discussion are:-

(i) The control of a large water resource system in real-time invariably results in a complex multivariable control problem. A feasible approach to this problem is to decompose the system into a number of subsystems which can be individually controlled in real-time.

(ii) The water resource system as a whole can be operated on the basis of a medium term policy (on say a weekly, monthly or seasonal time-step). If the medium-term policy outputs are regarded as the real-time control objectives for the respective subsystems, an effective coordinated short and medium-term management strategy can be achieved.

(iii) The implementation of the above strategy can be carried out with a regional control centre and outstation control stations for each subsystem. This approach conveniently fits into a framework of a hierarchy of mini/micro-computers and a regional computer centre.

(iv) The real-time control of a water resource system basically requires two types of algorithm, namely a prediction algorithm for the forecasting of river flows and a control algorithm for the control of the various control variables such as reservoir releases and pumping rates.

(v) A real-time forecasting algorithm known as the self-tuning predictor has been found to be suitable for the real-time forecasting of hydrological variables. In particular:-

(a) The predictor is obtained by deriving the minimum square error predictor for a process described by an ARMAX model.

(b) The parameters of the predictor are estimated at each time-step using recursive least-squares. The computer memory storage required is therefore very small.

(c) Because of the autoregressive structure of the model, cumulative errors are not allowed to build up.

(d) The algorithm converges rapidly, even with poor initial parameter estimates. The parameters converge, so that in the limit the minimum square error predictor is obtained.

(e) The self-tuning predictor can handle changes in system dynamics by introducing exponential weighting of the past data.

(f) The predictor can handle multiple-inputs/auxiliary variables with ease. System noise is also taken into account.

(g) Non-linear models may be developed by using non-linear inputs that may be functions of catchment characteristics, such as the perlog function.

(h) The model is easy to implement.

(vi) The self-tuning controller is a real-time control algorithm which has been found to have considerable potential for the real-time control of water resource systems. In particular:-

(a) The self-tuning controller is a feedback-feedforward controller, thus control signals are generated with reference to previous errors in control and observable but uncontrollable system inputs.

(b) The controller is assumed to be described by an ARMAX model with unknown parameters which are estimated using recursive least squares. The computer memory required is therefore minimal.

(c) The set-point (the output target for the system being controlled) can be variable.

(d) A penalty can be assigned to the rate of change of the control variable (or to the magnitude of the control variable, depending on the loss-function chosen).

(e) The convergence properties of the algorithm are good, and in the limit the controller that minimises the square of the chosen loss-function is usually obtained.

(f) The controller is applicable (referring to stability) to systems with large delays and non-linear systems.

(g) Changes in system dynamics can be handled by exponentially weighting the past data.

(h) The controller is easy to implement.

(vii) The simulation of components of a water resource subsystem is a valuable aid in investigating, developing and assessing real-time control strategies.

(viii) Real-time (and automatic) control is likely to be economically justifiable for the short-term management of large water resource systems.

(ix) While the self-tuning predictor and self-tuning controller have been discussed in the context of large water resource systems, they are equally applicable to smaller systems such as flood warning schemes.

(x) The use of forecasting and control modules within the framework of a short-term management strategy for a water resource system has considerable development and application potential.

7.2 RECOMMENDATIONS FOR FURTHER RESEARCH

A broad framework for the design and implementation of real-time forecasting and control procedures for a complex water resource system has been outlined in Section 2.2. Forecasting and control modules which can be adopted within this framework were described in Chapters 5 and 6. However, to ensure that a comprehensive forecasting and control system can be effectively implemented, further research effort should be applied to the following topics:-

(i) water quality considerations;

(ii) quantitative rainfall forecasting;

(iii) the medium term strategy;

(iv) the long-term planning aspect of water resource system design.

The above items do not constitute a comprehensive list of research topics. Instead, this list represents some of the more important topics related to the real-time forecasting and control of a water resource system, which have not been covered in this study.

Water Quality

In this study no reference has been made to the water quality aspect of water resource management. With the present emphasis on environmental conservation and protection, the design and operation of any water resource system must take into account both water quality and water quantity. The operational management of a water resource system would thus require the on-line measurement and real-time forecasting of water quality variables. Real-time control of wastewater treatment plants and real-time control of effluent discharge into rivers may also be desirable.

Considerable further research is required for the development of instrumentation for on-line water quality measurement. The particular problem here is also that the number of variables that have to be observed can be very large. It is envisaged that the workload imposed on the data acquisition system for water quality monitoring will far exceed the demands made for the monitoring of hydrological variables.

Research is also required into the real-time forecasting of water quality variables. For each water quality variable, a prediction model must be developed. A water quality variable could if necessary be controlled through a set-point, if the value of the variable (which may depend upon the dilution of a particular pollutant) could be controlled by either reservoir releases for dilution or by upstream effluent discharge control.

Quantitative Rainfall Forecasting

In Section 2.2 it was mentioned that the real-time forecasting and control of a water resource system is largely constrained by the maximum lead-times for which streamflow forecasting can be attempted. There is thus a great need for real-time forecasting of rainfall. At present there appears to be some scope in the use of radar derived rainfall observations which can be coupled to a model based on estimation theory, to produce predictions of rainfall quantity. Similarly, satellite imagery could also possibly be used, coupled with a prediction model, to provide rainfall forecasts. The

output of this prediction model must be compatible with the flow forecasting models which are to be used.

Medium-term Strategy

The real-time operation of the water resource system as described in Section 2.2 is dependent on the individual control of each subsystem. While real-time control to satisfy an objective based on minimising the mean square error between the set-point and the output variable (e.g. the control of streamflow at the control-station) was treated, two major aspects of this approach were unanswered. These were:-

(i) the set-point has to be chosen;

(ii) if there is more than one control loop (as a hierarchy of primary and secondary objectives) and only one loop can be controlled at a time, a decision making procedure is required to select the appropriate loop.

These steps may be taken at a medium-term level (which was suggested as being a time step of between several days to three months, depending on the system). The selection of set-points for each subsystem should be formulated as a global system optimization problem (using such techniques as dynamic programming). Research is required into how this could best be done.

It was also suggested that the hierarchy of loops selected on the basis of a decision table or rule-curve relate the system state to the appropriate (or optimal) loop. A general methodology based on optimization techniques would have to be developed to do this.

The above steps (i) and (ii) could also be performed in real-time (say as emergency measures) and procedures would also have to be developed for this. The real-time control so resulting could however be carried out with a larger time-step than the actual time-step of the subsystem forecasting models and controllers.

System Design and Long-term Planning

The use of real-time forecasting and control procedures could result in a future reduction in operating costs and an immediate increase in the standard of service provided by the water industry. It could also result in a reduction in required storage capacities. Flood plain management will also be dependent on the reliability of flood forecasting. Thus any immediate long-term planning for the optimal expansion of the water resource system to meet future demands should take into account the proposed real-time forecasting and control procedures and the benefits likely to be generated by their use.

APPENDIX A

OPTIMAL PREDICTION OF DISCRETE TIME STATIONARY PROCESSES

Prediction theory can be stated in many different ways which differ in the assumptions made on the process. Here, the following set of assumptions will be made:

(i) the output of the process to be predicted is stationary and Gaussian;

(ii) the best predictor is the one which minimises the variance of the prediction error;

(iii) the admissible predictor for $y(t+k)$ is any function of the past observation, i.e. $y(t)$, $y(t-1)$, ...

The quantity to be predicted $y(t)$ is represented by the ARMA model

$$A(z^{-1})y(t) = C(z^{-1})e(t) \qquad (a1)$$

where $e(t)$ is a Gaussian white noise variable distributed as $N(0,\sigma^2)$. The polynomials in the backward shift operator are defined as

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

$$C(z^{-1}) = 1 + c_1 z^{-1} + \ldots + c_n z^{-n} \qquad (a2)$$

The k-step ahead prediction error is given by

$$\varepsilon(t+k) = y(t+k) - \hat{y}(t+k|t). \qquad (a3)$$

The optimal predictor is the predictor that minimises the loss function

$$V = E\{\varepsilon(t)^2\}. \qquad (a4)$$

This can be obtained by introducing the identity (Åström, 1970, p.167).

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-k}G(z^{-1}) \qquad (a5)$$

where

$$F(z^{-1}) = 1 + f_1 z^{-1} + \ldots + f_{k-1} z^{-k+1}$$

$$G(z^{-1}) = g_o + g_1 z^{-1} + \ldots + g_{n-1} z^{-n+1} \qquad (a6)$$

noting that z^{-k} operates on $G(z^{-1})$ in equation a5 and $F(z^{-1})$ has k terms. Writing equation a1 as

$$y(t+k) = \frac{C(z^{-1})}{A(z^{-1})} e(t+k) \qquad (a7)$$

and replacing $C(z^{-1})$ by its identity (eqn. a5) gives

$$y(t+k) = F(z^{-1})e(t+k) + \frac{G(z^{-1})}{A(z^{-1})} e(t) \qquad (a8)$$

Eliminating e(t) in equation a8 using equation a1 gives

$$y(t+k) = F(z^{-1})e(t+k) + \frac{G(z^{-1})}{C(z^{-1})} y(t) \qquad (a9)$$

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)

$$E\{[y(t+k)-\hat{y}]^2\} = E\{[F(z^{-1})e(t+k)]^2 + E\{[\frac{G(z^{-1})}{C(z^{-1})} y(t) - \hat{y}]^2\} \qquad (a10)$$

Knowing the variance of e(t)

$$E\{[y(t+k) - \hat{y}]^2\} \geq \sigma^2(1 + f_1^2 + \ldots + f_{k-1}^2) \qquad (a11)$$

The predictor objective is that of minimum variance and this is satisfied when $\hat{y}$ is $\hat{y}(t+k|t)$ and

$$\hat{y}(t+k|t) = \frac{G(z^{-1})}{C(z^{-1})} y(t) \qquad (a12)$$

Eliminating y(t) using equation a3 and $C(z^{-1})$ using equation a5,

the optimal predictor is obtained as

$$\hat{y}(t+k|t) = \frac{G(z^{-1})}{A(z^{-1})F(z^{-1})}\,\varepsilon(t) \qquad (a13)$$

Finally, the penalty of the predictor is that it contains k-1 more parameters than the model it represents.

APPENDIX B

ASYMPTOTIC PROPERTIES OF THE SELF-TUNING REGULATOR

In Section 4.2 it was stated that for the self-tuning regulator, if the parameters $a_1, \ldots a_m$ and $b_2, \ldots b_\ell$ converge as $t \to \infty$ then the covariances

$$E\{y(t+\tau)y(t)\} = 0 \quad ; \quad \tau = k, \ldots k+m-1$$

$$E\{y(t+\tau)u(t)\} = 0 \quad ; \quad \tau = k, \ldots k+\ell-1 \qquad (b1)$$

This is based on the reasoning that the parameters are estimated using least-squares, and that the input to the (closed loop) system model is generated by the feedback control law, equation 4.15.

The least squares estimates for a sample size N are given by the solution of the following equations (see Appendix D, eqn. d5).

$$\left[\begin{array}{cc:cc}
\Sigma y(t)^2 & \ldots \Sigma y(t)y(t-m+1) & -\Sigma y(t)y(t-1) & \ldots -\Sigma y(t)u(t-\ell+1) \\
\Sigma y(t)y(t-1) & \Sigma y(t-1)y(t-m+1) & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\Sigma y(t)y(t-m+1) & \Sigma y^2(t-m+1) & -\Sigma y(t-m+1)u(t-1) & -\Sigma y(t-m+1)u(t-\ell+1) \\
\hdashline
-\Sigma y(t)y(t-1) & -\Sigma u(t-1)y(t-m+1) & \Sigma u^2(t-1) & \Sigma y(t-1)u(t-\ell+1) \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
-\Sigma y(t)u(t-\ell+1) \ldots & -\Sigma u(t-\ell+1)y(t-m+1) & \Sigma u(t-1)u(t-\ell+1) \ldots & \Sigma u^2(t-\ell+1)
\end{array}\right]
\left[\begin{array}{c} a_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ a_m \\ \hdashline b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_\ell \end{array}\right]$$

$$= \begin{bmatrix} -\Sigma y(t+k)y(t) & + b_1 \Sigma u(t)y(t) \\ -\Sigma y(t+k)y(t-1) & + b_1 \Sigma u(t)y(t-1) \\ \vdots & \\ -\Sigma y(t+k)y(t-m+1) & + b_1 \Sigma u(t)y(t-m+1) \\ \hline \Sigma y(t+k)u(t-1) & - b_1 \Sigma u(t)u(t-1) \\ \vdots & \\ \Sigma y(t+k)u(t-\ell+1) & - b_1 \Sigma u(t)u(t-\ell+1) \end{bmatrix} \qquad (b2)$$

where the summations are taken over N values. If efficient parameter estimates are obtained for a sufficiently large sample and it is assumed that the coefficients of the control law, equation 4.15, converge to constant values, then by replacing the $b_1 \Sigma u(t)$ terms in the righthand side of equations b2 by the control law, equation 4.15, the following equations are obtained

$$\begin{bmatrix} \Sigma y(t+k)y(t) \\ \vdots \\ \Sigma y(t+k)y(t-m+1) \\ \hline \Sigma y(t+k)u(t-1) \\ \vdots \\ \Sigma y(t+k)u(t-\ell+1) \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (b3)$$

With the assumptions of the above equations and using the control law equation 4.15 it follows that

$$\Sigma y(t+k)u(t) = 0 \qquad (b_4)$$

If the closed loop system is such that the output is ergodic in the second moments, then the above summation terms can be replaced by mathematical expectations and this gives the proof for equations b1.

APPENDIX C

THE MINIMUM VARIANCE REGULATOR

The criterion for the control problem is chosen so as to minimise the variance of output y(t) of the system represented by

$$A(z^{-1})y(t) = B(z^{-1})u(t-k) + C(z^{-1})e(t) \qquad (c1)$$

where e(t) is a Gaussian white noise variable distributed as $N(0,\sigma^2)$, k is a pure time delay and u(t) is the control input. The polynomials $A(z^{-1})$, $B(z^{-1})$ and $C(z^{-1})$ are defined by equations 4.2. To solve the problem the situation at time t is considered, when the measurements y(t), y(t-1), ... and the control actions u(t-1), u(t-2), ... are known. It follows from equation c_1 that the control input u(t) will influence the output k steps ahead, i.e. y(t+k). By using the identity

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-k}G(z^{-1}) \qquad (c2)$$

where $F(z^{-1})$ and $G(z^{-1})$ are polynomials of degrees k-1 and n-1 (and defined by eqns. 4.5), $C(z^{-1})$ can be eliminated from equation c_1 giving

$$y(t+k) = \frac{B(z^{-1})}{A(z^{-1})}u(t) + F(z^{-1})e(t+k) + \frac{G(z^{-1})}{A(z^{-1})} \qquad (c3)$$

Replacing e(t) from the above equation, again using equation c1, thus gives

$$y(t+k) = F(z^{-1}) + e(t+k) + \frac{G(z^{-1})}{C(z^{-1})}y(t) - \frac{G(z^{-1})B(z^{-1})}{A(z^{-1})C(z^{-1})}u(t-k) + \frac{B(z^{-1})}{A(z^{-1})}u(t) \qquad (c4)$$

The equation reduces to

$$y(t+k) = F(z^{-1})e(t+k) + \frac{G(z^{-1})}{C(z^{-1})}y(t) + \frac{B(z^{-1})F(z^{-1})}{C(z^{-1})}u(t) \qquad (c5)$$

By taking expectations

$$E\{y(t+k)^2\} = E\{[F(z^{-1})e(t+k)]^2\} + E\{[\frac{G(z^{-1})}{C(z^{-1})}\, y(t) + \frac{B(z^{-1})F(z^{-1})}{C(z^{-1})}\, u(t)]^2\} \qquad (c6)$$

as $e(t+1)$, $e(t+2)$, ... $e(t+k)$ are independent of $y(t)$, $u(t-1)$, ... and $u(t-1)$, $u(t-2)$, ...

For $E\{y(t+k)^2\}$ to be a minimum

$$\frac{G(z^{-1})}{C(z^{-1})}\, y(t) = -\frac{B(z^{-1})F(z^{-1})}{C(z^{-1})}\, u(t) \qquad (c7)$$

while the control error is given by

$$E\{y(t+k)^2\} \geq \sigma^2(1 + f_1^2 + \ldots\ldots f_{k-1}^2) \qquad (c8)$$

The minimum variance control law, from equation c7, is given by

$$u(t) = -\frac{G(z^{-1})}{B(z^{-1})F(z^{-1})}\, y(t) \qquad (c9)$$

APPENDIX D

THE RECURSIVE LEAST-SQUARES ALGORITHM

Consider the system model

$$y(t) = \phi(t)\theta + e(t) \qquad (d_1)$$

where $\phi(t)$ is the measurement vector containing the system input and outputs, θ is a vector of parameters and $e(t)$ is a Gaussian white noise sequence distributed as $N(0,\sigma^2)$. The vectors $\phi(t)$ and θ are as defined in equations 3.27.

The least-squares estimate of θ corresponding to the 1-step ahead prediction for a sample size N is obtained by minimising V_{LS} given by

$$V_{LS} = \sum_{t=1}^{N} (y(t) - \phi(t-1)\theta)^2 \qquad (d_2)$$

To minimise V_{LS}, the first derivative with respect to the parameters θ is equated to zero. This gives the 'normal' equations, i.e.

$$\nabla_\theta V_{LS} = 0 \qquad (d_3)$$

giving

$$\sum_{t=1}^{N} \phi^T(t-1)[y(t) - \phi(t-1)\theta(N)] = 0 \qquad (d_4)$$

so that

$$\theta(N) = [\sum_{t=1}^{N} \phi^T(t-1)\phi(t-1)]^{-1} \sum_{t=1}^{N} \phi^T(t-1)y(t) \qquad (d_5)$$

by replacing the summation terms of equation d_5 by the matrix $P(t)$ and the vector $B(t)$, the solution for θ at time t is given by

$$\hat{\theta}(t) = P(t)B(t) \qquad (d_6)$$

If $\hat{\theta}$ is to be computed at every time-step a recursive solution

avoiding matrix inversion can be used to obtain P(t). The following matrix inversion lemma is used.

$$P^{-1}(t) = P^{-1}(t-1) + \phi^T(t-1)\phi(t-1) \tag{d7}$$

as

$$\sum_{i=1}^{t} \phi^T(i-1)\phi(i-1) = \sum_{i=1}^{t-1} \phi^T(i-1)\phi(i-1) + \phi^T(t-1)\phi(t-1) \tag{d8}$$

and similarly

$$B(t) = B(t-1) + \phi^T(t-1)y(t) \tag{d9}$$

Equation d7 is post-multiplied by P(t-1) and pre-multiplied by P(t), such that

$$P(t-1) = P(t) + P(t)\phi^T(t-1)\phi(t-1)P(t-1) \tag{d10}$$

Post-multiplying equation d7 by $\phi^T(t-1)$ gives

$$P(t-1)\phi^T(t-1) = P(t)\phi^T(t-1) + P(t)\phi^T(t-1)\phi(t-1)P(t-1)\phi^T(t-1)$$

$$= P(t)\phi^T(t-1)[1+\phi(t-1)P(t-1)\phi^T(t-1)] \tag{d11}$$

Post-multiplying equation d11 by $[1+\phi(t-1)P(t-1)\phi^T(t-1)]^{-1}\phi(t-1)P(t-1)$ results in

$$P(t-1)\phi^T(t-1)[1 + \phi(t-1)P(t-1)\phi^T(t-1)]^{-1}\phi(t-1)P(t-1)$$

$$= P(t)\phi^T(t-1)\phi(t-1)P(t-1) \tag{d12}$$

Equations d10 and d11 give the updating equation for P(t) as

$$P(t) = P(t-1) - \frac{P(t-1)\phi^T(t-1)\phi(t-1)P(t-1)}{1 + \phi(t-1)P(t-1)\phi^T(t-1)} \tag{d13}$$

Now, from equation d10

$$P(t) = P(t-1) - P(t)\phi^{T}(t-1)\phi(t-1)P(t-1) \qquad (d14)$$

Post-multiplying equation d14 by $P(t-1)^{-1}$ gives

$$P(t)P(t-1)^{-1} = 1 - P(t)\phi^{T}(t-1)\phi(t-1) \qquad (d15)$$

whence, from equation d6,

$$B(t) = P(t)^{-1}\hat{\theta}(t) \qquad (d16)$$

From equations d9 and d16,

$$P(t)^{-1}\hat{\theta}(t) = P(t-1)^{-1}\hat{\theta}(t-1) + \phi^{T}(t-1)y(t) \qquad (d17)$$

which, on pre-multiplying by P(t) gives

$$\hat{\theta}(t) = P(t)P(t-1)^{-1}\hat{\theta}(t-1) + P(t)\phi^{T}(t-1)y(t) \qquad (d18)$$

From equations d15 and d18,

$$\hat{\theta}(t) = \hat{\theta}(t-1) - P(t)\phi^{T}(t-1)\phi(t-1)\hat{\theta}(t-1) + P(t)\phi^{T}(t-1)y(t) \qquad (d19)$$

This gives the updating equation for $\hat{\theta}(t)$, i.e.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\phi^{T}(t-1)\,[\,y(t) - \phi(t-1)\theta(t-1)] \qquad (d20)$$

APPENDIX E

ESTIMATION OF THE AUTO AND CROSS-CORRELATION FUNCTIONS

The autocorrelation function

An important guide to the properties of a time-series $y(t)$ is provided by the covariance between $y(t)$ and its value separated by τ time steps, $y(t+\tau)$. A coefficient of this function is known as the lag τ autocovariance function, and is defined as

$$\gamma_{yy}(\tau) = E\{y(t) - y)(y(t+\tau) - y)\} \qquad (e1)$$

where y is the mean value of $y(t)$. The magnitude of the autocovariance coefficients however depends upon the relative magnitude of $y(t)$, so that often the autocovariance function is standardized, to give the autocorrelation function $\rho_{yy}(\tau)$. Assuming stationarity, the variance $\sigma_y^2 = \gamma_{yy}(0)$ is the same at time t as at time $t+\tau$ so that

$$\rho_{yy}(\tau) = \frac{\gamma_{yy}(\tau)}{\gamma_{yy}(0)} \qquad (e2)$$

The autocorrelation function is symmetric about $\tau = 0$ and $\rho_{yy}(0) = 1$.

An estimate of the above autocovariance function for a sample size N can be given by

$$c_{yy}(\tau) = \frac{1}{N-1} \sum_{t=1}^{N-\tau} (y(t) - y)(y(t+\tau) - y) \qquad (e3)$$

The estimate of the autocorrelation function is then given by

$$r_{yy}(\tau) = \frac{c_{yy}(\tau)}{c_{yy}(0)} \qquad (e4)$$

If $y(1), \ldots y(N)$ are independent and identically distributed random variables with an arbitrary mean, it can be shown that

$$E\{r_{yy}^2(\tau)\} \simeq -1/N \qquad (e5)$$

and

$$E\{r_{yy}^2(\tau)\} \simeq 1/N \qquad (e6)$$

and that for large N , $r_{yy}(\tau)$ is (asymptotically) normally distributed under weak conditions (Kendall and Stuart, 1966). Thus the approximate 95% confidence limits for $r_{yy}(\tau)$ are given by $\pm 1/N \pm 1.95/\sqrt{N}$, which is often further approximated to $\pm 2/\sqrt{N}$. Observed values of $r_{yy}(\tau)$ falling outside these limits are 'significantly' different from zero at the 5% level.

The cross-correlation function

The lag τ cross-covariance function $\gamma_{xy}(\tau)$ is similar to the auto-covariance function, except that the relationship between two time series $x(t)$ and $y(t)$ is being considered. The cross-covariance coefficient between $x(t)$ and $y(t+\tau)$ is given by

$$\gamma_{xy}(\tau) = E\{(x(t) - \bar{x})(y(t+\tau) - \bar{y})\} \qquad (e7)$$

The cross-covariance coefficients depend on the relative magnitude of $x(t)$ and $y(t)$ and so for interpretation purposes (as for the autocorrelation function), the cross-correlation function $\rho_{xy}(\tau)$ is usually employed, where

$$\rho_{xy}(\tau) = \gamma_{xy}(\tau)/\sqrt{(\gamma_{xx}(0)\gamma_{yy}(0))} \qquad (e8)$$

Points to note include

$$\rho_{xy}(\tau) \neq \rho_{xy}(-\tau)$$

$$\rho_{xy}(\tau) = \rho_{yx}(\tau)$$

and

$$|\rho_{xy}(\tau)| \leq 1 \qquad (e9)$$

Applying the appropriate suffix notation to equation e3, an estimate

of the cross-covariance function is given by

$$c_{xy}(\tau) = \frac{1}{N-1} \sum_{t=1}^{N-\tau} (x(t) - \bar{x})(y(t+\tau) - \bar{y}) \qquad (e10)$$

and the estimated cross-correlation function is given by

$$r_{xy}(\tau) = c_{xy}(\tau)/\sqrt{(c_{xx}(0)c_{yy}(0))} \qquad (e11)$$

While the above estimator is consistent (Chatfield, 1975) it can also be shown that the variance of the estimates depend on the auto-correlations of the two series themselves (Box and Jenkins, 1970). For two uncorrelated white noise series however,

$$E\{r_{xy}(\tau)\} \simeq 0$$

and

$$E\{r_{xy}^2(\tau)\} \simeq 1/N \qquad (e12)$$

so that cross-correlation coefficient values falling outside the interval $\pm 1.96/\sqrt{N}$ are significantly different from zero at the 5% level.

REFERENCES

AMOROCHO, J.,(1963), Measures of the linearity of hydrologic systems, Journal of Geophys. Res., Vol. 68, No. 8, pp. 2237-2249.

AMOROCHO, J., and BRANDSTETTER, (1971), Determination of non-linear functional response functions in rainfall-runoff processes, Water Resources Research, Vol. 7, No. 3, pp. 1087-1101.

APPLEBY, F.V., (1965), Dissertation on engineering hydrology with special reference to recession, submitted for the Diploma of Imperial College, London.

APPLEBY, F.V., (1977), personal communication.

ASKEW, A.J., (1974), Chance-constrained dynamic programming and the optimisation of water resource systems, Water Resources Research, Vol. 10, No. 6, pp. 1099-1106.

ÅSTROM, K.J., (1970), Introduction to Stochastic Control Theory, New York, Academic Press.

ÅSTROM, K.J., (1974), A self-tuning parameter estimator, Dept. of Computing and Control, Publ. No. 74/55, Imperial College of Science and Technology, London.

ÅSTROM, K.J., and T. BOHLIN, (1966), Numerical identification of linear dynamic systems from normal operating records, Proc. IFAC Symp. on the Theory of Self Adaptive Systems, Teddington, England. Also in Theory of Self-Adaptive Control Systems (Ed. P.H. Hamond), New York, Plenum Press.

ÅSTROM, K.J., and P. EYKHOFF, 1971, System Identification - A Survey. Automatica, Vol. 7, pp. 123-162.

ÅSTROM, K.J., and B. WITTENMARK, (1971), Problems of identification and control, Journal Math. Analysis Applic., Vol. 34, pp. 90-113.

ÅSTROM, K.J., and B. WITTENMARK, (1973), On self-tuning regulators, Automatica, Vol. 9, pp. 185-199.

ATHANS, M., and P. FALB, (1966), Optimal Control, New York, McGraw-Hill.

BAR-SHALOM, Y., and E. TSE, (1974), Dual effect, certainty equivalence and separation in stochastic control, IEEE Trans. Automatic Control, Vol. AC-19, p. 494.

BECKER, L., W. W-G. YEH, D. FULTS and D. SPARKS, (1976), Operational models for Central Valley Project, Journal of the Water Resources Planning and Management Division, ASCE, Vol. 102, No. WR1, pp. 101-115.

BIDWELL, V.J., (1971), Regression analysis of non linear catchment systems, Water Resources Research, Vol. 7, No. 5, pp. 1118-1126.

BONEH, A., and M.H. DISKIN, (1973), Determination of optimal kernals for second order stationary surface runoff systems, Wat. Resour. Res., Vol. 9, pp. 311-325.

BORISSON, U., and R. SYDING, (1974), Self-tuning control of an ore-crusher, Proc. IFAC Symposium on Stochastic Control, Budapest, pp. 491-497.

BORISSON, U., and B. WITTENMARK, (1973), Moisture control of a paper machine - an application of a self-tuning regulator, Rept. 7337, Dept. Automatic Control, Lund Institute of Technology, Lund, Sweden.

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

BOX, G.E.P., and D.A. PIERCE, (1970), Distribution of residual auto-correlations in autoregressive-integrated moving average time-series models, Journal of the American Statistical Association, Vol. 65, No. 332, pp. 1509-1529.

BRAS, R.F. and I. RODRIGUEZ-ITURBE, (1976), Rainfall network design for runoff prediction, Water Resourc. Res., 12(6), pp. 1197-1208.

BRYSON, A.E., and Y.C. HO, (1969), Applied Optimal Control, Waltham, Mass., Publishing Co..

CENTRAL WATER PLANNING UNIT, (1977), Dee Weather Radar and Real-Time Hydrological Forecasting Project, Report by the Steering Committee, CWPU, England.

CHATFIELD, C. (1975), The Analysis of Time Series: Theory and Practice, London, Chapman and Hall.

CHATFIELD, C., and D.L. PROTHERO, (1973), Box-Jenkins seasonal forecasting: problems in a case study, Journal of the Royal Statist. Soc.,Series A, Vol. 136, pp. 295-336.

CHOW, V.T., (1959), Open-channel hydraulics, New York, McGraw Hill.

CLARKE, D.W., (1967), Generalized least squares estimation of the parameters of a dynamic model, IFAC Symposium on Identification in Automatic Control Systems, Prague.

CLARKE, D.W., S.N. COPE, and P.J. GAWTHROP, (1975), Feasibility study of the application of microprocessors to self-tuning controllers, Report No. 1137/75, Engng. Lab., Oxford University.

CLARKE, D.W., and P.J. GAWTHROP, (1975), Self-tuning controller, Proc. Institution of Electrical Engineers, Vol. 122, No. 9, pp. 929-934.

CLARKE, R.T., (1971), The representation of a short period of experimental catchment data by a linear stochastic difference equation, Warsaw Symposium on Mathematical Models in Hydrology, July, 1971, 15pp.

CLARKE,R.T., (1973), Mathematical models in hydrology, FAO, Drainage and Irrigation paper No. 19, Rome.

COLE, J.A., (1975), Assessment of surface water sources, Chap. 10, pp. 113-125, in Engineering Hydrology Today, Inst. Civ. Engnrs, London.

COLE, J.A., (1976), On-line forecasting for a regulated river - examples from the Welsh Dee, incorporating weather radar data, IIASA/WMO Workshop on the Recent Development in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

CRAWFORD, N.M., and R.K. LINSLEY, (1966), Digital simulation in hydrology: Stanford Watershed Model IV, Dept. Civ. Eng., Tech. Rep. 39, Stanford University.

DAVIES, N., C.M. TRIGGS and P. NEWBOLD, (1977), Significance levels of the Box-Pierce portmanteau statistic in finite samples, Biometrica, Vol. 64, No. 3, pp 517-522.

DISKIN, M.H., and A. BONEH, (1972), Properties of the kernals for time-invariant, initially relaxed, second order, surface runoff systems, Journal of Hydrology, Vol. 17, pp. 115-141.

DOOGE, J.C.I., (1973), Linear Theory of Hydrologic Systems, Tech. Bull. 1468, Agricultural Research Serice, U.S. Dept. Agriculture, Washington.

DORFMAN, R., (1962), Mathematical Models: The Multi-structure Approach in Design of Water Resource Systems, A. Maass, et al., eds., Cambridge, Mass, Harvard University Press.

DUDLEY, N.J., and BURT, O.R., (1973), Stochastic reservoir management and system design for irrigation, Water Resources Research, Vol. 9, No. 3, pp. 507-522.

DUONG, N., C.B. WINN, and G.R. JOHNSON, (1975), Modern concepts in hydrology, IEEE Trans. Systems, Man, and Cybernetics, Vol. SMC-5, No. 1, pp. 46-53.

FREEZE, R.A., (1972), Role of subsurface flow in generating surface runoff: 1, Baseflow contributions to Channel flow, Water Resources Research, Vol. 8, No. 3, pp. 601-623.

FULTS, D.M. and L.F. HANCOCK, (1972), Optimal operations model for Shasta-Trinity system, Journal of the Hydraulics Division, ASCE, Vol. 98, No. HY9, pp. 1497-1514.

FULTS, D.M., and L.F. HANCOCK, (1974), Managing systems operations - Central Valley Project, Journal of the Hydraulics Division, ASCE, Vol. 100, No. HY10, pp. 1419-1427.

FELDBAUM, A.A., (1960), Dual control theory, Automn. Remote Control, Vol. 21, pp. 874-880 and pp. 1033-1039.

GABLINGER, M. and D.P. LOUCKS, (1970), Markov models for flow regulation, Journal of the Hydraulics Division, ASCE, Vol. 96, No. HY1, pp. 165-181.

GANENDRA, T., (1976), A self-tuning predictor applied to river flow forecasting, IIASA/WMO Workshop on the Recent Developments in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

GANENDRA, T., (1977), A self-tuning controller applied to river regulation, IFIP Working Conference on Modelling and Simulation of Land, Air and Water Resources Systems, August 30-September 2, 1977, Ghent.

GELB, A., (1974), Applied Optimal Estimation, Cambridge, Mass., MIT Press.

HARRIS, C.J., (1976), Problems in system identification and control, Institute of Mathematics and its Applications, Bull., Vol. 12, pp. 139-150.

HEIDARI, M., V.T. CHOW, P.V. KOKOTOVIĆ and D.D. MEREDITH, (1971), Differential dynamic programming approach to water resources system optimisation, Water Resources Research, Vol. 7, No. 2, pp. 273-282.

HENDERSON, F.M., (1966), Open Channel Flow, New York, Macmillan.

HINO, M., (1970), Runoff forecasts by linear predictive filter, ASCE, Journal of Hydraulics Division, Hy 3, pp. 681-701.

HINO, M., (1973), On-line prediction of hydrological systems, Proc. XVth Conference IAHR, Istanbul, pp. 121-129.

HINO, M., T. SUKIGARA, and H. KIKKAWA, (1971), Nonlinear runoff kernals of hydrologic systems, in V. Yevjevich etc., Systems Approach to Hydrology, Water Res. Publ., Fort Collins.

HUFSCHMIDT, M.M. and M.B. FIERING, (1966), Simulation Technique for Design of Water Resource Systems, Cambridge, Mass. Harvard Univ. Press.

IEEE, (1972), Standard Dictionary of Electrical and Electronics Terms, New York, John Wiley.

IFAC, (1973), Identification and System Parameter Estimation, Proc. 3rd IFAC Symposium, The Hague/Delft, 12-15 June, 1973, ed. P. Eykhoff, New York, John Wiley.

JAMIESON, P.G., (1972), River Dee research program. I. Operating multi-purpose reservoir systems for water supply and flood alleviation, Water Resources Research, Vol. 8, No. 4, pp 899-910.

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

KALMAN, R.E., (1960), A new approach to linear filtering and prediction problems, Trans. ASME, J. Basic Engng., Vol. 82D, pp. 35-45.

KALMAN, R.E. and R.S. BUCY, (1961), New results in linear filtering and prediction theory, Trans. ASME, J. Basic Engng., Vol. 83D, pp. 95-108.

KASHYAP, R.L., and A.R. RAO, (1973), Real-time recursive prediction of river flows, Automatica, Vol. 9, pp. 175-183.

KATAYAMA, T., (1976), Application of maximum likelihood identification to a river flow prediction, IIASA/WMO Workshop on the Recent Developments in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

KENDALL, M.A., and A. STUART, (1961), The Advanced Theory of Statistics, Vol. 2, London, Griffin.

LAMBERT, A.O., (1972), Catchment models based on ISO-functions, Journal of the Institution of Water Engrs., Vol. 26, No. 8, pp. 413-422.

LAMBERT, A.O. and R.J. CAMERON, (1975), Control rules for long and short term objectives, Symp. on Weather Radar and Water Managment, Water Research Centre and Royal Radar Establishment, 23pp.

LJUNG, L., and B. WITTENMARK, (1974), Asymptotic property of self-tuning regulators, Rept. 7404, Lund Institute of Technology, Lund, Sweden.

LORENT, B., (1976), River flow prediction and simulation model, IIASA/WMO Workshop on the Recent Developments in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

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, Symp. on Weather Radar and Water Management, Water Research Centre and Royal Radar Establishment, 22 pp.

MANDEVILLE, A.N., (1975), Nonlinear conceptual catchment modelling of isolated storm events, PhD thesis, Univ. of Lancaster, Lancaster, England.

MEHRA, R.K. (1969), Identification of stochastic linear dynamic systems, Eighth IEEE Symp. on Adaptive Processes, Pennsylvania, November, 1969, pp. 6f1-6f5.

MEIER, W.L., Jr., and C.S. BEIGTLER, (1967), An optimization method for branching multi-storage water resource systems, Water Resources Research, Vol. 3, No. 3, pp. 645-652.

MEYER-ZURWEILE, J., (1973), Optimum release strategies for systems of flood protection reservoirs, IAHR Congress, Istanbul.

MOORE, R.J., and G. WEISS, (1976), Real-time parameter estimation of a non-linear catchment model using extended Kalman Filters, IIASA/WMO Workshop on the Recent Developments in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

MORAN, P.A.P., (1959), The Theory of Storage, London, Methuen.

MURPHY, A.H., (1976), Subjective quantification of uncertainty in real-time weather forecasts in the United States, IIASA/WMO Workshop on the Recent Developments in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

NATALE, L., and E. TODINI, (1976), A constrained parameter estimation technique for linear models in hydrology, Proc. Workshop on Mathematical Models in Hydrology, Centro Scientifico, IBM Italia, Pisa, December 1974, London, John Wiley.

OAKES, D.B., and J.M.A. PONTIN, (1976), Mathematical modelling of a chalk aquifer, Tech. Rep. TR24, Water Research Centre, Medmenham, England.

O'CONNELL, P.E., et al., (1977), Methods for evaluating the UK raingauge network, Institute of Hydrology, Wallingford, Oxon, 303 pp.

OZAKI, T., (1977), On the order determination of ARIMA models, Journal of the Royal Statist. Soc., Series C, Vol. 26, p. 290.

PRASAD, R., (1967), Non-linear simulation of a regional hydrologic system, PhD thesis, Univ. of Illinois, Urbana.

PRICE, R.K., (1973), Flood routing methods for British rivers, Proc. Institution of Civil Engineers, part 2, Vol. 55, pp. 913-930.

PRICE, R.K., (1975a), Mathematical model for river flows, 1 theoretical development, Rept. INT 127, Hydraulics Research Station, Wallingford, England.

PRICE, R.K., (1975b), A flow routing model for the River Dee, Rept. EX 712, Hydraulics Research Station, Wallingford, England.

QUIMPO, R.G., (1975), Stochastic identification of nonlinear hydrologic systems, IAHS Symp. on the Application of Mathematical Models in Hydrology and Water Resources Systems.

REMSON, I., G.M. HORNBERGER, and F.J. MOLTZ, (1971), Numerical methods in subsurface hydrology, New York, Wiley.

SHULTZ, G.A., and E.J. PLATE, (1976), Developing optimal operating rules for flood protection reservoirs, Journal of Hydrology, Vol. 28, pp. 246-264.

SODERSTROM, T., L. LJUNG and I. GUSTAVSSON, (1974), A comparative study of recursive identification methods, Rept. 7427, Dept. Automatic Control, Lund Institute of Technology, Lund, Sweden.

SZOLLOSI-NAGY, A., (1976), Introductory remarks on the state space modelling of water resource systems, Research Memoranda RM-76-73, IIASA, Laxenburg, Austria.

THAMES CONSERVANCY, (1971), Report on the Lambourn Valley pilot scheme, 1967-1969, Reading, England.

TODINI, E., and D. BOUILOOT, (1975), A rainfall-runoff Kalman filter model, in simulation in Water Resources, Ed. G.C. Vansteenkiste, Amsterdam, North-Holland.

TODINI, E., A. SZÖLLÖSI-NAGY, and E.F. WOOD, (1976), Adaptive state/parameter estimation algorithms for real-time hydrologic forecasting; A case study, IIASA/WMO Workshop on the Recent Developments in Real-Time Forecasting/Control of Water Resources Systems, October 18-20, 1976, Laxenburg, Austria.

TODINI, E., and J. WALLIS, (1977), Using CLS for daily of longer period rainfall-runoff modelling, Proc. Workshop on Mathematical Models in Hydrology, Centro Scientifico, IBM Italia, Pisa, December 1974, London, John Wiley.

TOYODA, J., N. TORIUMI, and Y. INONE, (1969), An adaptive predictor of river flows for on-line control of water resource systems, Automatica, Vol. 5, pp 175-181.

YEVJEVICH, V.M., (1974), Regression and correlation analysis, Section 8, Part II: In V.T. Chow, ed., Handbook of Applied Hydrology, McGraw Hill, New York.

YOUNG, G.K., (1967), Finding reservoir operating rules, Journal of Hydraulics Division, ASCE, Vol. 93, No. HY6, pp. 297-321.

YOUNG, P.C., (1968), The use of linear regression and related procedures for the identification of dynamic processes, Proc. Seventh IEEE Symposium on Adaptive Sytems, New York, IEEE.

YOUNG, P.C., (1970), An instrumental variable method for real-time identification of a noisy process, Automatica, Vol. 6, pp. 271-287.

YOUNG, P.C., S.H. SHELLSWELL, and C.G. NEETHLING, (1971), A recursive approach to time-series analysis, Dept. of Engng., Univ. of Cambridge, Rep. No. CUED/B-CONTROL/TR16; 60pp.

YOUNG, P.C., and P. WHITEHEAD, (1977), A recursive approach to time-series analysis for multivariate systems, Int. Journal of Control, Vol. 25, No. 3, pp. 457-482.

ZAND, S.M., and J.A. HARDER, (1973), Application of nonlinear system identification to the Lower Mekong River, Southeast Asia, Water Resources Research, Vol. 9, No. 2, pp. 290-297.

WELLSTEAD, P.E., (1976), A review of applications of self-tuning systems, Institution of Electrical Engineering, Colloqiuium of Self-Tuning and Adaptive Systems, 17 May, 1976, IEE Digest No. 1976/35, pp. 3/1-3/2.

WITTENMARK. B. (1973), A self-tuning regulator, Rept. 7311, Dept. Automatic Control, Lund Institute of Technology, Lund, Sweden.

WITTENMARK, B., (1974), A self-tuning predictor, IEEE Trans. on Automatic Control, Vol. AC-19, No. 6, pp 848-851.

WITTENMARK, B., (1975),Stochastic control methods: a survey, Int. Journal of Control, Vol. 21, No. 5, pp. 705-730.

WITSENHAUSEN, H.S., (1971), Separation of estimation and control for discrete time systems, Proc. Inst. of Electrical and Electronic Engrs., Vol. 59, p. 1557.

WHITEHEAD, P.G. and YOUNG, P.C., (1974), A dynamic stochastic model for water quality in part of the Bedford-Ouse river system, Proc. IFIP Working Conference on Modelling and Simulation of Water Resource Systems, Ghent.

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

WOOD, E.F., (1979), On-line flow routing, Proc. Extended workshop on real-time hydrological forecasting and control, July 1977, Institute of Hydrology, Wallingford, England (to be published).

WORLD METEOROLOGICAL ORGANIZATION, (1975), Intercomparison of conceptual models used in operational hydrology forecasting, Operational hydrology Rep. 7, WMO No. 429, Geneva.