RM-75-22

AN ADAPTIVE IDENTIFICATION AND PREDICTION ALGORITHM FOR THE REAL-TIME FORECASTING OF HYDROLOGIC TIME SERIES

A. Szöllösi-Nagy

May 1975

An Adaptive Identification and Prediction Algorithm for the Real-Time Forecasting of Hydrologic Time Series*

A. Szöllösi-Nagy**

Abstract

In order to achieve the effective control of water resources systems, one must know the future behaviour of the inputs to that particular system. Because of the uncertainties inherent in water resources processes the prediction algorithm, to be constructed, should include stochastic elements, too. Moreover, the algorithm should be recursive to avoid cumbersome computations and to be able for real-time forecasting.

In the paper we present a method which is applicable for both linear and nonlinear hydrologic systems having not completely time-invariant properties. The algorithms are based on the state space description of the processes involved and utilize the Kalman stochastic filtering technique. Due to the unknown nature of noise processes, the basic algorithms were changed to be adaptive. Using the algorithms the joint handling of water quantity and quality data becomes feasible.

Introduction

In order to achieve effective control of water resource systems, one must know the future behaviour of the inputs to the particular systems. This is the ancient challenging task of the human being, because man's encounter with the prediction problem is as old as civilization itself.

The first successful scientific attack dates back to the early 1940's when Wiener and Kolmogorov solved the problem independently for the case of linear dependent stationary processes, which requires solution of the Wiener-Hopf equation. In 1966, Eagleson et al. [4] were the first to apply this technique for runoff prediction. Their paper initiated the avalanche of articles dealing with the various tricky modifications of Wiener's procedure for practical hydrologic forecasting

---

*Submitted to the International Symposium and Workshops on the Application of Mathematical Models in Hydrology and Water Resources Systems, Bratislava, September 1975.

**On leave from the Research Institute for Water Resources Development, Budapest.

problems (see e.g. [6], [22] and [24]). The classical Wiener-Hopf technique, however, has some serious drawbacks: it can be applied only for strictly linear and time invariant systems with stationary input/output processes. Some efforts had been made to extend the theory (see e.g. [7]); due to the computational burden, however, they practically failed. Additional difficulties were raised by the use of spectral factorization, and from a practical point of view, by the necessity of using a relatively large computer to store all the data.

Obviously, for real-time operation of water resource systems, small computers are preferable. Hence, our prediction algorithms must be suited for these small computers. But how? The answer is simple: using recursive prediction algorithms in which there is no need to store all the past measurements for the purpose of predicting the future behaviour of the time series in question. Moreover, these algorithms would offer the following advantages:

(1) The treatment of the information of each measurement in a sequential manner allows for on-line implementation (e.g. by means of data acquisition by automatic measurement devices connected in real-time mode with a central processor); and

(2) Time variable parameters and different types of disturbances can easily be treated.

Hence, a suitable prediction scheme should preferably satisfy the following requirements:

(1) it should be mathematically tractable;

(2) it should be easily implemented for relatively small computers;

(3) it should be generally applicable;

(4) it should yield an 'optimum' prediction;

(5) it should be adaptable to varying environmental conditions;

(6) it should yield an acceptable convergence.

The hydrologic prediction schemes used nowadays unfortunately generally fail to meet one or more of these requirements.

In the early 1960's R.E. Kalman [10] developed an optimal sequential estimation technique, usually referred to as the Kalman filter, which has proved extremely useful in dealing with the description of stochastically excited dynamic systems.*

---

*In this respect, the reader is referred to the extremely rich literature, e.g. [1], [16], [18] and [20].

The Kalman filtering technique is based on the state space, time domain formulation of the processes involved, and with slight modifications offers a procedure as a candidate for satisfying the above requirements of a suitable hydrologic prediction scheme.

In this paper we briefly outline the basis of the Kalman filtering technique and propose a simple state-space-based model for the recursive adaptive estimation of the impulse response of a hydrologic system. Discrete time models are considered. The proposed algorithms can be applied to slightly non-linear and time varying systems using a proper moving data window. Having obtained the optimal time varying impulse response(s), the well-known techniques can be used for predicting the output process(es).

State Space Representation of Hydrologic Processes

Consider a water resource system (Figure 1), the behaviour which, evolving on the discrete-time set $T = \{t_k: k = 0, 1,2,...\}$, can be described by

$$\underline{x}(t_{k+1}) = \mathcal{A}[\underline{x}(t_k), \underline{u}(t_k), \underline{w}(t_k)] \quad (1)$$

$$\underline{z}(t_k) = T[\underline{x}(t_k), \underline{v}(t_k)] \quad (2)$$

where $\underline{x}(t_k)$ is the n-vector of the states of the system at the discrete time $t_k \in T$; $\underline{u}(t_k)$ is the s-vector of control variables or known system inputs; $\underline{w}(t_k)$ is the r-vector of uncertain disturbances 'driving' the system; $\underline{z}(t_k)$ is the m-vector of measurements on the system; $\underline{v}(t_k)$ is the m-vector of uncertain disturbances corrupting the observations; and $\mathcal{A}$ and T are certain functionals characterizing the properties of that particular system. Eq. (1) is called the state equation, and Eq. (2) the measurement equation (as the measurement noise $\underline{v}(\cdot)$ is sometimes referred to as measurement uncertainty, while some components of $\underline{w}(\cdot)$, or the entire $\underline{w}(\cdot)$ itself, might be referred to as model uncertainty). Considering the simple example of a reservoir system consisting of n reservoirs, $\underline{x}(t_k)$ might be sought as a vector composed of the values of the amount of stored water of each reservoir at time t_k; $\underline{u}(t_k)$ as a vector of water releases (control variables); $\underline{w}(t_k)$ as the vector of natural (uncontrolled stochastic) inflows to the reservoirs, and $\underline{z}(t_k)$ as the vector of measured outflows from the reservoirs.

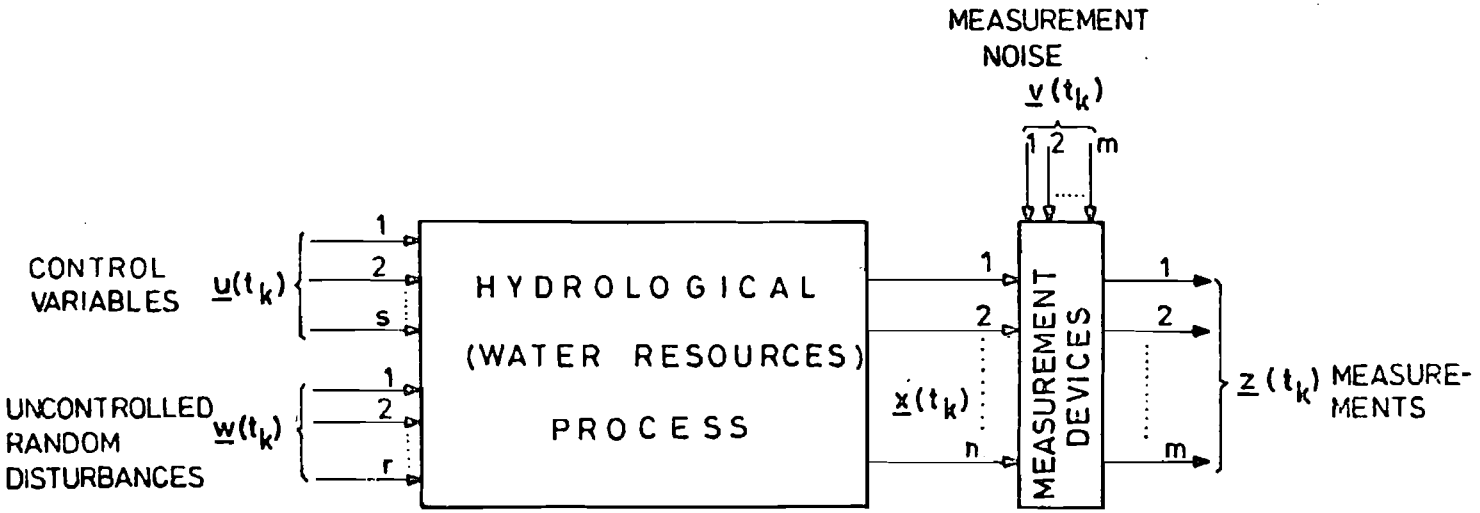

FIGURE 1. HYDROLOGIC PROCESS IDENTIFICATION

In this case, the state vector $\underline{x}(\cdot)$ refers to actual physical states, namely to the amount of stored water in the system; but, as will be shown later, it is not at all necessary to associate the state vector with 'physical' states. In other words, one can choose amongst different types of state variables to describe the same process.

Dealing with the above type of models, one must determine the structure of the system, in other words, the functionals $\mathcal{A}$ and T. This is the problem of system identification [18]. Having identified the system, the next step is to find the 'best' prediction of the state vector (which may sometimes contain the output process depending upon the choice of the state variables) $\ell > 0$ time periods ahead, based upon knowledge of the measurement on the system at t_k

$$\hat{\underline{x}}(t_{k+\ell} \mid t_k) = \Omega[\underline{x}(t_k), \underline{z}(t_k)] \quad , \tag{3}$$

where ℓ is the lead time of the prediction, Ω denotes the prediction algorithms, and the circumflex refers to the predicted (estimated) value. Obviously, the goodness of prediction must be evaluated through a given loss (cost) function, $L(\cdot)$. Now, the prediction problem can be formulated as follows: given the set of measurements $\underset{\sim}{Z}_k = \{\underline{z}(t_i): i = 1,2,\ldots,k\}$, find and estimate $\hat{\underline{x}}(t_{k+\ell} \mid \underset{\sim}{Z}_k)$ of $\underline{x}(t_{k+\ell})$, $\ell > 0$, subject to the condition that this estimation (prediction) minimize the chosen loss function.

We mention in advance that the identification and prediction algorithms will be imbedded here into the same general adaptive algorithms.

In this paper we consider linear lumped parameter water resource systems where $\mathcal{A}$ and T are linear functionals. In other words, the processes are assumed to be represented by the linear vector difference equation

$$\underline{x}(t_{k+1}) = \underline{\Phi}_k \underline{x}(t_k) + \underline{\Gamma}_k \underline{w}(t_k) + \underline{\Lambda}_k \underline{u}(t_k) \tag{4}$$

where, beyond the variables already defined, $\underline{\Phi}_k \triangleq \underline{\Phi}(t_{k+1}, t_k)$ is the n x n nonsingular state transition matrix which, in the case of an unforced system, maps the state vector from time t_k to the state vector at time t_{k+1}; $\underline{\Gamma}_k \triangleq \underline{\Gamma}(t_k)$ is the n x r system noise coefficient matrix, and $\underline{\Lambda}_k \triangleq \underline{\Lambda}(t_k)$ is the n x s control matrix. Note that in general these matrices are time varying.

As for the stochastic model uncertainty $w(t_k)$, without loss of generality it is assumed to be a Gaussian white noise sequence with zero mean

$$\mathcal{E}\{\underline{w}(t_k)\} = \underline{0}$$

and covariance matrix

$$\mathcal{E}\{\underline{w}(t_k)\ \underline{w}^T(t_j)\} = \underline{Q}_k\ \delta_{kj}$$

where $\mathcal{E}\{\cdot\}$ denotes the expected value operator, T the matrix transposition, δ_{kj} the Kronecker delta, and $\underline{Q}_k \triangleq \underline{Q}(t_{k+1}, t_k)$ the r x r noise covariance matrix, i.e. $\underline{w}(t_k) \sim N(\underline{0}, \underline{Q}_k)$. Also, it is assumed that the measurement equation (cf. Eq. (2)) is linear and has the form

$$\underline{z}(t_k) = \underline{H}_k\ \underline{x}(t_k) + \underline{v}(t_k) \quad . \tag{5}$$

Here $\underline{H}_k \triangleq \underline{H}(t_k)$ is the m x n measurement matrix, and the measurement uncertainty $\underline{v}(t_k)$ is also assumed to be a Gaussian white sequence with zero mean

$$\mathcal{E}\{\underline{v}(t_k)\} = \underline{0}$$

and covariance matrix

$$\mathcal{E}\{\underline{v}(t_k)\ \underline{v}^T(t_j)\} = \underline{R}_k\ \delta_{kj}$$

where the m x m noise covariance matrix $\underline{R}_k \triangleq \underline{R}(t_k)$ is assumed to be positive-definite. That is, $\underline{v}(t_k) \sim N(\underline{0}, \underline{R}_k)$. Moreover, it is assumed that the noise processes are uncorrelated with one another, i.e.

$$\mathcal{E}\{\underline{w}(t_k)\ \underline{v}^T(t_j)\} = \underline{0} \quad , \qquad \forall\, k,j \quad .$$

Further, we will utilize the separation theorem (see e.g. Bryson and Ho, [2]) which states that, for linear systems with quadratic cost functions and subject to additive white Gaussian noise inputs, the optimum stochastic controller is realized by cascading an optimal estimator (predictor) with a deterministic optimum controller (Figure 2). According to this principle, the optimal stochastic control of a water resource system can be decoupled into two parts. Now, we concentrate on the first problem, the state estimation/prediction problem. Therefore, the terms in Eq. (1) and (4) consisting of the control function $\underline{u}(\cdot)$ will be omitted from now on.

One can argue about the basic assumptions of the noise processes being Gaussian white sequences with known covariance matrices. In particular, it is hard to say that the latter values are known in dealing with hydrologic time series. To overcome this difficulty, an adaptive noise covariance matrix algorithm will be introduced. As for handling 'colored' noises, if the state vector might be properly augmented with the dependent part of the processes, the resulting residual is a white sequence (for details see Porebski [15]).

It is a well-known fact that the autoregressive (AR) models and moving-average (MA) models, or their combinations, the ARMA and ARIMA models, have found a fruitful application area in describing the behaviour of hydrologic time series. There are tremendous amounts of literature to prove this; however, almost each paper offers a different approach for handling the models. It can be shown that all those time series models could be included, as special cases, under the umbrella of the general state space model. As an example, consider the m^{th} dimensional discrete-time autoregressive model p^{th} order, $AR_m(p)$:

$$\underline{y}(t_{k+1}) + \sum_{j=1}^{p} \tilde{\underline{\phi}}_j(t_k)\, \underline{y}(t_{k-j+1}) = \tilde{\underline{\Theta}}(t_k)\, \underline{w}(t_k)$$

$$\underline{z}(t_k) = \underline{y}(t_k)$$

where, beyond the known notations, the matrices $\tilde{\underline{\phi}}_j(\cdot)$ and $\tilde{\underline{\Theta}}(\cdot)$ contain the AR parameters, and the vector $\underline{y}(\cdot) = [y_1(\cdot), y_2(\cdot), \ldots, y_m(\cdot)]^T$ represents m (possibly correlated) water resource processes such as runoff, soil moisture content, water use, BOD, DO, toxic materials in the water, etc., depending upon the objective of the study. For the sake of notational simplicity initial conditions are ignored and the $p = 2$ case is considered here. For the state space representation of the $AR_m(2)$ process we define the state vector as

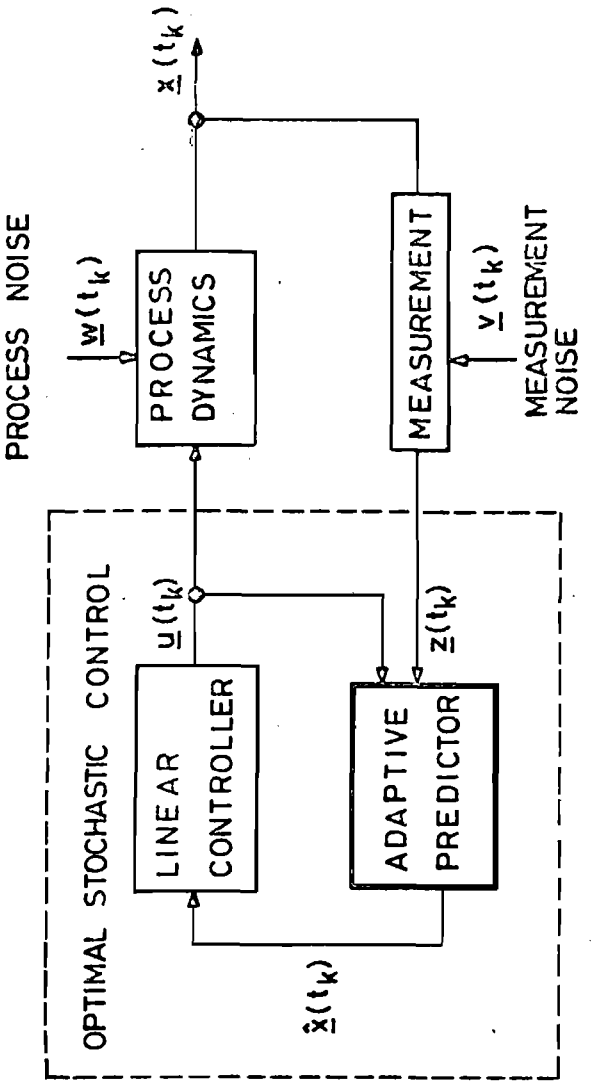

FIGURE 2. SEPARATION PRINCIPLE

$$\underline{x}(t_k) = \begin{bmatrix} \underline{x}_1\ (t_k) \\ \underline{x}_2\ (t_k) \end{bmatrix} = \begin{bmatrix} \underline{y}\ (t_{k-1}) \\ \underline{y}\ (t_k) \end{bmatrix}$$

and, using the AR parameter matrices, define the following matrices:

$$\underline{\Phi}_k = \begin{bmatrix} \underline{O} & \vdots & \underline{I} \\ \cdots\cdots & \vdots & \cdots\cdots \\ -\tilde{\underline{\Phi}}_2(t_k) & \vdots & -\tilde{\underline{\Phi}}_1(t_k) \end{bmatrix}$$

$$\underline{\Gamma}_k = \begin{bmatrix} \underline{O} \\ \cdots\cdots \\ \tilde{\underline{\Theta}}(t_k) \end{bmatrix} \qquad \underline{H} = \begin{bmatrix} \vdots \\ \underline{O} \vdots \underline{I} \\ \vdots \end{bmatrix} \quad .$$

Then the state space model

$$\underline{x}(t_{k+1}) = \underline{\Phi}_k \underline{x}(t_k) + \underline{\Gamma}_k \underline{w}(t_k)$$

$$\underline{z}(t_k) = \underline{H} \underline{x}(t_k)$$

is completely equivalent to the $AR_m(2)$ model. That this is really a special case is seen when the above state space model is compared with Eq. (4) and (5). A similar formulation can be obtained for MA,ARMA, processes. It should be noted again, however, that the above state space formulation of an AR process is not unique, in the sense that if another form is chosen for the state vector the matrices $\underline{\Phi}$, $\underline{\Gamma}$, $\underline{H}$ will change but the input-output behaviour of the system will not. In other words, the choice of a particular set ($\underline{\Phi}$, $\underline{\Gamma}$, $\underline{H}$) corresponds to the choice of a coordinate system [20]. However, the proper choice of the state vector has great significance from the point of view of practical computations on the one hand, and of system controllability and observability on the other [21].

The Adaptive Sequential Prediction Algorithms

Assume that a prior estimate $\hat{\underline{x}}(t_k | t_{k-1})$ of the system state $x(t_k)$ is given at time t_k which is based on previous measurements up to t_{k-1}. Then we seek an updated estimate $\hat{\underline{x}}(t_k | t_k)$

which takes into account the new measurement $\underline{z}(t_k)$ at time t_k. (For the notations and timing see Figure 3.) Consider the updated estimation as being the linear combination of the previous state and the new (noisy) measurement

$$\hat{\underline{x}}(t_k|t_k) = \tilde{\underline{K}}_k \hat{\underline{x}}(t_k|t_{k-1}) + \underline{K}_k \underline{z}(t_k) \quad , \tag{6}$$

where $\tilde{\underline{K}}_k \triangleq \tilde{\underline{K}}(t_k)$ and $\underline{K}_k \triangleq \underline{K}(t_k)$ are time varying weighting matrices as yet unspecified. As a matter of fact, we wish to minimize, in a certain sense, the prediction error

$$\tilde{\underline{x}}(t_k|t_k) = \hat{\underline{x}}(t_k|t_k) - \underline{x}(t_k) \quad . \tag{7}$$

Substituting Eq. (5) into Eq. (6) and utilizing the properties of the noise process, it can readily be seen that Eq. (6) will be an unbiased estimation only if $\tilde{\underline{K}} = \underline{I} - \underline{K}_k \underline{H}_k$. Hence, the state estimation $\hat{x}(t_k|t_k)$, using the new measurement $\underline{z}(t_k)$, is

$$\hat{\underline{x}}(t_k|t_k) = \hat{\underline{x}}(t_k|t_{k-1}) + \underline{K}_k[\underline{z}(t_k) - \underline{H}_k \hat{\underline{x}}(t_k|t_{k-1})] \tag{8}$$

where $\underline{K}_k$ is still unspecified, and the initial condition at $t = t_o$ for the state estimation is given by

$$\hat{\underline{x}}(t_o|t_o) = \mathcal{E}\{\underline{x}(t_o)\} = \hat{\underline{x}}(t_o) \quad .$$

As a measure of the goodness of the estimation, we use the n x n covariance matrix $\underline{P}(\cdot)$ of the prediction error defined as

$$\underline{P}(t_k|t_k) = \mathcal{E}\{\tilde{\underline{x}}(t_k|t_k)\ \tilde{\underline{x}}^T(t_k|t_k)\} \quad , \tag{9}$$

which is obviously symmetric, and its trace is the mean square length of the vector $\tilde{\underline{x}}(\cdot)$. Its initial condition is given by

$$\underline{P}(t_o|t_o) = \mathcal{E}\{(\underline{x}(t_o) - \hat{\underline{x}}(t_o))\ (\underline{x}(t_o) - \hat{\underline{x}}(t_o))^T\}$$

$$= \text{var}\ \{\underline{x}(t_o)\} = \underline{P}(t_o) \quad .$$

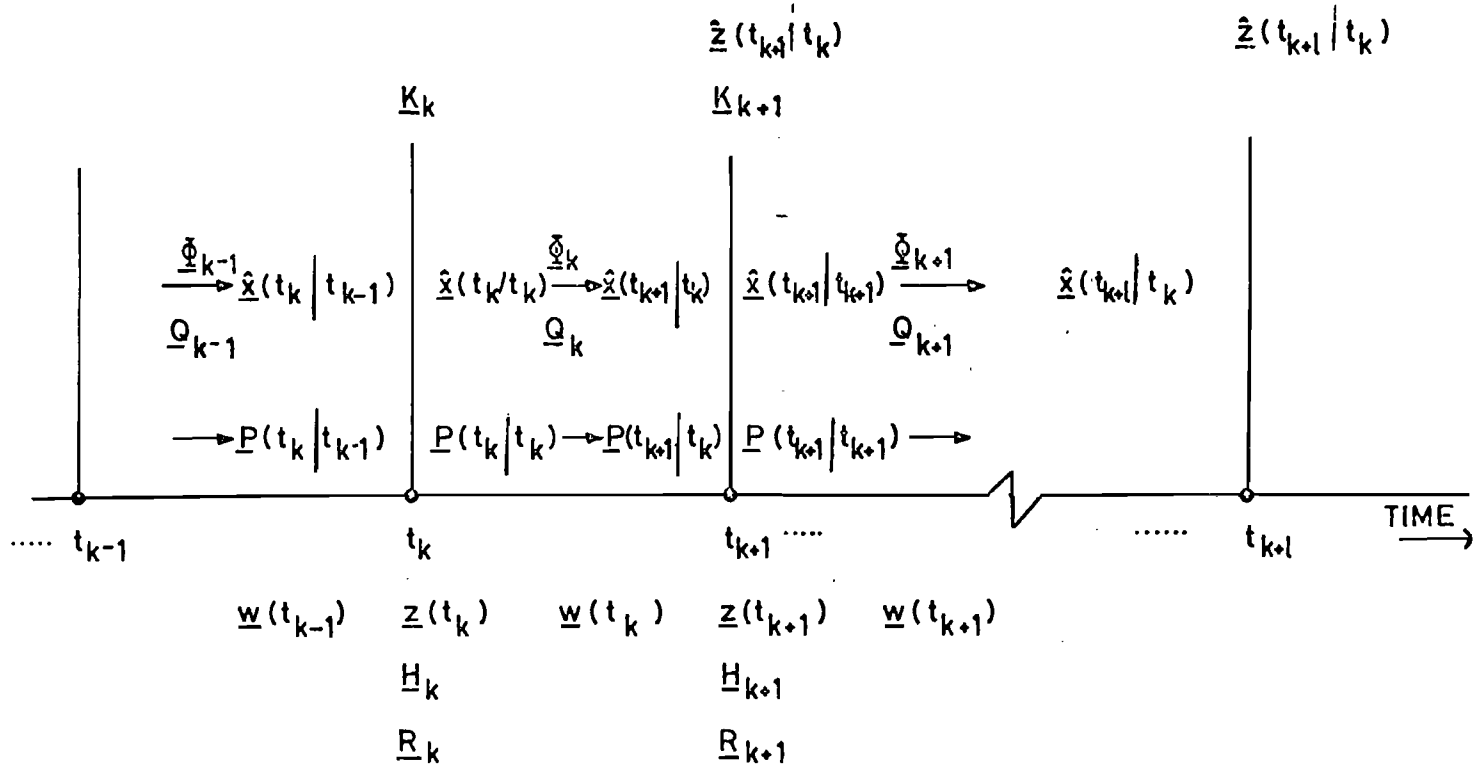

FIGURE 3. DISCRETE TIMING DIAGRAM

It can also easily be seen that the covariance matrix of $\underline{\tilde{x}}(t_k|t_k)$ can be projected from that of $\underline{\tilde{x}}(t_k|t_{k-1})$ as

$$\underline{P}(t_k|t_k) = (\underline{I} - \underline{K}_k \underline{H}_k)\, \underline{P}(t_k|t_{k-1})\, (\underline{I} - \underline{K}_k \underline{H}_k)^T + \underline{K}_k \underline{R}_k \underline{K}_k^T \quad . \tag{10}$$

Now, we define the loss function as the following quadratic form:

$$L_k \triangleq L(\underline{\tilde{x}}(t_k|t_k)) = \underline{\tilde{x}}^T(t_k|t_k)\, \underline{S}\, \underline{\tilde{x}}(t_k|t_k) \quad ,$$

where $\underline{S}$ is any positive semi-definite matrix; for the sake of simplicity let $\underline{S} = \underline{I}$, the identity matrix. Having defined the loss function we seek that estimate $\underline{\hat{x}}(t_k|t_k)$ of $\underline{x}(t_k)$--in other words, that form of the yet unspecified $\underline{K}_k$--which minimizes the expected loss (or Bayesian risk)

$$B_k = \mathcal{E}\{L(\underline{\tilde{x}}(t_k|t_k))\} \quad .$$

Since B_k is the trace of the error covariance matrix (cf. Eq. (9)) the problem is to minimize the Euclidean norm $||\underline{P}(t_k|t_k)||$ of $\underline{P}(t_k|t_k)$, i.e. the length of the estimation error vector. Using the properties of matrix derivatives, it can be seen that the weighting matrix $\underline{K}_k$ can be obtained from

$$\frac{\partial}{\partial \underline{K}_k} ||\underline{P}(t_k|t_k)|| = \underline{0}$$

as

$$\underline{K}_k = \underline{P}(t_k|t_{k-1})\, \underline{H}_k^T\, [\underline{H}_k\, \underline{P}(t_k|t_{k-1})\, \underline{H}_k^T + \underline{R}_k]^{-1} \quad , \tag{11}$$

which is referred to as the Kalman gain matrix. Now, the next step is the extrapolation of the state variable. Consider the one-step-ahead case, when $\ell = 1$. In the process model, Eq. (4), $\underline{w}(\cdot)$ is a white noise sequence, so no more information on it is contained in $\underline{z}(\cdot)$; thus the best prediction of $\underline{w}(\cdot)$ that can be

made from $\underline{z}(\cdot)$ is its mean value, i.e. 0. Consequently, the one-step-ahead prediction of the state vector, given observations up to t_k, is

$$\hat{\underline{x}}(t_{k+1}|t_k) = \underline{\Phi}_k \hat{\underline{x}}(t_k|t_k) \quad . \tag{12}$$

The propagation of prediction errors, i.e. $\underline{P}(t_k|t_k) \rightarrow \underline{P}(t_{k+1}|t_k)$, can be determined by computing the predicted error covariance matrix as

$$\underline{P}(t_{k+1}\ t_k) = \mathcal{E}\{\tilde{\underline{x}}(t_{k+1}|t_k)\ \tilde{\underline{x}}^T(t_{k+1}|t_k)\} \quad .$$

Using Eq. (12) and (4) and utilizing the fact that the prediction error and model error are independent of each other, we obtain

$$\underline{P}(t_{k+1}|t_k) = \underline{\Phi}_k \underline{P}(t_k|t_k)\ \underline{\Phi}_k^T + \underline{\Gamma}_k \underline{Q}_k \underline{\Gamma}_k^T \quad . \tag{13}$$

Using the formulas in the order of Eq. (12), (13), then with k: = k + 1 in (11), (8) and (10) the celebrated Kalman filter is obtained. The algorithms should be used sequentially, k = 1,2,..., starting with the given initial conditions at time t_o.* The complete algorithms, together with the initial conditions, are summarized in Table 1. Kalman has shown that the algorithms are convergent and stable [11].

Up to this point, we assumed that the noise covariance matrices $\underline{Q}_k$, $\underline{R}_k$ at time t_k are known in the estimation algorithms. But in dealing with water resource time series, this is far from being true; it is necessary to predict $\underline{Q}_k$ and $\underline{R}_k$ based upon measurements at the previous stage. Hence, to take into account the changing structure of uncertainties, an adaptive algorithm should be constructed for estimating the noise covariances, starting with arbitrary initial guesses. Since the noise covariance matrix $\underline{R}(t_k)$ is assumed to be independent of time, the one-step-ahead prediction of it is

---

*It might be mentioned that the same algorithms are obtained by maximizing the a posteriori probability $P(\underset{\sim}{X}_k|\underset{\sim}{Z}_k)$ where $\underset{\sim}{X}_k = \{\underline{x}(t_i) : i = 1,2,\ldots,k\}$ and $\underset{\sim}{Z}_k$ is as before. For a detailed discussion, consult Sage [16].

System Model	$\underline{x}(t_{k+1}) = \underline{\Phi}_k \, \underline{x}(t_k) + \underline{\Gamma}_k \, \underline{w}(t_k) \quad , \quad \underline{w}(t_k) \sim N(\underline{o}, \underline{Q}_k)$
Measurement Model	$\underline{z}(t_k) = \underline{H}_k \, \underline{x}(t_k) + \underline{v}(t_k) \quad , \quad \underline{v}(t_k) \sim N(\underline{o}, \underline{R}_k)$
Initial Conditions	$\mathcal{E}\{\underline{x}(t_o)\} = \hat{\underline{x}}(t_o) = \hat{\underline{x}}(t_o \mid t_o)$
	$\mathcal{E}\{(\underline{x}(t_o) - \hat{\underline{x}}(t_o))(\underline{x}(t_o) - \hat{\underline{x}}(t_o))^T\} = \text{var}\{\underline{x}(t_o)\} = \underline{P}(t_o) = \underline{P}(t_o \mid t_o)$
Other Assumption	$\mathcal{E}\{\underline{w}(t_k) \, \underline{v}^T(t_j)\} = \underline{o}, \quad \forall \, k,j$
State Prediction	$\hat{\underline{x}}(t_{k+1} \mid t_k) = \underline{\Phi}_k \, \hat{\underline{x}}(t_k \mid t_k)$
Predicted Error Covariance Matrix	$\underline{P}(t_{k+1} \mid t_k) = \underline{\Phi}_k \, \underline{P}(t_k \mid t_k) \, \underline{\Phi}_k^T + \underline{\Gamma}_k \, \underline{Q}_k \, \underline{\Gamma}_k^T$
Predictor Gain Algorithm	$\underline{K}_{k+1} = \underline{P}(t_{k+1} \mid t_k) \, \underline{H}_{k+1}^T [\underline{H}_{k+1} \, \underline{P}(t_{k+1} \mid t_k) \, \underline{H}_{k+1}^T + \underline{R}_{k+1}]^{-1}$
State Estimation Using the New Measurement	$\hat{\underline{x}}(t_{k+1} \mid t_{k+1}) = \hat{\underline{x}}(t_{k+1} \mid t_k) + \underline{K}_{k+1} [\underline{z}(t_{k+1}) - \underline{H}_{k+1} \, \hat{\underline{x}}(t_{k+1} \mid t_k)]$
Error Covariance Matrix Algorithm	$\underline{P}(t_{k+1} \mid t_{k+1}) = (\underline{I} - \underline{K}_{k+1} \, \underline{H}_{k+1}) \underline{P}(t_{k+1} \mid t_k) (\underline{I} - \underline{K}_{k+1} \, \underline{H}_{k+1})^T + \underline{K}_{k+1} \, \underline{R}_{k+1} \cdot \underline{K}_{k+1}^T$

DISCRETE KALMAN FILTER ALGORITHMS

TABLE 1

$$\hat{\underline{R}}(t_k|t_{k-1}) = \hat{\underline{R}}(t_{k-1}|t_{k-1}) \quad . \tag{14}$$

For the sequential estimation of $\hat{\underline{R}}(\cdot)$ Sage and Husa [17] developed a suboptimal adaptive estimation algorithm:

$$\hat{\underline{R}}(t_k|t_k) = \frac{1}{t_k}\,[t_{k-1}\,\hat{\underline{R}}(t_{k-1}|t_{k-1}) + \underline{\nu}(t_k)\,\underline{\nu}^T(t_k) - \underline{H}_k\,\underline{P}(t_k|t_{k-1})\,\underline{H}_k^T] \quad , \tag{15}$$

where

$$\underline{\nu}(t_k) = \underline{z}(t_k) - \underline{H}_k\,\hat{\underline{x}}(t_k|t_{k-1}) \tag{16}$$

is known as 'innovation sequence' (Kailath [9]) for the suboptimal estimator. The innovation process $\underline{\nu}(\cdot)$ is a white noise sequence, i.e. heuristically there is no information left in $\underline{\nu}(\cdot)$ if $\hat{\underline{x}}(\cdot)$ is an optimal estimation (Mehra [12]). A similar expression can be obtained for the adaptive estimator of the model noise covariance. Sage and Husa have also shown in their paper cited that the suboptimal estimation rapidly converges to the optimal one when t_k is increasing. It should be mentioned that there are numerous adaptive algorithms ([18, 12]) on the market, but for our purpose the above seems to be the most effective, at least from a computational point of view.

Adaptive Prediction of Linear Hydrologic Systems

It is well known (see e.g. Dooge [3]) that a fairly large class of hydrologic systems (e.g. rainfall excess/surface runoff, runoff/runoff transformations of flood routing, etc.) can be described by a convolution type of model

$$y(t) = h(t) * u(t)$$

where $u(t)$ is the input of the system (either controllable or not), $h(t)$ is the impulse response of the system and $y(t)$ is the output process; the asterisk denotes the convolution. In practice, however, we have only noise corrupted measurements

$$z(t) = y(t) + v(t)$$

where v(t) is an unknown noise process (Figure 4). Hence for linear time invariant lumped systems

$$z(t) = \int_{-\infty}^{\infty} h(\tau)\ u(t-\tau)\ d\tau + v(t) \quad , \tag{17}$$

where in case of physically realizable systems the upper bound of the integration is t. Note that although the system was assumed linear, in case of slight non-linearities, the noise process v(·) might be sought as a term including those 'small' non-linear disturbances.

Considering discrete-time systems with finite memory q, Eq. (17) can be written as

$$z(t_k) = \sum_{j=0}^{q} h(t_j)u(t_{k-j}) + v(t_k) \quad , \tag{18}$$

and by defining the vectors

$$\underline{H}_k = [u(t_k),\ u(t_{k-1}),\ldots,u(t_{k-q})] \quad ,$$

$$\underline{x}(t_k) = [h(t_o),\ h(t_1),\ldots,h(t_q)]^T$$

Eq. (18) becomes

$$z(t_k) = \underline{H}_k\, \underline{x}(t_k) + v(t_k) \quad . \tag{19}$$

This equation can be looked upon as a measurement equation for the above-defined state vector $\underline{x}(\cdot)$; cf. Eq. (5). The missing state equation can also be introduced without much difficulty. It was assumed that the system is time invariant, i.e. its impulse response h(·) does not change with time. Using the state vector defined above, this statement can be formulated as

$$\underline{x}(t_{k+1}) = \underline{x}(t_k) \tag{20}$$

which plays the role of the state equation.

Although it was assumed that the system is truly time variant, it should be stressed that the above formulation can be used for describing slightly time variant systems which,

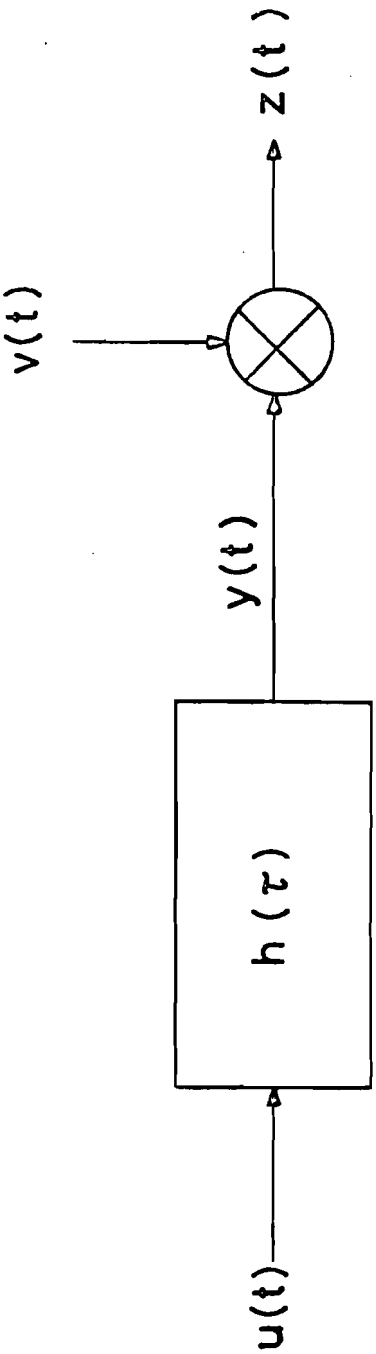

FIGURE 4. LINEAR SYSTEM MODEL

due to seasonal changes, are most common in hydrology. This concept is illustrated in Figure 5 where the system behaviour is considered to be time invariant within a well defined 'data window.' This data window, of course, is of a moving type. As to the length of the moving data window, it is essentially equal to the memory of the system and might be estimated from cross-correlation analysis of the input/output processes. The moving data window creates the basis of the sequential prediction.

If we assume that the noise sequence $v(t_k)$ is Gaussian white with $v(t_k) \sim N(0,R_k)$, then it is still an open question how to determine its variance. This can be done by the adaptive algorithm of Saga and Husa previously discussed, or even more easily because of the special structure of the state space model. The specialities are $\underline{\Phi}_k = \underline{I}$, $\underline{\Gamma}_k = \underline{0}$, and hence the state prediction is

$$\hat{\underline{x}}(t_{k+1}|t_k) = \hat{\underline{x}}(t_k|t_k) \quad ,$$

and the predicted error covariance matrix is in the form of

$$\underline{P}(t_{k+1}|t_k) = \underline{P}(t_k|t_k) \quad .$$

Since $v(\cdot)$ is a zero mean white noise sequence, the optimal one-step-ahead Bayes (minimum variance) prediction of the output process, based upon observations up to t_k, is

$$\hat{z}(t_{k+1}|t_k) = \underline{H}_k \hat{\underline{x}}(t_k|t_k) \quad .$$

The complete sequential prediction algorithms are summarized in Table 2. Note that to use the recursive algorithms, the initial conditions $\hat{\underline{x}}(t_o)$, var $\{\underline{x}(t_o)\}$ and R_o must be specified (or rather assumed).

In order to illustrate the utility of the proposed algorithms, a simulation exercise was elaborated. A given impulse response was assumed, and using that and an arbitrary input sequence, the output process was calculated through the simple discrete convolution. Then a Gaussian white noise sequence was generated with zero mean and variance 0·1. This sequence was then added to the output process; the resulting noise corrupted sequence and the original input sequence were further analyzed to see whether the algorithm does or does not give

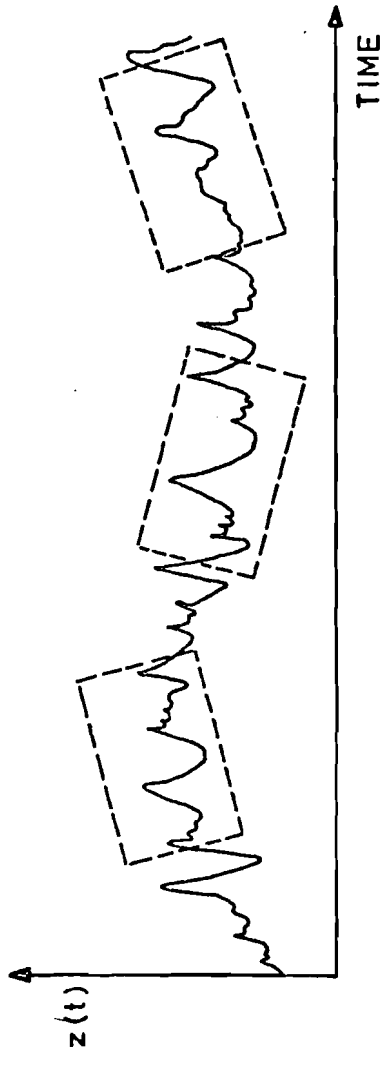

FIGURE 5. TIME VARYING MEASURMENT SEQUENCE WITH MOVING DATA WINDOW

Process Model	$\underline{x}(t_{k+1}) = \underline{x}(t_k)$
Measurement Model	$z(t_k) = \underline{H}_k\ \underline{x}(t_k) + v(t_k)\ ,\quad v(t_k) \sim N(o, R_k)$
Initial Conditions	$\hat{\underline{x}}(t_o) = \mathcal{E}\{\underline{x}(t_o)\}\ ,\quad \underline{P}(t_o) = \text{var}\ \{\underline{x}(t_o)\}$
State Prediction	$\hat{\underline{x}}(t_{k+1}\vert t_k) = \hat{\underline{x}}(t_k\vert t_k)$
Output Prediction	$\hat{z}(t_{k+1}\vert t_k) = \underline{H}_k\ \hat{\underline{x}}(t_k\vert t_k)$
Predicted Error Covariance Matrix	$\underline{P}(t_{k+1}\vert t_k) = \underline{P}(t_k\vert t_k)$
Predictor Gain Algorithm	$\underline{K}_{k+1} = \underline{P}(t_k\vert t_k)\ \underline{H}_{k+1}^T[\underline{H}_{k+1}\ \underline{P}(t_k\vert t_k)\ \underline{H}_{k+1}^T + \hat{R}_k]^{-1}$
State Estimation Using the New Measurement	$\hat{\underline{x}}(t_{k+1}\vert t_{k+1}) = \hat{\underline{x}}(t_k\vert t_k) + \underline{K}_{k+1}[z(t_{k+1}) - \underline{H}_{k+1}\ \hat{\underline{x}}(t_k\vert t_k)]$
Error Covariance Matrix Algorithm	$\underline{P}(t_{k+1}\vert t_{k+1}) = (\underline{I} - \underline{K}_{k+1}\ \underline{H}_{k+1})\underline{P}(t_k\vert t_k)(\underline{I} - \underline{K}_{k+1}\ \underline{H}_{k+1})^T + \underline{K}_{k+1}\ \hat{R}_{k+1}\cdot \cdot \underline{K}_{k+1}^T$
Measurement Noise Variance Algorithm	$\hat{R}_{k+1} = \frac{1}{t_{k+1}}\ [t_k\ R_k + v(t_{k+1})\ v^T(t_{k+1}) - \underline{H}_{k+1}\ \underline{P}(t_k\vert t_k)\underline{H}_{k+1}^T]$
Innovations Algorithm	$v(t_{k+1}) = z(t_{k+1}) - \underline{H}_{k+1}\ \hat{\underline{x}}(t_k\vert t_k)$

DISCRETE SUBOPTIMAL ADAPTIVE PREDICTION ALGORITHMS
TABLE 2

'back' the impulse response assumed. As an example Figure 6 shows the situation concerning a particular ordinate of the impulse response. The constant line (a) means the 'true' third ordinate of the impulse response, h_3, while curve (b) shows its estimated values using the prior knowledge (if it is available) of the variance; curve (c) shows how its estimated values evolve when there is no prior knowledge, i.e. an initial guess for the variance had been considered and the adaptive noise variance estimation technique was used. It is clear from the figure that whatever the initial guess is, the estimation procedure is convergent as the number of measurement data increases. The history of the adaptive sequential noise variance estimation is depicted in Figure 7. In fact, the same conclusion might be drawn.

Summary and Conclusions

The paper outlined the state space formulation of hydrologic/water resource systems. Prediction algorithms have been proposed which satisfy the requirements of the suitable prediction scheme laid down in the introduction since:

(1) Using time domain formulation, the usual frequency-domain-based computations can be avoided on the one hand and the problem becomes mathematically tractable on the other;

(2) Due to the recursiveness of the algorithms the scheme can easily be implemented even for small computers and are applicable for real-time on-line forecasting, always taking into consideration the newest information gathered;

(3) Due to the state space formulation, it is generally applicable to most general hydrologic time series (water quantity and/or quality); thus the joint handling/prediction of multidimensional time series (which might include some economic data) becomes feasible even in the presence of different kinds of uncertainties;

(4) The algorithms give optimal prediction in Bayes' sense (Bayesian minimum variance estimators);

(5) The requirement of adaptivity to changing environmental conditions are fulfilled, as through a moving data window slight modifications in the model parameters are allowed;

(6) The algorithms are convergent and stable under very general conditions.

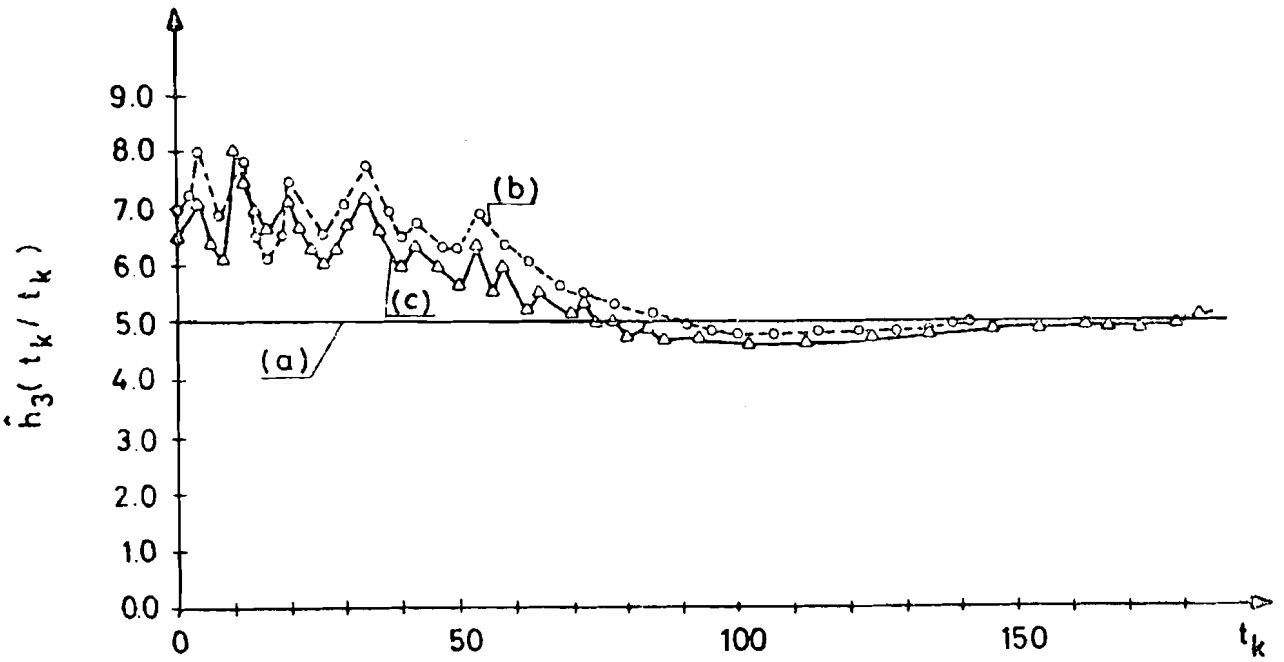

FIGURE 6. SEQUENTIAL ESTIMATION OF THE 3rd IMPULSE RESPONSE ORDINATE, h_3

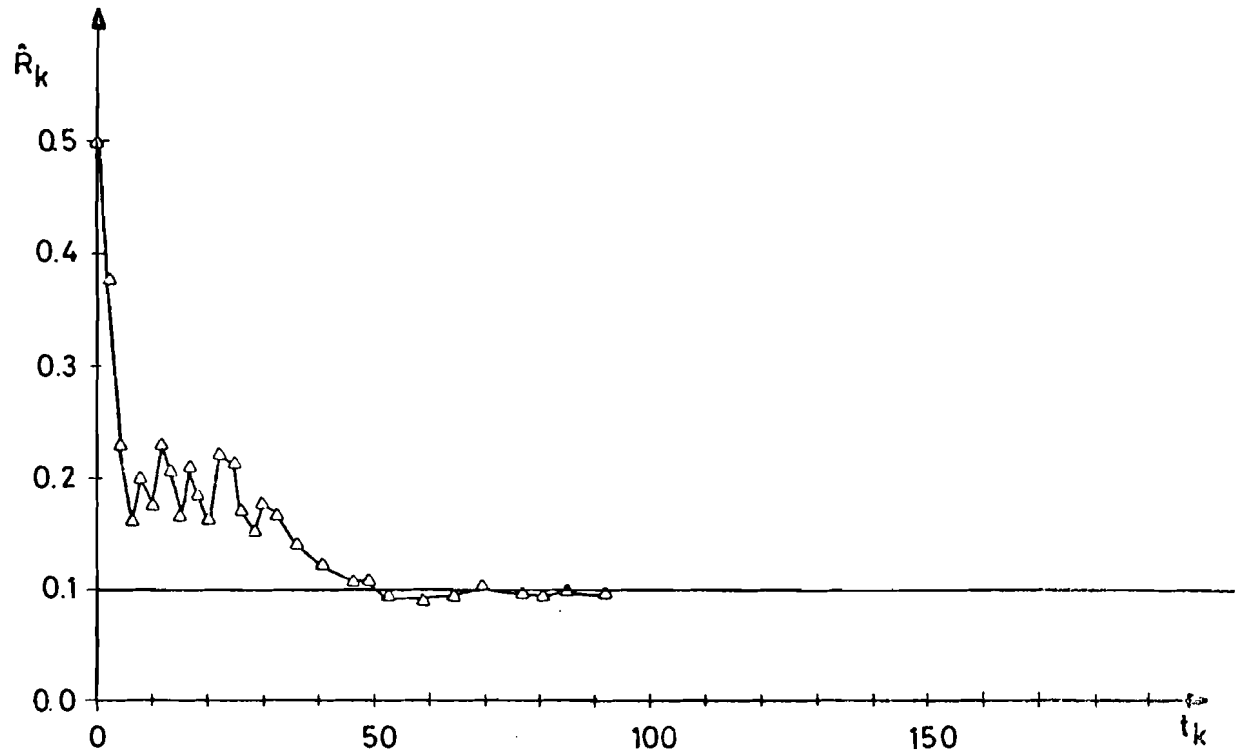

FIGURE 7. ADAPTIVE SUBOPTIMAL SEQUENTIAL ESTIMATION OF THE NOISE VARIANCE

To illustrate the above properties, an example was presented using simulated data. The results obtained indicate the practical applicability of the proposed procedure.

As a final remark, it might be mentioned that the procedure can be extended to include the identification/prediction of stochastic non-linear hydrologic system. This could be done, for example, by augmenting the state vector with the ordinates of the higher-order impulse responses and then taking advantage of the non-linear filtering techniques. But a lot of effort still remains to be made in the future towards the solution of these problems.

References

[1] Aoki, M., Optimization of Stochastic Systems, New York, Academic Press, 1967.

[2] Bryson, A.E. and Ho, Y.C., Applied Optimal Control, Waltham, Mass., Blaisdell Publishing Co., 1969.

[3] Dooge, J.C.I., Linear Theory of Hydrologic Systems, USDA Techn. Bulletin No. 1468, Washington, D.C., October 1973.

[4] Eagleson, P.S., Mejia, R. and March, F., "Computation of Optimum Realizable Unit Hydrograph", Water Resources Research, 2, 4 (1966).

[5] Harrison, P.J. and Stevens, C.F., "A Bayesian Approach to Short-Term Forecasting", Oppl. Res. Quart., 2, 4 (1971).

[6] Hino, M., "Analytic Solution of Wiener-Hopf Equation for Runoff Prediction", Proc. IAHR XIII. Congr., Kyoto, 1969.

[7] Ivakhnenko, A.G., and Lapa, V.G., Cybernetics and Forecasting Techniques, New York, American Elsevier Publ. Co. Inc., 1967.

[8] Jazwinski, A.H., "Adaptive Filtering", Automatica, 5, (1969) 475-485.

[9] Kailath, T., "An Innovations Approach to Least-Squares Estimation - Part I.: Linear Filtering in Additive White Noise", IEEE Trans. Automatic Control, AC-13, (December, 1968), pp. 646-655.

[10] Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems", J. Basic. Eng., 82D, (March, 1960), 35-45.

[11] Kalman, R.E., "New Methods in Wiener Filtering Theory", in Proc. First. Symp. on Engineering Application of Random Function Theory and Probability (ed. J.L. Bogdanoff and F. Kozin), New York, John Wiley and Sons Inc., 1963.

[12] Mehra, R.K., "On the Identification of Variances and Adaptive Kalman Filtering", Proc. 1969, Joint Automatic Control Conf., (1969), pp. 494-505.

[13] Miller, R.W., "Asymptotic Behaviour of the Kalman Filter with Exponential Aging", AIAA Journal, 9, 4 (April, 1971), 767-768.

[14] Naylor, T.H. and Seaks, T.G., "Box-Jenkins Methods: An Alternative to Econometric Models", Int. Stat. Rev., 40, 2 (1972), 123-137.

[15] Porebski, W., "Kalman Inverse Filter with Discrete Time Observations", Bull. de l'Academia Polonaise des Sci., 12, 20, 1 (1972), 15-19.

[16] Sage, A.P., Optimum Systems Control, Englewood Cliffs, N.J., Prentice-Hall, Inc., 1968.

[17] Sage, A.P. and Husa, G.W., "Adaptive Filtering with Unknown Prior Statistics", Proc. 1969 Joint Automatic Control Conf., (1969), pp. 760-769.

[18] Sage, A.P. and Melsa, J.L., System Identification, New York, Academic Press, 1971.

[19] Sawaragi, Y., Sunahara, Y. and Nakamizo, T., Statistical Decision Theory in Adaptive Control Systems, New York, Academic Press, 1967.

[20] Schweppe, F.C., Uncertain Dynamic Systems, Englewood Cliffs, N.J., Prentice-Hall, Inc., 1973.

[21] Sorenson, H.W., "Controllability and Observability of Linear, Stochastic, Time-discrete Control Systems", in Advances in Control Systems, (ed. C.T. Leondes), New York, Academic Press, 6, (1968), pp. 95-158.

[22] Szöllösi-Nagy, A., "Predicting the Discharge Time Series of Small Rivers and Brooks Using Wiener's Theory of Extrapolation", Proc. ICID IX. European Regional Conf., R:1.3/10, Budapest, 1973.

[23] Urbanek, K., Lectures on Prediction Theory, Berlin, Springer Verlag, Lect. Notes on Math., No. 44, 1967.

[24] Viscaino, A. and Bribiesca, J., "A Statistical Method to Search for a Rainfall-Runoff Relationship", Proc. Int. Symp. on Hydrology, Fort Collins, Col., 1967.