ANALYSIS OF HYDROLOGIC UNCERTAINTY

John C. Schaake, Jr.*

ABSTRACT

Suggested in this paper are some ideas for using deterministic models and stochastic theory together to reduce uncertainty in hydrologic information. The proposed approach is to apply the results of filtering theory which produces not only hydrologic estimates but also gives the covariance of the estimation errors. A hypothetical hydrological system is used to illustrate the filter approach in a simple river forecasting situation. Comparison of filter theory estimation errors with conventional estimation errors showed the filter theory errors were substantially smaller than conventional estimation errors. Details of the filter model for the hypothetical system are presented in an appendix. Potential applications of filter theory are suggested.

* United States National Weather Service, Silver Spring, Maryland, U.S.A.

ANALYSIS OF HYDROLOGIC UNCERTAINTY

by John C. Schaake, Jr.

INTRODUCTION

Measurement errors, model errors and natural variability of meteorological inputs to hydrologic systems are causes of uncertainty in hydrologic information. Although methods exist to analyze some of these uncertainties, improved techniques are needed. Methods exist, for example: to study the accuracy of hydrologic instrumentation; to describe the natural hydrologic variability of meteorological inputs to hydrologic systems either with probability distributions or with stochastic processes; and, to assess the accuracy of hydrologic models empirically by comparing computed results with observed data. Most extensively developed are univariate techniques for analysis of a single random variable or a single time series. Least extensively developed are multivariate techniques for analyses of physically inter-related multiple time series. Improved techniques are needed: (1) to assess the accuracy of models where computed results cannot be observed; (2) to permit measurements of variables such as snow water equivalent, soil moisture, or discharge (which may be computed model outputs) to be used more rationally to improve the accuracy of model results; (3) to make more rational use of data from more than one source (e.g., rainfall data from rain gages, radar and satellites); (4) to permit redundant information to influence hydrologic model results in proportion to the accuracy of the information.

Suggested in this paper are some ideas for using deterministic models and stochastic theory together to meet these needs. To be practical these ideas need extensive development and trial application. They are applied below to a simplified, hypothetical catchment. Potentially, they could help:

- develop improved forecast update procedures
- make better use of snow course data in river forecasting
- make better use of quantitative precipitation forecasts in river forecasting
- make better use in hydrology of satellite data which may contain much more uncertainty than ground data
- develop optimal network design procedures
- make better joint use of radar and rain gages to observe precipitation
- make use of runoff data jointly with point rainfall data to get improved estimates of mean areal precipitation

HYPOTHETICAL SYSTEM

Consider a hypothetical catchment and systems for measuring the mean areal precipitation over the catchment and the downstream discharge. This is illustrated in Figure 1.

Let

I_t = True mean areal precipitation (MAP) over the catchment during the interval <t-1,t>

P_t = Measurement of I_t from the rain gage network. This is a weighted average of point rainfall measurements.

k = Catchment parameter

O_t = True downstream discharge rate at time t

Q_t = Measurement of O_t from the stream gage.

Assume, for example, that I_t varies stochastically according to the autoregressive process

$$I_t = \rho I_{t-1} + \sigma_I \sqrt{1-\rho_I{}^2}\ \omega_{1_t} \tag{1}$$

where

ω_{1_t} = independently distributed standard normal deviate

σ_I = standard deviation of I_t

ρ_I = serial correlation coefficient of I_t

This is not strictly physically realistic for two principal reasons: (1) the autoregressive process is stationary whereas rainfall is not, and (2) I_t has, for convenience, a zero mean and can assume negative as well as positive values. (To be more practical it could have been defined as a deviation about an underlying storm pattern, but that would complicate the example.)

Assume also, for example, that the catchment functions with no evapotranspiration nor other loss of water, as a single linear storage reservoir according to

$$O_t = (1-k)\ O_{t-1} + kI_t \tag{2}$$

Since I_t can assume negative values, O_t also can be negative.

The true values of I_t and O_t can never be known exactly. They can only be measured and modeled, but measurements always include measurement error. Assume these measurements are made as

$$P_t = I_t + V_1{}^t \tag{3}$$

where

$V_1{}^t$ = a serially uncorrelated random normal measurement error with zero mean and variance σ_p^2.

and as

$$Q_t = O_t + V_2{}^t \tag{4}$$

where

$V_2{}^t$ = a serially uncorrelated random normal measurement error with zero mean and variance $\sigma_I{}^2$.

ESTIMATION OF THE TRUE STATE OF THE HYPOTHETICAL SYSTEM

The true values of I_t and O_t in the hypothetical system can never be known; but they can be estimated on the basis of measurements and model computations. Variables I_t and O_t are called the <u>state</u> of the system. The measurements of I_t and O_t are P_t and Q_t, respectively. The estimates of I_t and O_t are $\hat{I}_t$ and $\hat{O}_t$.

Conventional Approach

Conventionally, in hydrology, when measurements are available they are used directly to make required estimates. For example, available discharge measurements are used to estimate the true discharge without considering rainfall data nor rainfall-runoff models. Mathematically, this conventional estimate is

$$\hat{O}_t = Q_t \tag{5}$$

As another example, available point rainfall measurements are used to estimate MAP. Mathematically, this conventional estimate is

$$\hat{I}_t = P_t \tag{6}$$

Available discharge measurements, generally, are not used to improve $\hat{I}_t$; and, if rainfall data do not exist, conventionally no estimate can be made of I_t.

Equations 5 and 6 are introduced to help relate conventional hydrologic methods with the ideas to be presented. An important point is that

$$I_t \neq P_t \tag{7}$$

and

$$O_t \neq Q_t \tag{8}$$

The conventional approach in hydrology when rainfall data are available and streamflow data are not is to use a catchment model to estimate streamflow from rainfall. This is illustrated in Figure 2. The conventional approach to assess the accuracy of model estimates is to compare estimates with measured values. This also is shown in Figure 2.

Before presenting the proposed approach, estimation accuracy must be discussed briefly because there are subtle, but theoretically important differences between the conventional accuracy definition illustrated in Figure 2 and the definition to be proposed.

Measure of Accuracy

One of the most important needs in hydrology is to be able to say how accurate are given estimates, $\hat{I}_t$ and $\hat{O}_t$. This requires

° a definition of accuracy

° methods to evaluate given estimates

One approach to accuracy definition is to regard the true state, say I_t, as a random variable that has a probability density function (PDF) distributed about the estimated value $\hat{I}_t$. This is illustrated in Figure 3. This PDF has a mean $\mu_{I_t|P_t,Q_t}$ and standard deviation $\sigma_{I_t|P_t,Q_t}$ which may be used to measure the bias and variability of the estimate. The PDF in Figure 3 is a conditional PDF because $\hat{I}_t$ is based on measurements P_t and Q_t up to the current time.

A method exists to quantify $\mu_{I_t|P_t,Q_t}$ and $\sigma_{I_t|P_t,Q_t}$ but it requires a key assumption to be made. An assumption is needed because I_t and O_t cannot be observed, so conventional statistical methods cannot be used to estimate these statistics. The assumption is that the true relationship between all of the state variables is known. This may involve both deterministic and stochastic components. Whereas the true relationships between state variables in a natural process can never be known exactly, models of these relationships are frequently used to make estimates of state variables. If such models are acceptable for making estimates, perhaps they could be acceptable also for estimating the accuracy of the estimates.

Proposed Approach

The proposed approach is to apply the results of filtering theory which produces the estimate

$$\hat{I}_t = \mu_{I_t|P_t,Q_t} \text{ and } \hat{O}_t = \mu_{O_t|P_t,Q_t} \tag{9}$$

Jazwinski[1] presents a clear and complete discussion of filtering, including the necessary mathematical background in stochastic processes. Jazwinski also proves that estimates equal to the conditional mean, as in Equation (9) also are minimum variance estimates for all filtering problems both linear and non-linear. That is, these estimates minimize $\sigma^2_{I_t|P_t,Q_t} + \sigma^2_{O_t|P_t,Q_t}$.

The filtering approach is illustrated in Figure 4. In many respects the filtering approach is similar to the conventional approach, but some additional aspects appear. Most notable is that state estimates $\hat{I}_t$, $\hat{O}_t$ are not made directly by the catchment model but by the Kalman Filter. Also, this filter produces the covariance of the errors in the state estimates. (In other words, the Kalman filter is designed to compute both $\mu_{I|P,Q}$ and $\sigma_{I|P,Q}$.) In the hypothetical system, the covariance of the estimation errors forms a 2x2 matrix

$$\underset{\sim}{S}_t = \begin{pmatrix} \sigma^2_{I_t|P_t,Q_t} & \sigma_{I_tO_t|P_t,Q_t} \\ \sigma_{I_tO_t|P_t,Q_t} & \sigma^2_{O_t|P_t,Q_t} \end{pmatrix}. \tag{10}$$

The off-diagonal elements are non-zero if the estimates of the different state variables are correlated, which usually is the case.

The additional input information required by the filtering approach to produce not only a state estimate but also the covariance of the estimation errors are (1) covariance data for the measurement errors, (2) a stochastic model of the natural variations of I_t, and (3) system noise covariance data (i.e., parameters of the stochastic rainfall model).

The details of the Kalman filter model for the hypothetical system appear in Appendix A.

EXAMPLE APPLICATION IN RIVER FORECASTING

Suppose measurements P_t and Q_t were made for $t=1,\ldots,5$ and estimates of I_t and O_t are wanted at these same times. Also, a forecast of O_t is wanted for $t=6,\ldots,10$.

The "conventional" approach would be to estimate

$$\hat{I}_t = P_t \tag{11}$$

$$\hat{O}_t = Q_t \tag{12}$$

for $t \leq 5$ and to use a model of the catchment to forecast O_t using $\hat{I}_t = P_t$ as input up to $t = 5$ and $\hat{I}_t = 0$ thereafter. Results of this conventional approach will be compared below to results from the filtering approach.

<u>Example Data</u>

Table 1 presents assumed numerical values for the parameters of the hypothetical dynamical and measurement systems.

Table 1 Assumed Parameter Values

Parameter	Value
k	.2
ρ_I	.5
σ_I	5
σ_P	1
σ_Q	.25

Parameter k controls catchment lag time; ρ_{I_1} is related to the persistence of the precipitation; σ_I is related to the natural variability of precipitation; σ_P is the measurement error of the rain gage network; and σ_Q is the measurement error in discharge measurements (caused by uncertainty in the rating curve). The relative values of σ_P and σ_Q were chosen to preserve the fact that discharge measurements are much more accurate than precipitation measurements. The relative values of

σ_I and σ_P were chosen to preserve the fact that precipitation measurement errors are small compared to natural variability from time to time of precipitation amounts.

A time series of observations are need to apply the filter, so the hypothetical model was "operated" using the parameters in Table 1 and tables of standard normal deviates to assign values to the noise and error vectors. The values assigned to observed rainfall and discharge are given in Table 2.

Table 2 Rainfall and Discharge Measurements

Sampling Interval	Rainfall P_t	Discharge Q_t
1	4.66	.99
2	2.81	1.35
3	4.25	1.67
4	1.91	1.93
5	5.53	2.47

Filter Approach Estimation Errors

The Kalman filter was used to make estimates of both I_t and O_t from $t = 0$ to $t = 10$. Since the present time is $t = 5$, estimates beyond $t = 5$ were forecasts. Figure 5 shows how the standard deviation of error in the discharge estimate $\sigma_{O_t|P_t,Q_t}$ changed with time. Initially, it began equal to the measurement error, σ_Q, because the best possible estimate at $t = 0$ was the measured discharge. Then, as time progressed, the filter model used the measured discharge and rainfall data together to get better discharge estimates than could be gotten from the measured data alone. By $t = 5$, $\sigma_{O_t|P_t,Q_t}$ had reduced to 69 percent of the measurement error, σ_Q. Beyond $t = 5$ measurements of rainfall and discharge were unavailable, and $\sigma_{O_t|P_t,Q_t}$ increased. Ultimately, the water stored in the catchment was depleted, there was little information in the model unique to any particular time, and the estimation error approached the long-term climatic standard deviation of discharge, σ_O. At time $t = 10$, $\sigma_{O_t|P_t,Q_t}$ reached 88 percent of σ_O.

The filter gives estimates of I as well as of O. Figure 6 shows how the standard deviation of error in the rainfall estimate $\sigma_{I_t|P_t,Q_t}$ changed with time. It followed a pattern very similar to $\sigma_{O_t|P_t,Q_t}$ starting equal to the measurement error, σ_P and decreasing until at $t = 5$ it was equal to 81 percent of the measurement error. Beyond $t = 5$, there were no more measurements so the rainfall estimate was based soley on persistence. Very rapidly after $t = 5$, $\sigma_{I_t|P_t,Q_t}$ increased until it reached the long-term climatic standard deviation of rainfall, σ_I.

Conventional Approach Estimation Errors

Figure 7 shows how the standard deviation of the error in conventional discharge estimates changed with time. Prior to and including $t = 5$, the estimates were based only on the discharge measurements. Therefore, the estimation error was equal to the discharge measurement errorwhich was much less than the catchment model error but substantially greater than the estimation error for this period by the filter approach.

Beyond time $t = 5$ forecasts of O_t were made using the catchment model with input values of P_t up to $t = 5$. At time $t = 5$ the catchment model had complete rainfall information and the standard deviation of the model estimate of O_5 was equal to 0.33 which is 133 percent greater than the discharge measurement error and 193 percent of the error in the filter estimate of O_5. (Recall that use of the model and discharge measurements together in the filter approach reduced the estimation error of O_5 to 69 percent of the measurement error.)

Because there was a shift of estimation technique at $t = 5$ there was also a discontinuity in measurement error at $t = 5$.

The estimation error in discharge forecasts from the catchment model also increased and approached the long-term climatic standard deviation of discharge as occurred in the filter approach.

Comparison of Filter and Conventional Approaches

Figure 7 shows the filter discharge estimate error was less than the conventional discharge estimate error for the entire period $0 < t \leq 10$. The ratio of the filter error to the conventional error is shown in Figure 8. Note that the ratio is always less than 1.0 since the filter error was always less than the conventional error. Note also that the improvement is greatest in the immediate future beyond the present time and up to the present time.

SUMMARY AND CONCLUSIONS

Presented were some ideas for using deterministic and stochastic hydrologic models together to reduce uncertainty in hydrologic information. A hypothetical river forecasting example was used to compare these ideas with a conventional deterministic model approach. The proposed ideas involve use of Kalman filter theory which depends on deterministic models to represent physical processes while also making full use of much additional knowledge as well.

The approach taken to present the filter theory ideas was to select a very simple example involving a basic deterministic model - the linear storage reservoir; a basic stochastic model - the lag-one autoregressive process; and basic measurement error models. While still maintaining enough physical sense to be worth analyzing, a numerical example was used to examine some implications of this theory for river forecasting. This provides a basis for beginning to look at advantages and limitations of Kalman filter theory and for creating more complete, more physically meaningful models.

Estimates of present and future states may be produced by the Kalman filter. Also produced is the covariance matrix of the estimation errors. Kalman filter theory is more explicit about differences between <u>true states</u>, <u>estimated states</u> and <u>measured states</u> than conventional methods. The estimation error is the difference between the estimated state and the true state, which can never be known exactly. The filter estimates minimize the total variance of the estimation errors. The presence of measurement error is accounted for.

Although the example illustrated the original Kalman filter which was for stationary, time-invariant, linear problems, the filter is easily applied to non-stationary, time-varying problems and the extended Kalman theory can be used when non-linearities are important.

Some of the limitations of the hypothetical system in the example are:

- All relationships are linear.
- All parameters are assumed known and the implications of not knowing these are not explored.
- The "base event" is subtracted out because interest is in the estimation errors which depend mainly on deviations about some base. In principle this is correct, but there is no discussion of how to establish the base event nor are possible uncertainties introduced by the approach to base event estimation explored.
- The rainfall variation model is autoregressive, which may be adequate as a first approximation; but implications of more complex rainfall interrelationships have not been explored.
- In the example catchment model, only direct runoff is considered. This suggests I_t should be more precisely defined as "effective rainfall" and O_t as direct runoff. Therefore, the measurement errors must include uncertainty from the soil moisture accounting model.

Many of these limitations can be overcome by adding some more complexity to the example. It remains to be seen if all of them can be overcome.

REFERENCES

1. Jazwinski, A.H., Stochastic Processes and Filtering Theory, Academic Press, 1970.

2. Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems," Trans. ASME, Ser D: J. Basic Eng. 82, 35-45, 1960.

$\underline{F} = [F_{ij}]$ = state transition matrix from t to t+Δt; i=1, N; j=1,N.
In the example:

$$\underline{F} = \begin{pmatrix} \rho_I & 0 \\ k\rho_I & 1-k \end{pmatrix} \tag{A2}$$

The state transition matrix explains how $\underline{X}^t$ would change if there were no stochastic inputs to the system.

Two matricies and one vector are associated with the stochastic inputs to the system:

$\underline{W}^t = [W_i^t]$ = system noise vector, i=1,N

$\underline{G} = [G_{ij}]$ = system noise coefficient matrix;
i=1, N: j=1,N

$\underline{U} = [U_{ij}]$ = system noise covariance matrix,
(i=1,N; j=1,N) (i.e. U(1,1) is the
variance of ω_{it}),

In the example, the elements of $\underline{G}$ and $\underline{U}$ are:

$$\underline{W}^t = \begin{pmatrix} \omega_i^t \\ 0 \end{pmatrix} \tag{A3}$$

$$\underline{G} = \begin{pmatrix} \sigma_I \sqrt{1 - \rho_I^2} & 0 \\ k\sigma_I \sqrt{1 - \rho_I^2} & 0 \end{pmatrix} \tag{A4}$$

and

$$\underline{U} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{A5}$$

The dynamical system model is, then

$$\underline{X}^{t+\Delta t} = \underline{F}\,\underline{X}^{t} + \underline{G}\ \underline{W}^{t+\Delta t} \tag{A6}$$

which is equivalent in abstract matrix notation to Equations 1 and 2. The matrix $\underline{U}$ is the covariance matrix of the vector $\underline{W}$.

Measurement System Model

$\underline{M}$ = number of measured variables (in this case M=2)

$\underline{Y}^t = [Y_i^t]$ = measurement vector at time t; i=1,M
In this case

$$\underline{Y}^t = \begin{pmatrix} P_t \\ Q_t \end{pmatrix} \tag{A7}$$

$\underline{H} = [H_{ij}]$ = measurement matrix; i=1,M; j=1,N
In this case

$$\underline{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{A8}$$

The measurement matrix explains how $\underline{Y}^t$ would depend on $\underline{X}^t$ if there were no measurement errors. One matrix and one vector are associated with the measurement errors:

$\underline{V}^t = [V_i^t]$ = measurement error vector; i=1,M
In this case:

$$\underline{V}^t = \begin{pmatrix} V_1^{\ t} \\ V_2^{\ t} \end{pmatrix} \tag{A9}$$

$\underline{R} = [R_{ij}]$ = measurement error covariance vector; i=1,M; j=1;M.
In this case

$$\underline{R} = \begin{pmatrix} \sigma_P^2 & 0 \\ 0 & \sigma_Q^2 \end{pmatrix} \tag{A10}$$

Matrix $\underline{R}$ is the covariance matrix of vector $\underline{V}^t$.

The measurement model is, then

$$\underline{Y}^t = \underline{H}\,\underline{X}^t + \underline{V}^t \tag{A11}$$

which is equivalent in abstract matrix notation to Equations 3 and 4.

Filter Computations

Let the desired estimate of X_i^t be

$$\hat{X}_i^t = \text{state estimate at time } t\text{, given } Y^t \text{ up to and including time } t;\ i=1,N$$

In this case

$$\underline{\hat{X}}^t = \begin{pmatrix} \hat{I}_t \\ \hat{0}_t \end{pmatrix} \tag{A12}$$

The uncertainty in this estimate is given by the covariance matrix

$$S_{ij}^t = \text{covariance matrix of error in } \hat{X}_i^t;\ i=1,N;\ j=1,N$$

Step 1 - Matrix $\underline{S}^t$ is to be computed for $t>o$, but initial values of $\underline{\hat{X}}^t$ and $\underline{P}^t$ must be given at time $t=o$. These correspond to prior estimates of the initial state and the uncertainty in whether the system is in the estimated initial state. In the example, the initial state is assumed to be

$$\hat{I}_o = 0 \tag{A13}$$

$$\hat{0}_o = 0 \tag{A14}$$

based on the measurements at that time. Accordingly, $\underline{S}^o$ was taken to be the same as the measurement error covariance matrix

$$\underline{S}^o = \underline{R} \tag{A15}$$

which seemed reasonable because $\underline{H}$ is an identity matrix in this case.

Step 2 - Compute the forecast state at $t^{+\Delta t}$ given the measurements $\underline{Y}^t$ up to and including time t,

$$\underline{X1}^{t+\Delta t} = \underline{F}\,\underline{\hat{X}}^t \tag{A16}$$

This forecast is the best estimate that can be made of $\underline{X}^{t+\Delta t}$ if $\underline{Y}^{t+\Delta t}$ does not exist.

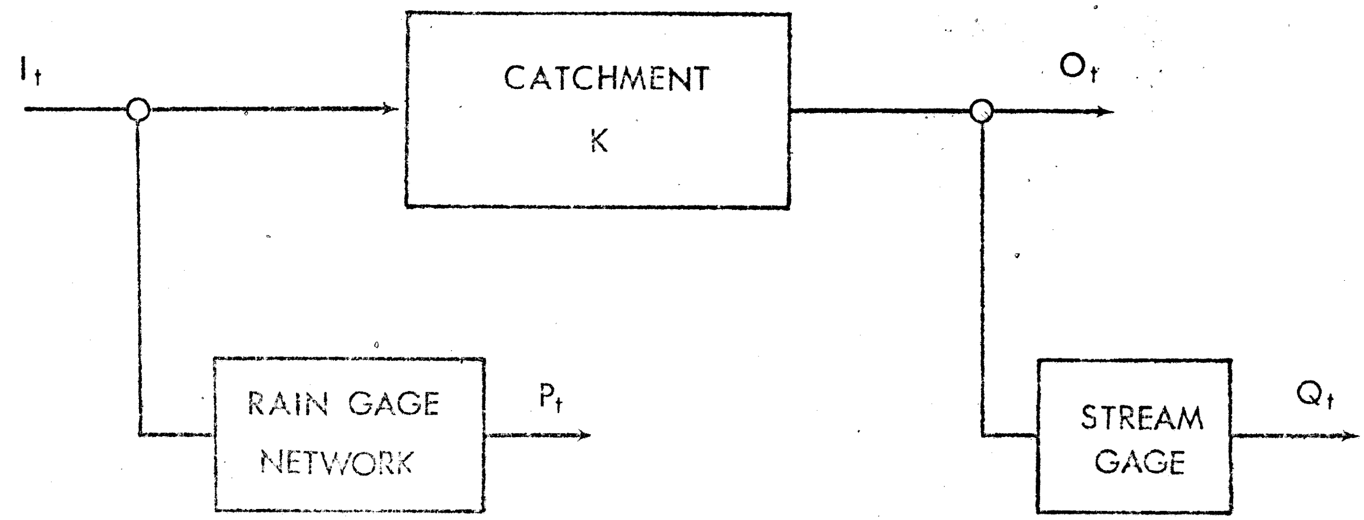

FIGURE 1. HYPOTHETICAL HYDROLOGIC SYSTEM

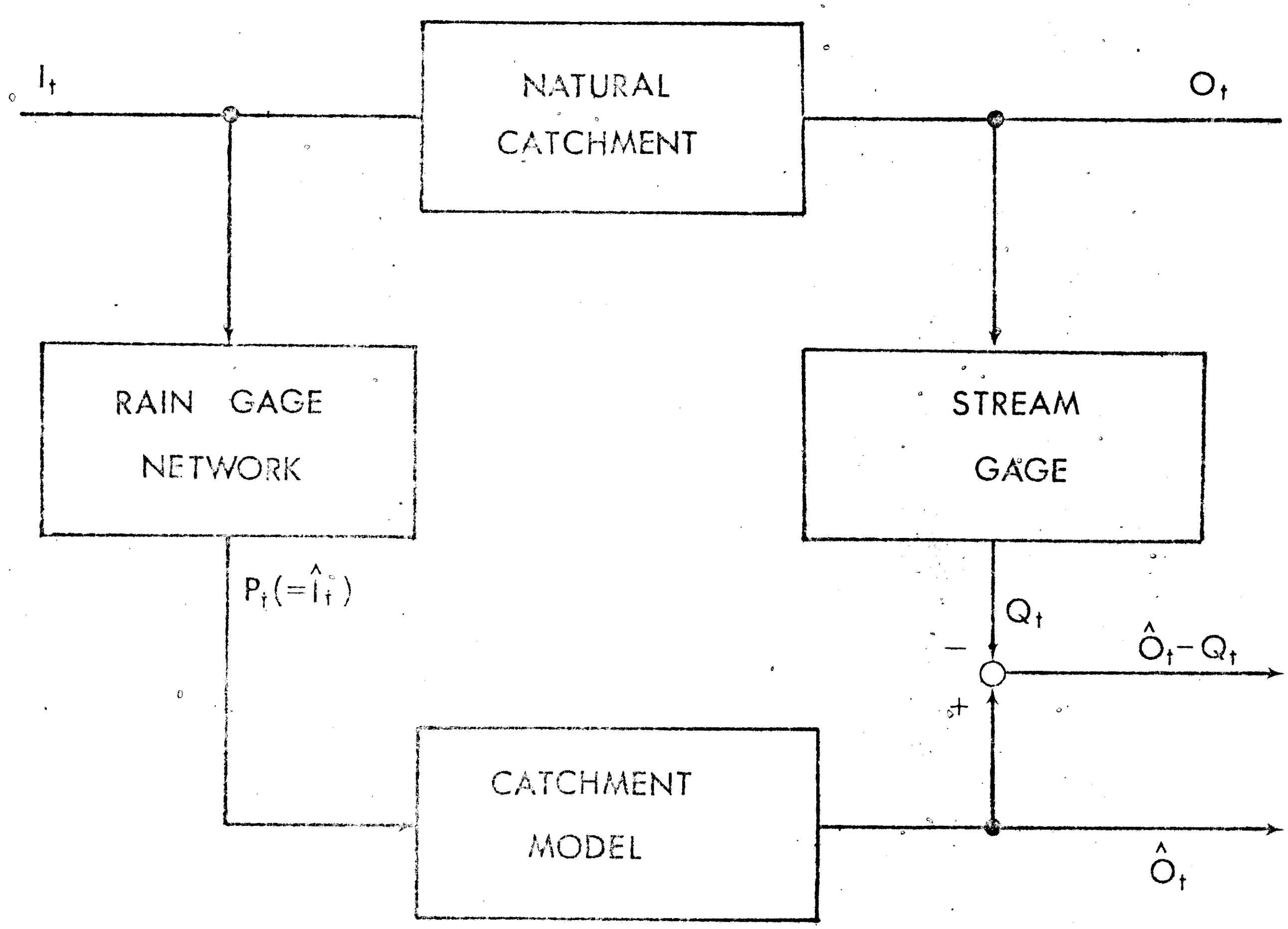

FIGURE 2. CONVENTIONAL DETERMINISTIC MODEL APPROACH

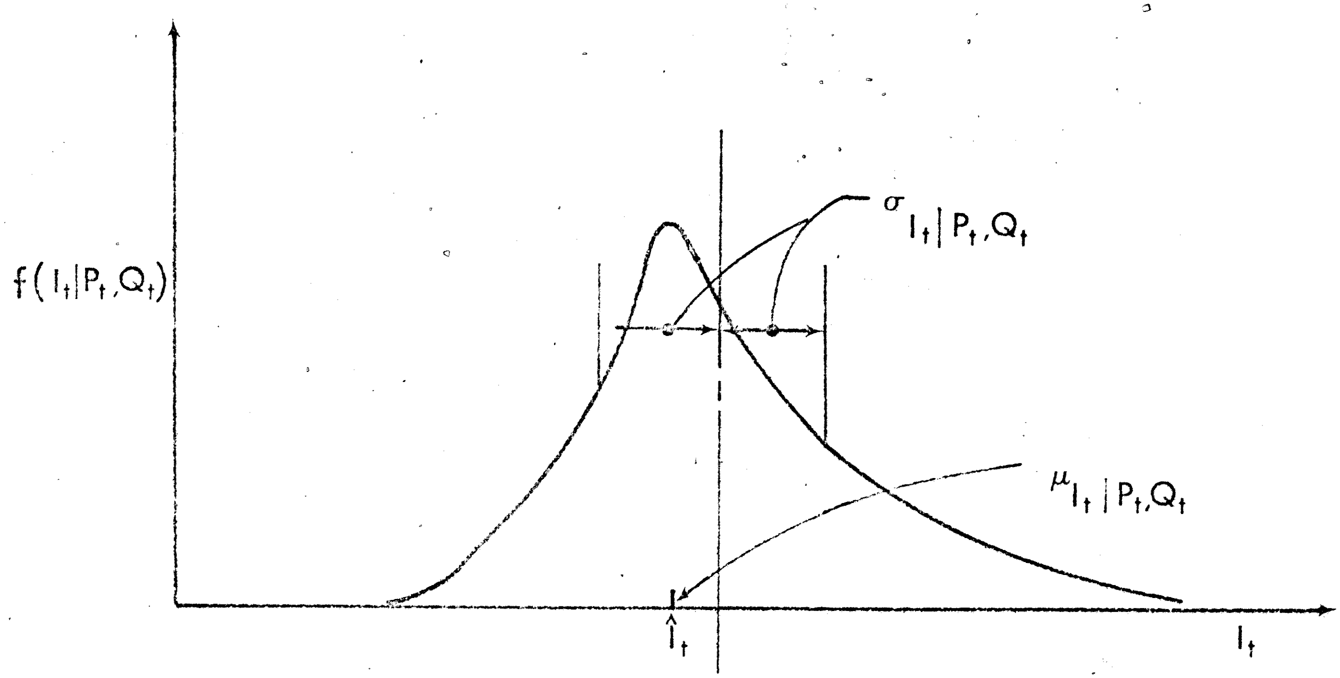

FIGURE 3. PROBABILITY DENSITY FUNCTION OF I_t, GIVEN THE MEASUREMENTS P_t AND Q_t

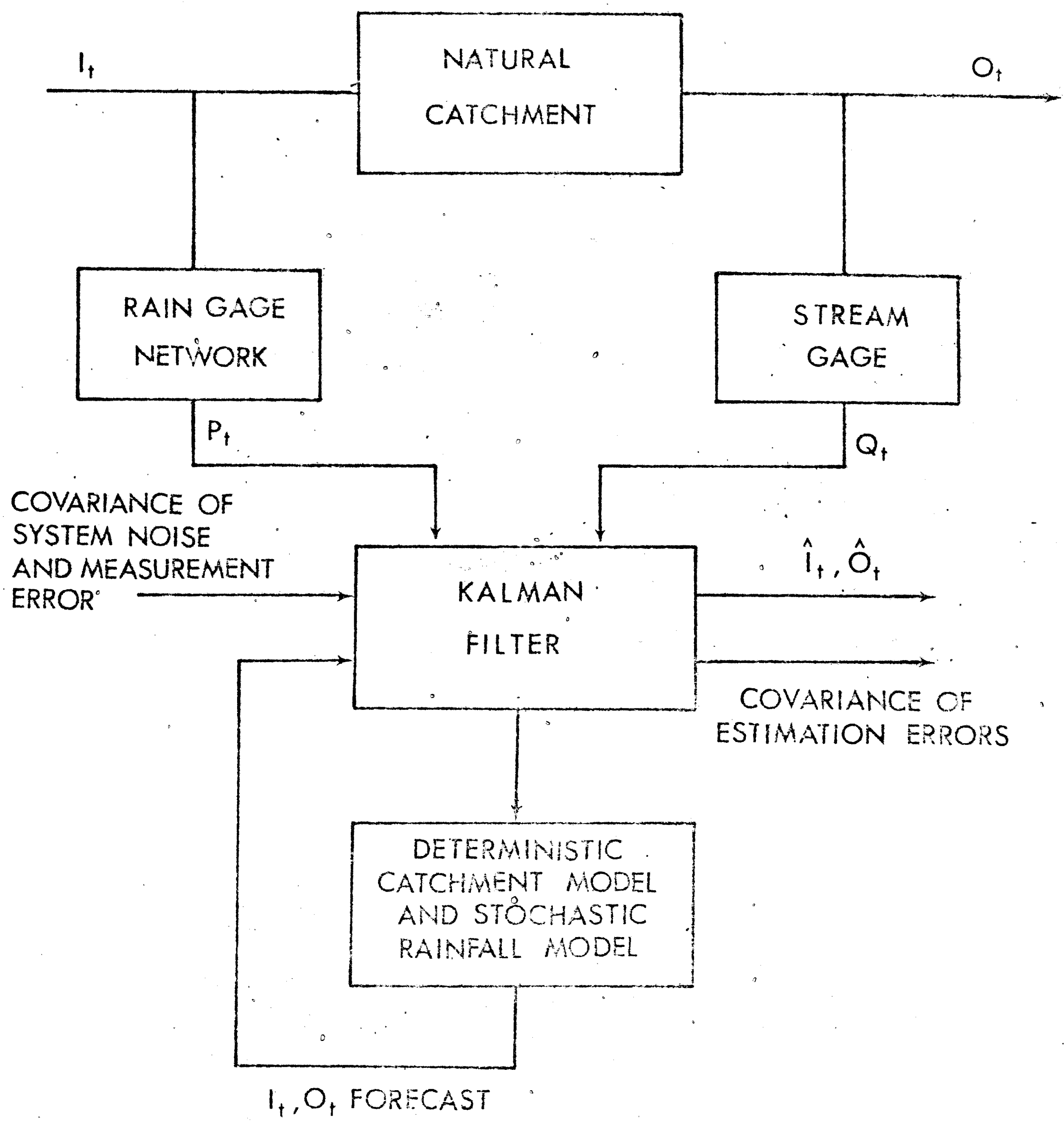

FIGURE 4. FILTER THEORY APPROACH

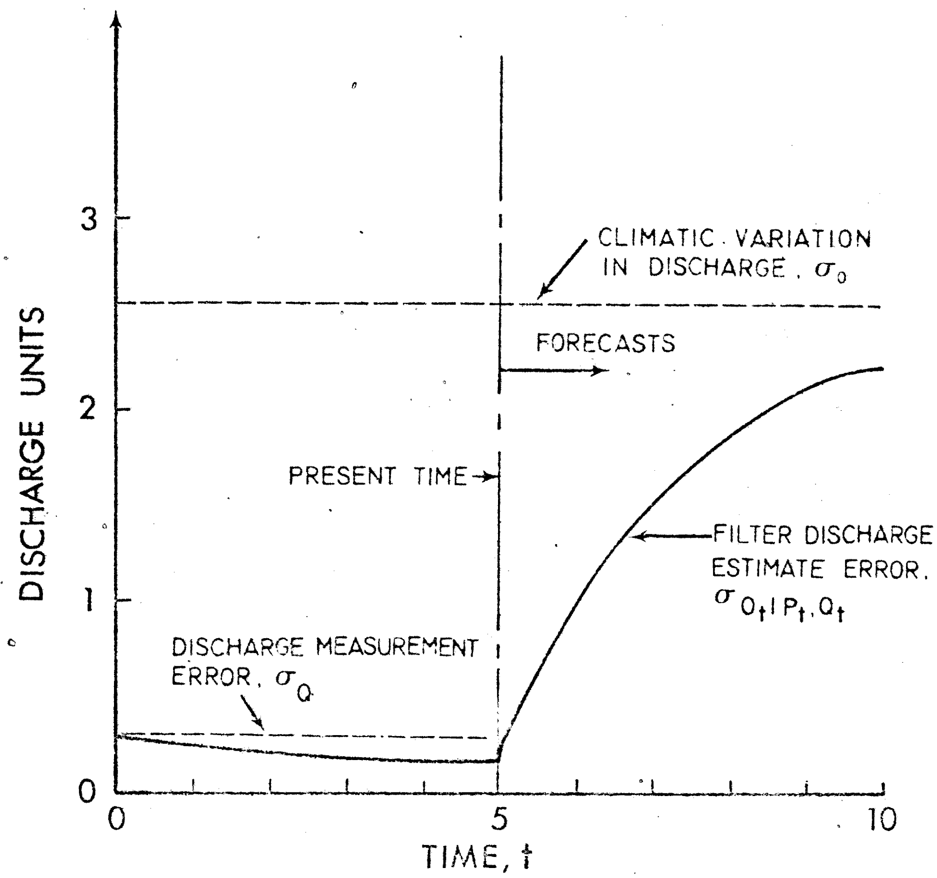

FIGURE 5. CHANGE WITH TIME IN STANDARD DEVIATION OF FILTER DISCHARGE ESTIMATION ERROR, $\hat{O}_t - O_t$

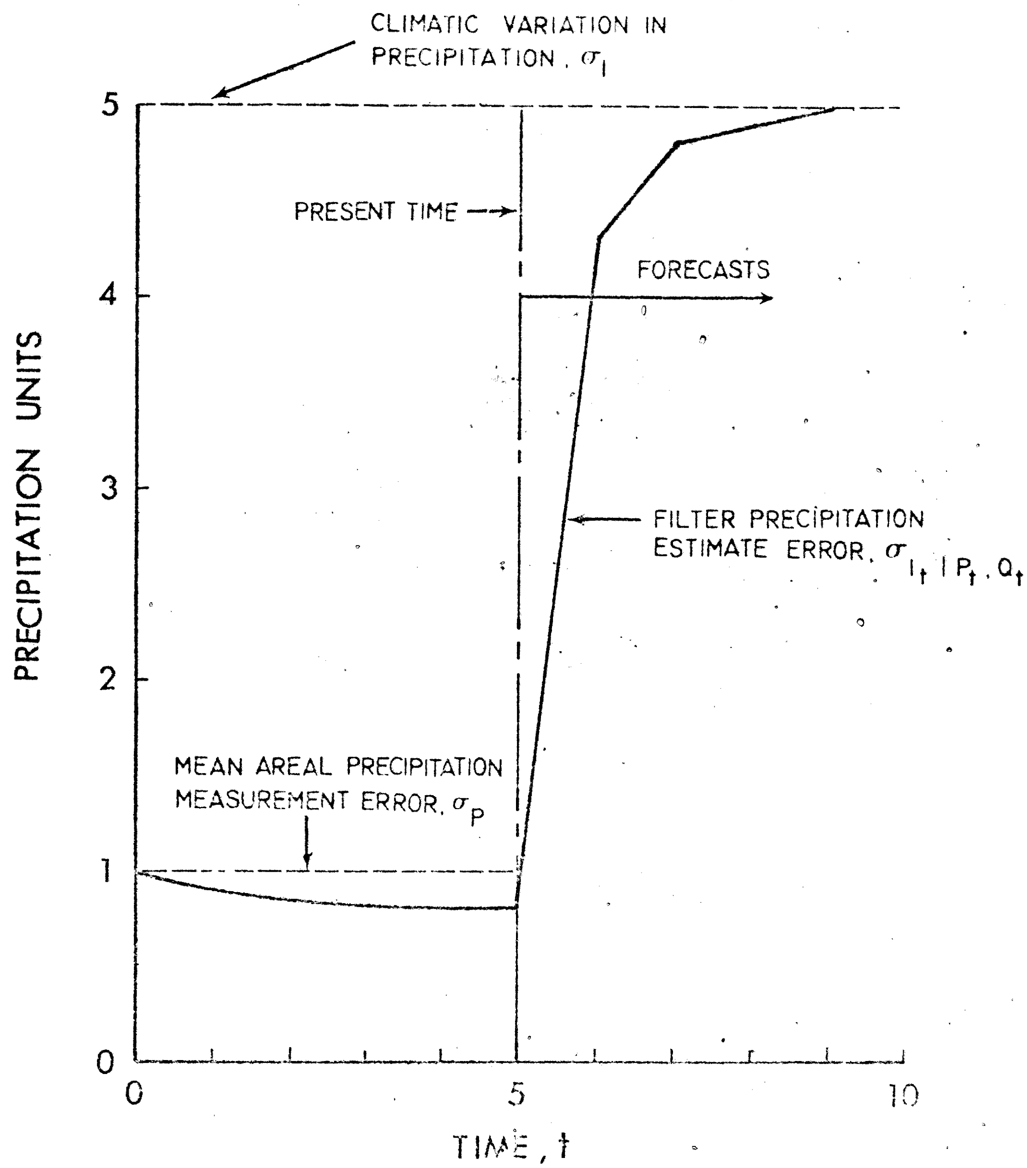

FIGURE 6. CHANGE WITH TIME IN STANDARD DEVIATION OF FILTER PRECIPITATION ESTIMATION ERROR, $\hat{I}_t - I_t$

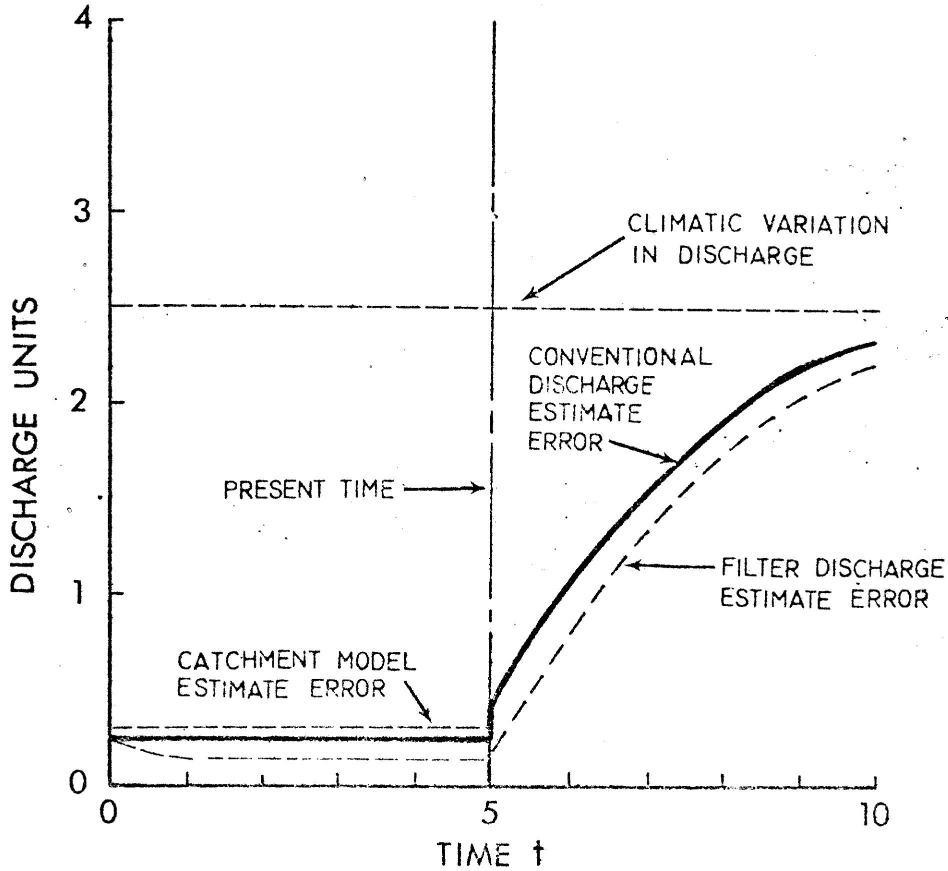

FIGURE 7. CHANGE WITH TIME IN STANDARD DEVIATION OF CONVENTIONAL DISCHARGE ESTIMATION ERROR, $\hat{Q}_t - Q_t$

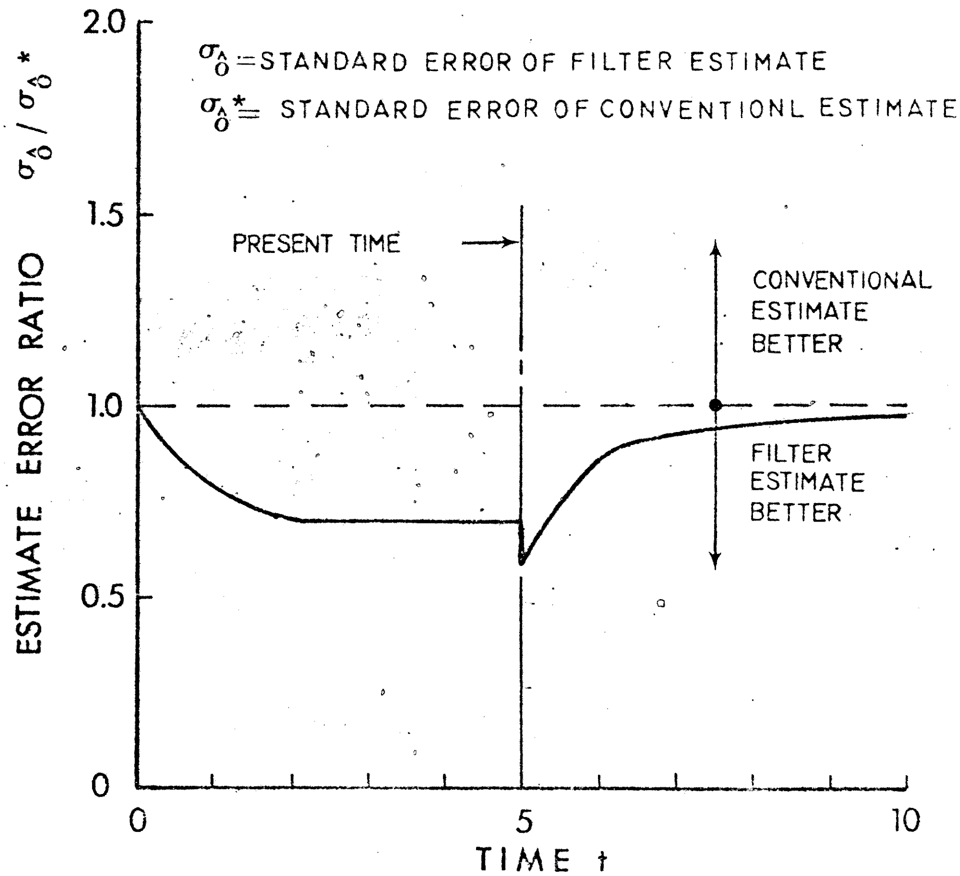

FIGURE 8. RATIO OF FILTER DISCHARGE ESTIMATION ERROR TO CONVENTIONAL ESTIMATION ERROR