Journal of Hydroscience and Hydraulic Engineering
Vol. 9, No. 2, January, 1992, 73-80

THE STYLE OF ON-LINE RAINFALL-RUNOFF FORECASTING INFORMATION

By

T. TAKASAO
M. SHIIBA
C.Y. LIU

Department of Civil Engineering, Kyoto University,
Kyoto, 606, Japan

and

Y. KANBAYASHI

NEWJEC Inc., Tokyo Office,
1-3-3, Uchisaiwai-cho, Chiyoda-ku, Tokyo, 100, Japan

SYNOPSIS

Although the methodology of on-line rainfall-runoff forecasting has been basically established, there are still many important problems when it is actually implemented in river management offices.

This paper especially focuses on the idea that on-line rainfall-runoff forecasting information should be presented in a style which easily weaves into existing management systems, and proposes a new algorithm of on-line forecasting to achieves this.

This method can give information on the whole hydrograph including the probability that the hydrograph peak has already occurred, the probability distribution of the total volume of future discharge, the probability that the flood will surpass the warning water level, and the probability distribution of the time to peak of discharge and the peak volume.

INTRODUCTION

Although the methodology of on-line rainfall-runoff forecasting has been basically established, there are still many important problems when it is used in actual river management offices. This paper especially focuses on the point that on-line rainfall-runoff forecasting information should be given in a style which easily weaves into existing management systems. Also, this paper introduces a new algorithm of on-line forecasting for this purpose.

The authors have so far developed some methods of on-line runoff forecasting so far with the following hypothetical framework:

1. Given the forecasted values and forecast error covariance matrices of future rainfall series at every time step.

2. Forecast the runoff discharge and give the forecasted value's error variance with some constant lead time, for example, one or two hours.

However, rainfall forecasting information is not always provided in the above mentioned format. In most cases, as Takeuchi (1) points out, it is given in a gross statement such as "100mm – 200mm rainfall is expected from tonight to early in the tomorrow morning". Moreover, although the methods developed under the above framework can give the designated time steps ahead information, they cannot give information regarding the whole hydrograph. For instance, these methods are not useful to answer the following questions:

- Whether the peak has already passed or not? What is the probability that the peak has passed?
- How much is the total volume of future discharge?
- What is the probability that the flood will surpass the warning water level from now on in the future?
- What are the expected values or the variances of the peak discharge and the time to peak? What are the probability distributions of these quantities?

However as far as the authors have investigated, it is the answers to these questions that are needed for managing flood-control dams and flood warning. It is desirable, regarding operations of usual dams, that release be made according to a comparison between the current empty volume of the dam and the forecasted future total flood volume, and it is important to obtain information on such items as the time to peak of the inflow hydrograph and the probability that the flood will surpass the warning water level. Using the on-line runoff forecasting algorithm presented by Takara *et al.* (3), we can also obtain information about the future hydrograph to some extent when the lead-time of forecasting is prolonged, because there is not a limit for the forecasting lead-time. However, we cannot obtain probabilistic information on quantities that are defined relating the whole hydrograph, such as the time to peak, the peak discharge and the total discharge volume.

In this paper, after a modification of the fundamental frameworks which were described above, a new runoff forecasting method is proposed for providing information which can be more easily and realistically used.

LINEAR LEAST-SQUARES ESTIMATION AND THE PREREQUISITE FORMULATIONS OF RAINFALL-RUNOFF FORECASTING

Linear least-Squares Estimation

Let X and Z be jointedly distributed n-dimensional and p-dimensional random vectors, respectively. If the value of Z is obtained, we can estimate the value of X, because X depends on Z. The estimate equation $X^*[Z]$ which gives the minimum value of the mean square estimate error among all linear equations of Z, and its estimate error variance $Var\{X - X^*[Z]\}$ can be expressed as follows:

$$X^*[Z] = \overline{X} + \Sigma_{XZ}\Sigma_{ZZ}{}^{-1}(Z - \overline{Z}) \tag{1}$$

$$Var\{X - X^*[Z]\} = \Sigma_{XX} - \Sigma_{XZ}\Sigma_{ZZ}{}^{-1}\Sigma_{XZ}{}^{\mathrm{T}} \tag{2}$$

where $\overline{X} = E(X)$ denotes the mean vector; $\Sigma_{XZ} = Cov\{X, Z\}$ denotes the variance matrix; the right upper T denotes transposition.

Consider an n_1-dimensional random vector X_1 which is determined by

$$X_1 = \Phi X + b + w \tag{3}$$

where w is a random vector which is uncorrelated with X and Z, and has mean 0 and variance matrix R; Φ is an $n_1 \times n$-dimensional non-random matrix; b is an n_1-dimensional non-random vector. In terms of Eqs. 1 and 2, the linear least-squares estimate of X_1 and its estimate error variance can be expressed as

$$X_1{}^*[Z] = \overline{X_1} + \Sigma_{X_1Z}\Sigma_{ZZ}{}^{-1}(Z - \overline{Z}) = \Phi X^*[Z] + b \tag{4}$$

$$Var\{X_1 - X_1{}^*[Z]\} = \Sigma_{X_1X_1} - \Sigma_{X_1Z}\Sigma_{ZZ}{}^{-1}\Sigma_{X_1Z}{}^{\mathrm{T}} = \Phi\, Var\{X - X^*[Z]\}\Phi^{\mathrm{T}} + R \tag{5}$$

Especially, if Z, X and w are normally distributed, then X_1 is also normally distributed, and the conditional distribution given observation Z is normal with mean $X_1^*[Z]$ and variance $Var\{X_1 - X_1^*[Z]\}$, and hence, the probabilistic calculation about X_1 is possible. Even if these requirements are not strictly satisfied, the calculation can be still made under the assumption that they are normally distributed. It may seem unappropriate to assume that a quantity relating to a rare phenomenon like a flood complies satisfactorily with normal distribution. However, the assumption of the distribution being normal is still realistic, because we make forecasting calculation while a storm or a flood is in progress, and hence the phenomenon is not rare one in that occasion.

The Prerequisite Conditions of Rainfall-Runoff Forecasting

By defining a state vector appropriately, most of the runoff models which have been developed so far take the following state-space form:

$$\text{State Equation} \quad dx_i(t)/dt = f_i(x(t), r), \qquad i = 1, \ldots, N_x \tag{6}$$

$$\text{Output Equation} \quad y(t) = g(x(t)) \tag{7}$$

where $x(t)$ is an N_x-dimensional vector; $x_i(t)$ is the ith component of $x(t)$; r is the rainfall intensity; $y(t)$ is the discharge at time t; f_i and g are scalar functions.

If we introduce a modeling error term and an observation error term to the above model, we can obtain a stochastic runoff model described by the following equations.

$$\text{State Equation} \quad dx_i(t)/dt = f_i(x(t), u_k) + G_{x_i}p(t), \quad k-1 < t < k, \quad i = 1, \ldots, N_x \tag{8}$$

$$dp_i(t)/dt = -(1/\tau_i)p_i(t) + v_{p_i}(t), \quad i = 1, \ldots, N_p \tag{9}$$

$$\text{Output Equation} \quad y(t) = g(x(t)) + G_y p(t) \tag{10}$$

where $p(t)$ is an N_p-dimensional noise vector representing the model error and the observation error; the ith component $p_i(t)$ is a first-order exponentially correlated process with time constant τ_i and varince $\sigma^2_{p_i}$; $v_{p_i}(t)$ is a white, zero mean, continuous process noise; G_{x_i} is the ith row vector of the $N_x \times N_p$-dimensional non-random matrix G_x; G_y is an N_p-dimensional row vector. Moreover, it is assumed that the rainfall intensity r takes a constant value u_k within an unit time interval, and the error due to disregarding the variation of rainfall intensity within an unit time interval is incorporated to the error vector $p(t)$.

Through the use of the above formulation, the prerequisite conditions of runoff forecasting are:

- We filter the state vector $(x(t), p(t))$ by using the obtained observation information at every time step, and take the filtered estimate values as initial conditions to predict the future state values and outputs. (In the following, we suppose that $t = 0$ represents the current time and we are given the estimate values and the estimate error variance matrix of the state vector $(x(0), p(0))$ which were filtered with past observation information.)
- We suppose that the forecasted value $\hat{\underline{u}}$ and its forecast error variance matrix $R_{\underline{u}} = Var\{\underline{u} - \hat{\underline{u}}\}$ of future M time units rainfall intensities $\underline{u} = (u_1, u_2, \ldots, u_M)$ are available, where M is supposed to be a rather large value which can contain the whole rainfall duration. If the probability models from recent researches (ref. (4)) about precipitation fields are fully developed, then the mean value and variance matrix obtained from these probability models at the beginning of rainfall can be used as the information necessary here. Forecast information of short time intervals from radar rain gauges or general information like "From tonight to early in the tomorrow morning ..." (this case will be discussed later), eventually acts as additional information to be joined with $\hat{\underline{u}}$ and $R_{\underline{u}}$.
- Using some linearization method and difference approximation methods, we can transfrom Eqs. 8 and 9 into the following discrete-time linearized equations.

  $$\mathbf{x}((j+1)\Delta t) = A_j \mathbf{x}(j\Delta t) + B_j + v_{xj}, \quad j = 0, 1, \ldots, M \times K - 1 \tag{11}$$

  $$y(k) = A_{yk}\mathbf{x}(k) + B_{yk} + v_{yk}, \quad k = 1, \ldots, M \tag{12}$$

  where each unit time is divided into K equal intervals with duration $\Delta t = 1/K$; $\mathbf{x}(t) = (x(t)^{\mathrm{T}}, p(t)^{\mathrm{T}}, \hat{\underline{u}}^{\mathrm{T}})^{\mathrm{T}}$ is the extended $N_{xpu}(= N_x + N_p + M)$-dimensional state vector; v_{xj} is the N_{xpu}-dimensional noise vector with mean 0 and variance matrix Q_j; v_{yk} is a noise with mean 0 and variance Q_{yk}; A_j, B_j, A_{yk}, B_{yk} are non-random coefficient matrices or column vectors.

  Here, it should be noted that, in order to avoid treating infinite dimensional quantities, we do not consider the problem of forecasting all the values of $y(t)$ during $0 < t \leq M$, but instead we consider the values at discrete times $y(k)$, $k = 1, 2, \ldots, M$.

HANDLING RAINFALL FORECASTING INFORMATION

Providing rainfall forecasting information of short time intervals using radar rain gauges has been tried with some degree of accuracy (Ref. (6)). If, besides the rainfall probability model given at the beginning of the rainfall, the rainfall forecast information with short time intervals is provided, it can be handled as additional information. For example, suppose that in addition to previously obtained information $\hat{\underline{u}}$ and $R_{\underline{u}}$, the one hour ahead forecast value $\hat{u}_1{}'$ of rainfall u_1 and its forecast error variance $Var\{u_1 - \hat{u}_1{}'\}$ were given. In this case, treating this information as if we had made an observation of the future rainfall $\hat{\underline{u}}$ through the observation equation $y_{obs} = u_1 + \epsilon$ and had gotten the observation value $\hat{u}_1{}'$, where ϵ is a noise with a mean of 0 and a variance of $Var\{u_1 - \hat{u}_1{}'\}$, we filter the observation value and update $\hat{\underline{u}}$ and $R_{\underline{u}}$.

As for the problem of incorporating forecasting information like "About 100mm rainfall is expected from tonight to next early morning", one can consider it as a filtering problem with some hypothetical observation information in the following manner.

- Let L_1 and L_2 denote "tonight" and "early in the tomorrow morning", respectively. Suppose that

$$\sum_{k=L_1}^{L_2} u_k + v_1 = y_{obs1}; \qquad \sum_{k=L_2+1}^{M} u_k + v_2 = y_{obs2} \tag{13}$$

  are observation equations in which v_1, v_2 are observation noises.

- Let observation value y_{obs1} be 100mm, y_{obs2} be a small value, eg. 5mm because it will not rain after "early in the tomorrow morning". Expressing the uncertainty of the informaiton by the variances of v_1 and v_2, one can filter the observation values and update the estimate value $\hat{\underline{u}}$ and $R_{\underline{u}}$ of the future rainfall intensity.

Moreover, in order to take the uncertainty of the rainfall period into the account, we can suppose p_1, ..., p_I are the probabilities of rainfall periods ($L_1(i)$, $L_2(i)$), $i = 1, \ldots, I$, then we can calculate the mean rainfall period by using the posterior probability distribution with these probabilities p_i as weights. When adding new information on rainfall to the rainfall information, one should note that, in the above treatment, the noises of observation equations are supposed to be uncorrelated with prior probability distributions of the values to be estimated.

ALGORITHM FOR FORECASTING THE STOCHASTIC STRUCTURE OF RUNOFF SERIES

In order to obtain the joint probability distribution of the runoff series $y(1), \ldots, y(M)$, we propose an algorithm which can give the estimate values and their estimate error covariance matrix of $y(1), \ldots, y(M)$ at the last step of the forecasting calculation. This algorithm differs from the algorithms which the authors have presented so far in that it gives not only the variances of estimate errors but also the covariances between estimate errors at different times. For this purpose, we extend further the state vector $\mathbf{x}(t)$ and define it as follows.

$$\mathbf{X}(t) = (\mathbf{x}(t)^{\mathrm{T}}\ y(1)\ \cdots y(k))^{\mathrm{T}}, \qquad 0 \le k \le t < k+1 \le M \tag{14}$$

By this extension, $\mathbf{X}(t)$ becomes an $(N_{xpu}+k)$-dimensional column vector. Moreover, it follows from Eq. 11 that the transition of the state vector $\mathbf{X}(t)$ from $t = j\Delta t$ to $t = (j+1)\Delta t$ can be expressed as

$$\mathbf{X}((j+1)\Delta t) = \begin{bmatrix} A_j & 0 \\ 0 & I_k \end{bmatrix} \mathbf{X}(j\Delta t) + \begin{bmatrix} B_j \\ 0 \end{bmatrix} + \begin{bmatrix} v_{xj} \\ 0 \end{bmatrix} \tag{15}$$

where k is an integer such that $k \le j\Delta t < k+1$. Moreover, at time k, the state vector is extended as

$$\mathbf{X}(k+) = \begin{bmatrix} I_{N_{xpu}+k-1} \\ A_{yk} \end{bmatrix} \mathbf{X}(k) + \begin{bmatrix} 0 \\ B_{yk} \end{bmatrix} + \begin{bmatrix} 0 \\ v_{yk} \end{bmatrix} \tag{16}$$

In Eqs. 15 and 16, I is an unit matrix and the subscript of I denotes its dimension. Because these transition equations have the same form as Eq. 3, using Eqs. 4 and 5, the estimate values and the estimate error covariance can be calculated sequentially.

The previous calculation will be repeated until the estimate value $\widehat{\mathbf{X}}(M+)$ of $\mathbf{X}(M+)$ and its estimate error variance $Var\{\mathbf{X}(M+) - \widehat{\mathbf{X}}(M+)\}$ are obtained. By taking out the latter M dimensional vector of $\widehat{\mathbf{X}}(M+)$ and the lower right M-dimensional square matrix of $Var\{\mathbf{X}(M+) - \widehat{\mathbf{X}}(M+)\}$, we get the estimate series of $y(1), \ldots, y(M)$ and their estimate error covariance. If the probability distribution of $y(1), \ldots, y(M)$ is normal, then the joint probability distribution of $y(1), \ldots, y(M)$ can be determined, because the normal distribution is identified by the mean vector and the covariance matrix.

As a result, the following calculations can be performed:

- calculation of $P_{rob}\{y(K) < Q_{p0},\ k = 1, \ldots, M\}$ where Q_{p0} is the maximum discharge up to the current time,
- calculation of the probability distribution of $\sum_{k=1}^{M} y(k)$,
- calculation of $P_{rob}\{y(k) > Q_w$ for some k such that $1 \leq K \leq M\}$ where discharge Q_w corresponds to the warning water level, and
- calculation of the probability distributions of $y(k_p) = \max\{y(1), \ldots, y(M)\}$ and k_p.

If these calculations cannot be carried out analytically, we can still get an approximate value using random simulations.

PRELIMINARY EXAMINATION BY NUMERICAL SIMULATION

In order to test the validity of the algorithm proposed in the previous section, a numerical simulation was made on a nonlinear reservoir series model as follows:

$$\begin{aligned} ds_1/dt &= -K_1 {s_1}^m + \alpha_1 r + p_1(t) \\ ds_2/dt &= -K_2 {s_2}^m + \beta_2 K_1 {s_1}^m + \alpha_2 r + p_2(t) \\ &\cdots \\ ds_n/dt &= -K_n {s_n}^m + \beta_n K_{n-1} {s_{n-1}}^m + \alpha_n r + p_n(t) \\ y(t) &= A K_n {s_n}^m + w \end{aligned} \tag{17}$$

where the transition equation of n-dimensional noise vector $p(t)$ is the same as Eq. 9; w is a white normal error series; $s_1, \ldots, s_n$ are storage heights of the reservoirs; A is the basin area; $K_1, \ldots, K_n$, $\alpha_1, \ldots, \alpha_n$, $\beta_2, \ldots, \beta_n$ are constants.

The series of expected discharge, the probability distributions of the peak discharge and the time to peak were calculated by the method presented in this paper, and is plotted in Figs. 1 ~ 4. Linearization by Taylor series expansion around the estimate values was used. After calculating the expected values of runoff series and their covariances, we generated normal random numbers which have those expected values and covariances. In Figs. 1 ~ 4, we also plotted the results of the Monte Carlo simulation in which pseudo-random numbers were generated for all random noises in Eq. 17 and then the equation was solved with these generated values. We generated the same number of discharge series, but the calculation time spent by the method proposed in this paper was one thirtieth of the time spent by the Monte Carlo simulation.

For this calculation, we assumed $n = 5$, $m = 5/3$, $A = 120\text{km}^2$, $K_i = 22^{-5/3}\text{mm}^{-2/3}\text{h}^{-1}$, $\beta_i = i/(i+1)$, $\tau_i = 5\text{h}$, ${\sigma_{p_i}}^2 = 0.4(\text{mm/h})^2$, $i = 1, \ldots, 5$, and also gave appropriate values to the variance of w, the expected values and their error variance of rainfall r, the initial estimate values and the estimate error variance of the state vector.

The probability distributions of the peak discharge and the time of peak obtained by using the method of this paper show little difference from those obtained with the Monte Carlo simulation. Therefore, it was verified that stochastic forecasting of temporal runoff variation is possible by the method proposed in this paper, however the influence of the variation of rainfall intensity with time and its covariance should be examined in detail.

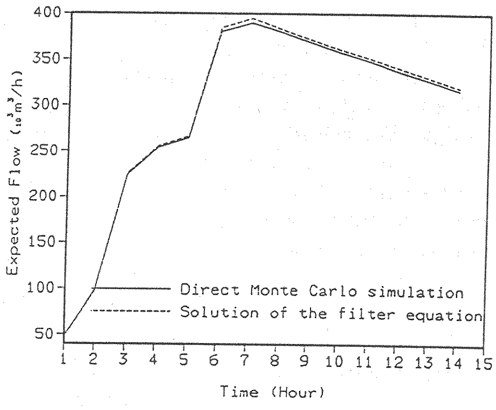

Fig. 1 The series of expected runoff discharge

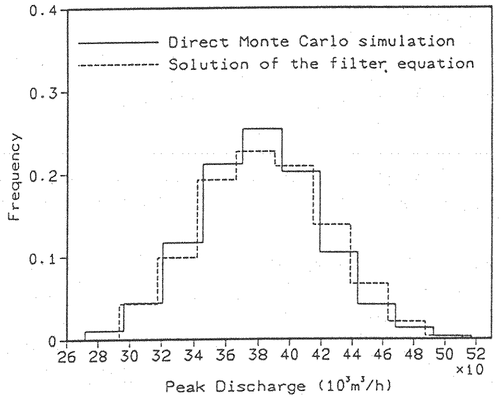

Fig. 2 The probability distribution of the peak discharge

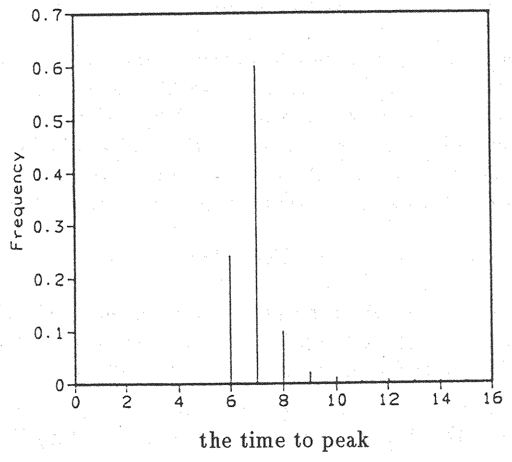

Fig. 3 The probability distribution of the time to peak obtained by the Monte Carlo simulation

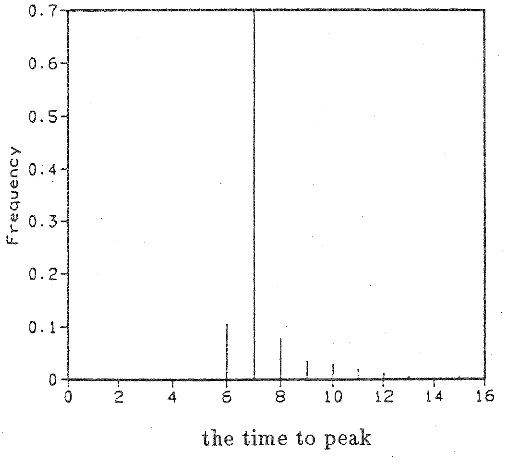

Fig. 4 The probability distribution of the time to peak obtained from the filter equation

CONCLUSION

In this work, we examined the practical problem of providing the forecasting information of rainfall and runoff in a style convenient for real management work, and proposed a new on-line forecasting algorithm. At the same time, its validity was demonstrated by a simple numerical simulation. In this simulation, although only the probability distributions of the peak discharge and the time to peak were calculated, the probability distributions of total discharge volume, and the probability of surpassing the warning water level within a designated time period can be also calculated in the same way.

REFERENCES

1. Takeuchi, K. and N. Hayashi : Investigating practical use of preliminary release type with rainfall prediction, Scientific Research Report of 1989 (Principal Investigator: T. Takasao), Study on Real-Time Forecasting and Control of Storm Runoff, pp.68-74, 1989.
2. Takasao, T., M. Shiiba and T. Hori : Fundamental study on an expert system for flood disaster prevention, The Annuals of the Disaster Prevention Research Institute Kyoto University, No.31 B-2, pp.357-368, 1988.
3. Takara, K., T. Takasao and M. Shiiba : Practical techniques in stochastic real-time prediction of flood runoff, Proceedings of the 28th Japanese Conference on Hydraulics, pp.415-422, 1984.
4. Rodriguez-Iturbe, I. and P.S. Eagleson : The mathematical models of rainstorm events in space and time, Water Resources Research, 23(1), pp.181-190, 1987.
5. Takasao, T., M. Shiiba and N. Tomisawa : Building of a runofff prediction system on the basis of the statistical second-order approximation theory, The Annuals of the Disaster Prevention Research Institute Kyoto University, No.27 B-2, pp.255-273, 1984.
6. Kanbayashi, Y.: A study on probability run-off forecasting method for flood control using radar rain gauge, Proceedings of JSCE, No.411/II-2, pp.169-175, 1989.

(Received August 2, 1990; revised December 20, 1991)