Embedding Kalman filter into a Distributed Hydrological Model

Sunmin KIM[1)], Yasuto TACHIKAWA[2)], and Kaoru TAKARA[2)]
1) Graduate school of Urban and Environment Engineering, Kyoto University
Yoshida Honmachi, Sakyo-ku, Kyoto, 606-8501, JAPAN
E-mail: sunmin@rdp.dpri.kyoto-u.ac.jp
2) Disaster Prevention Research Institute, Kyoto University
Uji, Kyoto 611-0011, JAPAN

Abstract

This paper describes a methodology for embedding a Kalman filter into a distributed hydrological prediction system for real-time flood prediction. In most cases it is complicated to formulate a Kalman filter algorithm in the system structures of distributed models. However, the Kalman filter theory is successfully coupled with a distributed hydrological model to update spatially distributed state variables by using several techniques proposed here. To acquire the total water storages of a basin from discharge observations at the outlet, a Q-S curve is used as an observation equation. The ratio method is introduced to update the distributed storage amount of a basin, maintaining the spatially distributed water storage pattern of the basin by multiplying the distributed state variables in the model by a specified ratio. A Monte Carlo simulation is adopted to predict state variables and error covariance propagations. The methodology for deciding system noise is also discussed. A distributed model coupled with the Kalman filter theory gives updated simulation results with improved forecasting accuracy.

1. INTRODUCTION

To obtain an accurate prediction in a real-time rainfall-runoff simulation it is essential to set model parameters effectively and to give properly assessed initial state variables. Modern real-time observation systems make it possible to access various hydrological data while real-time simulation is carried out. Prediction accuracy should be improved when the parameters or state variables are updated with real time observed data. Also, the uncertainty caused by improper model parameters, initial state variables and input data can be reduced if filtering theory, such as the Kalman filter, is incorporated in a hydrological model.

R.E. Kalman (1960) published his famous paper describing a recursive solution, which was later named as the Kalman filter, to discrete data linear filtering problems. Having potential for broader use, Kalman filter has been enhanced as Extended Kalman filter for nonlinear systems. The Kalman filter is an optimal recursive data processing algorithm to estimate the state variables for minimizing the error statistically. It combines all available observation data, plus prior knowledge about the system and measuring devices, to produce an estimate of the desired variables in such a manner that the error is minimized statistically (Maybeck, 1979). A more detailed description about Kalman filter theory can be found at Jazwinski (1970) and a good discussion of the filter with several application cases to the hydrological system is given by Bras and Rodriguez-Itulbe (1985).

Since Hino (1974) initially adapted the Kalman filter theory to a hydrological system, numerous studies have been carried out to use the filter theory in the field of hydrology. Takaso *et al*. (1989) describes real-time flood forecasting based on a stochastic state-space formulation of rainfall-runoff systems coupled with the Kalman filtering-prediction theory and its application. In the research, the storage function method was used to couple with the filter taking the storage amounts of sub-basins as the state vector. Lee and Singh (1999) showed upgraded simulation results of the tank model when the state vector of the Kalman filter is composed of the model's parameters. The storage function method and the tank model are often used lumped models in Japan, Korea and many other countries for flood forecasting and watershed modeling. While the Kalman filter has been applied to many lumped models for better simulation or more accurate forecasting, it has hardly ever been applied to distributed hydrological models. One of the main reasons is that unlike lumped models, in most cases it is complicated to formulate the Kalman filter algorithm in the system structures of distributed models.

A large number of state variables based on a fine grid cell hydrologic system also makes it harder to apply the Kalman filter.

In this research, to avoid the computational burden for updating each state variable, several techniques are introduced for applying the Kalman filter to a distributed hydrological model. The Q-S curve which is determined under steady state assumption on a study basin is used for the observation equation to update the simulated total storage amount of the basin with discharge observations. To consider the spatial pattern of the updated storage amount in every grid-cell of a distributed hydrological system, the ratio method is adopted. The ratio method is used for updating spatially distributed storage amount in the model by multiplying by a ratio calculated from the updated total storage amount and the simulated storage amount. For the prediction algorithm, stochastic analysis is adopted to predict state variables and error covariance of the next updating step. Monte Carlo simulation is an effective technique to analyze the effect of error covariance propagation. The methodology for deciding system error variance is also discussed.

2. COUPLING OF CDRMV3 WITH KALMAN FILTER

The study mainly focuses on a coupling method of the Kalman filter to the Cell based Distributed Runoff Model Version 3 (CDRMV3, http://fmd.dpri.kyoto-u.ac.jp/~flood/product/cellModel/cellModel.html). The state variable to be updated is the total storage amount and its spatial distribution is calculated using water depth at all computational nodes in the model. The water depths are easily converted to storage amount by multiplying the cell area. The parameters of the CDRMV3 are calibrated before applying the Kalman filter and do not change when state variables are updated. Uncertainties caused by systems and observations are considered in the error covariance of the filter, though uncertainty caused by rainfall forecasting is not accounted for in this study. Radar observed rainfall data which is calibrated by ground gauges are used as forecast rainfall data.

2.1 Brief Model Description of CDRMV3

CDRMV3 is a one dimensional physically based distributed hydrologic model developed at Flood Disaster Research Laboratory of Disaster Prevention Research Institute, Kyoto University. The model solves the Kinematic wave equation using Lax Wendroff scheme on every node in a cell (Kojima *et al.*, 2003). Discharge and water depth propagate to the next cell according to a predefined routine order determined in accordance with DEM data. An advantage of the CDRMV3 is that the stage-discharge relationship of each cell reflects the topographic and physical characteristics of its own cell. Specified stage-discharge relationship, which incorporates saturated-unsaturated flow mechanism, is included in each cell (Tachikawa *et al.*, 2004). Because of the variable slope and roughness coefficient, each cell has its own relationship:

$$q(h) = \begin{cases} v_C d_C \left(h/d_C\right)^{\beta}, (0 \le h < d_C) \\ v_C d_C + v_a\left(h - d_C\right), (d_C \le h < d_S) \\ v_C d_C + v_a\left(h - d_C\right) + \alpha\left(h - d_S\right)^m, (d_S \le h) \end{cases} \tag{1}$$

where $v_C = k_C i$, $v_a = k_a i$, $\alpha = \sqrt{i/n}$.

k_c : hydraulic conductivity of unsaturated layer
k_a : hydraulic conductivity of saturated layer
n : roughness coefficient
i : slope of grid

The stage-discharge relationship is expressed by three equations corresponding to the water levels divided into three layers (see Equation 1). When the water depth h is lower than the depth of unsaturated layer ($0 \le h < d_c$), flow is described by Darcy's law with a degree of saturation, $(h/d_c)^{\beta}$, and a hydraulic conductivity k_c. If the h increases, flow from the saturated layer is considered with a different hydraulic conductivity k_a of saturated layer. After the water depth is greater than the soil layer d_s, overland flow is added by using the Manning's equation. According to this mechanism, the equations between discharge per unit width *q* and water depth *h* are formulated. Figure 1 shows the graphical relationship between *q* and *h*. More detail on the specified state-discharge relationship and the model structure can be found in Tachikawa *et al*. (2004).

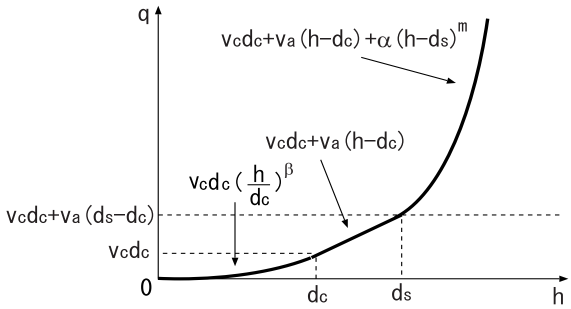


Figure 1. Relationship between unit width discharge and water depth in the CDRMV3

The model is applied to the Kamishiiba basin (211km^2) of Kyushu area. Four different flood types of the basin are selected for this study. Flood period and maximum discharge of the four events are shown in Table 1.

Table 1. Flood events used in the study

EVENT	Flood Term	Max Discharge
Event 979	Sep 15~19, 1997	1203.0 m^3/s
Event 996	Jun 24~Jul 3, 1999	210.0 m^3/s
Event 998	Aug 1~7, 1999	489.0 m^3/s
Event 999	Sep 22~27, 1999	644.0 m^3/s

2.2 Measurement Update Algorithm

In the measurement update algorithm of the Kalman filter, an observation equation which specifies a relation between observed data and state values is necessary. The observation vector y_k can be described as a linear combination of a state vector x_k as Equation 2. The observations are affected by white noise w_k which has a covariance matrix R_k. The $m \times n$ matrix *H* relates the state vector to the observation. In the measurement update algorithm, the state vector *x(k|k-1)* and error covariance *P(k|k-1)* which are estimated through system at time step *k-1*, are updated at time step *k* with use of the $n \times m$ matrix *K*. The matrix *K* which is called Kalman gain is chosen to minimize the updated error covariance *P(k|k)*. In the algorithm, ^ indicates estimated value and ^{T} indicates the transpose of a matrix.

Observation equation

$$y_k = H_k x_k + w_k \,; \quad w_k \sim N(0, R_k) \tag{2}$$

Measurement update algorithm

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K_k (y_k - H_k \hat{x}(k|k-1)) \tag{3}$$

$$P(k|k) = P(k|k-1) - K_k H_k P(k|k-1) \tag{4}$$

$$K_k = P(k|k-1) H_k^T (H_k P(k|k-1) H_k^T + R_k)^{-1} \tag{5}$$

The difference, y_k-H_k*x(k│k-1)*, which is called the residual or innovation reflects the discrepancy between the estimated observation H_k*x(k│k-1)* and the actual observation y_k. If the total storage amount to be updated is measured directly, the residuals are easily obtained. However observed quantities are discharge or river stage rather than distributed storage amount. When we think of the storage amount as a state variable, we can easily come up with a nonlinear relationship between discharge and total water storage in a subject basin as shown in Equation 6. In the storage function method, when it is coupled with the Kalman filter, the observation equation is adopted after linearization of Equation 6.

$$S(t) = KQ(t)^P \tag{6}$$

where $S(t)$: total water storage in a basin
$Q(t)$: discharge at the outlet
K, P : constants.

Unlike the storage function method the CDRMV3 does not have general relation between observed discharge and the storage amount. The relationship between the discharge at the outlet and the total amount of storage represents a loop shape as shown in Figure 2, whose shapes are different for each flood event. However, it is still possible to get a relation in a specific case like a steady state condition. After reaching the steady state condition with a given constant rainfall on the subject basin, the total storage amount that corresponds to the given rainfall intensity can be acquired by multiplying cell area with water depths of every cell and sum up these entire amounts. The cell size in this study is 250m×250m. Applying various rainfall intensities, the Q-S curve that is used for the observation equation is obtained as shown in Figure 2. Even though states during a runoff simulation are not steady, the difference between the two curves from the steady state and unsteady state seems acceptable.

The reason for needing the observation equation is to get the matrix *H* in the measurement update algorithm. The conversion matrix *H* relates total storage amount and observed discharge at time step *k*. More specifically, it stands for the gradient of the Q-S curve in accordance with simulation results at updating time step. As only one observation is available in the Kamishiiba basin, *H* is a scalar value in this study.

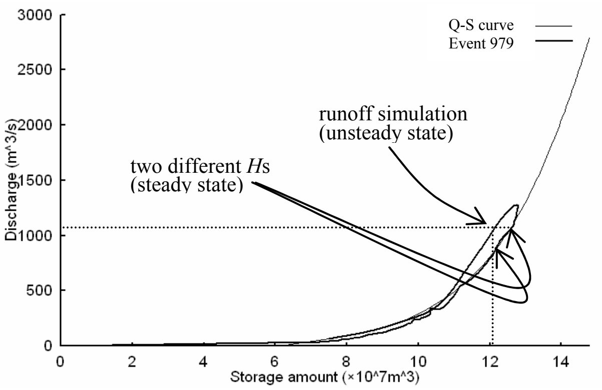


Figure 2. Two different curves between steady and unsteady state

In processing the measurement update algorithm, a couple of problems exist to be considered. These are basically caused by the steady state assumption when making use of the Q-S curve. Because the relationship is determined under the steady state assumption, there is always some departure towards the unsteady state, even if it is small.

At first, two different *H* values could be given at each time step as seen in the Figure 2. At each time step, discharge and total storage amount are obtained from simulation results. Those two values from unsteady state simulation may not match on the Q-S curve. This mismatch often gives two different *H*; one is from the simulated discharge and the other is from the simulated storage amount. However, through several tests, it is checked that those two different *H* values do not make a recognizable difference to the filtered results. For this reason, an arithmetic average of those two *H* values is used in the application of the CDRMV3.

Another problem occurs while getting residuals in the measurement update algorithm. According to the conventional equations of the Kalman filter, the residuals are calculated by the use of the observation equation. As shown in Equation 7, converting state variables needs the *H* value which contains the steady state error described already. On the other hand, residuals can be directly calculated from the difference between observed discharge and simulated discharge as in Equation 8. Filtered results from those two residuals show significant difference. In conclusion, the residuals from directly using simulated discharge gives much better filtered results. More detail discussion is provided in Section 3.

$$(y_k - H_k \hat{x}(k|k-1)) \Rightarrow (OQ_k - H_k \times SS_k + A_k) \quad (7)$$

$$(y_k - H_k \hat{x}(k|k-1)) \Rightarrow (OQ_k - SQ_k) \quad (8)$$

where OQ_k : observed discharge
SQ_k : simulated result of discharge
SS_k : simulated result of total storage amount
H_k: conversion matrix, gradient of the Q-S curve
A_k: optional input for the linearization

After updating the total storage amount through the measurement update algorithm, the updated storage amount should be distributed to each cell in a subject basin. One efficient way to update each cell's storage amount is using a specific ratio calculated from the updated total storage amount and the simulated storage amount. The calculated ratio is applied to all water depths of each cell in the model, which has the same spatial distribution pattern with the simulation result before updating as shown in Figure 3. For example, if the simulated storage amount at a specific time step is 1.03476E+8 m^3 and the updated value is 0.97292E+8 m^3, all water depths of each cell at this time are multiplied by the ration 0.9402, and then the simulation starts again using each cell's updated state variable. This method which is named as the ratio method, offers efficient and effective updating skill of state variables considering its spatial distribution pattern (Kim *et al.*, 2004).

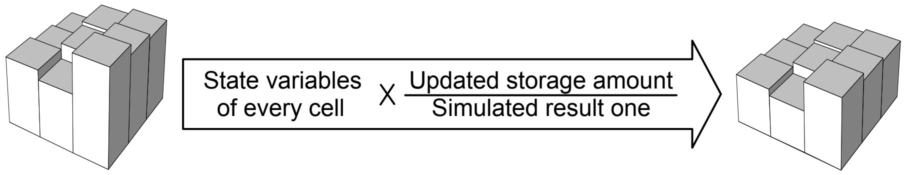


Figure 3. Concept of the ratio method for updating spatially distributed state variables

2.3 Time Update Algorithm

The Kalman filter is an algorithm to optimize the state vector *x* of a discrete time controlled process which is governed by a linear difference equation. The $n \times n$ matrix *F* in the system equation relates the state at previous time step *k* to the state at current step *k+1*. The systems are continuously affected by white noise, v_k, with covariance matrix Q_k respectively. The matrix B_k relates optional control input to the state *x*. The time update algorithm is for projecting forward the current state and error covariance to obtain the estimation for the next time step. The estimated error covariance *P* is a $n \times n$ matrix.

System equation

$$x_{k+1} = F_k x_k + B_k + v_k\,; \;\; v_k \sim N(0, Q_k) \quad (9)$$

Time update algorithm

$$\hat{x}(k+1|k) = F_k \hat{x}(k|k) + B_k \quad (10)$$

$$P(k+1|k) = F_k P(k|k) F_k^T + Q_k \quad (11)$$

In the CDRMV3, a complicated relation exists between the present state variables and the next state variables; in this case the storage amount and its spatial distribution. Each cell responds interdependently to the next step's state variable according to its present state variable and other input data such as rainfall. So it is impractical to formulate the system matrix F_k, which is essential to update the error covariance $P(k+1|k)$, even though it is possible theoretically. Rather than use the conventional concept of the Kalman filter theory as shown in the schematic drawing (a) in Figure 4, the Monte Carlo simulation technique (drawing (b) in Figure 4) is useful to solve this kind of problem.

By the concept of the Monte Carlo simulation, many sets of random variables are generated at time step k; hundreds of storage values were generated in this study. Each random variable is defined by any possible storage amount within a range of probability distribution, $N(X_k, \sigma_k)$; where, $\sigma_k=P_k^{0.5}$.

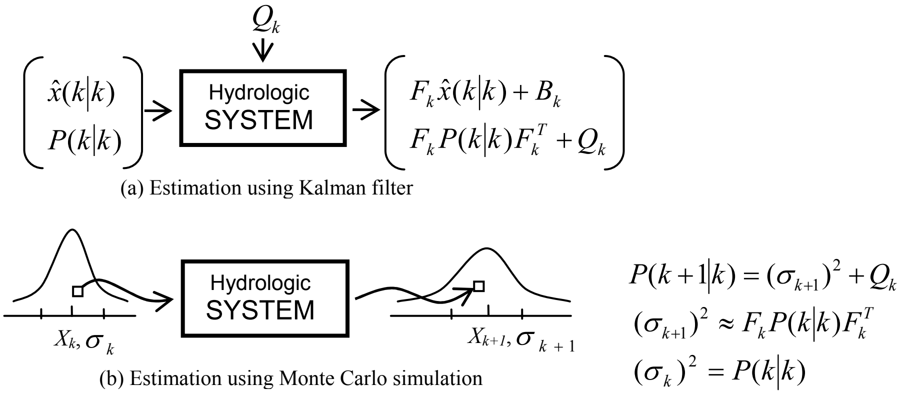


Figure 4. Concepts of time update algorithm

The ratio method is used again at this point to reset the water stage at each cell by multiplying by the ratio of generated storage amount to the mean storage amount X_k. After a simulation repeatedly calculates multiple input sets until the next update time step $k+1$, another probability distribution, $N(X_{k+1},\sigma_{k+1})$, is calculated from the simulated results. Now, the estimated state X_{k+1} is the mean value of the probability distribution and the estimation error covariance $F_kP_kF_k^T$ is regarded as $(\sigma_{k+1})^2$. Adding the system error covariance Q_k completes the error covariance P_{k+1} at time step $k+1$. Estimation error covariance $F_kP_kF_k^T$ means propagation of the error covariance P_k through the simulation, and Q_k stands for a new generated or added system covariance during simulation from time step k to the next time step $k+1$. The newly added system covariance is caused by system structure or new input data such as rainfall. The methodology to determine the system error covariance, Q_k, is discussed in the following section.

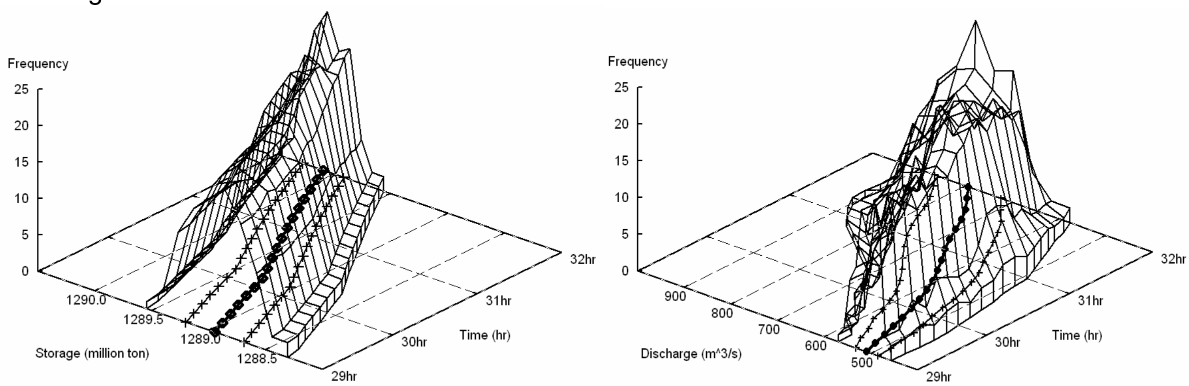


Figure 5. Propagations of probability density of storage amount and discharge (Event 979)

Figure 5 shows an example of the probability distribution propagation during three hours from Monte Carlo simulation results. The three lines on the bottom represent the variation of mean value and standard deviation from the distribution at each calculation time step. While the distributions of storage amount follow the normal distribution pattern, the distributions of discharge which are composed by each simulation results set from the storage amount variables do not always stick to the normal distribution. This phenomenon is caused by the nonlinear relationship between storage and discharge. The different distribution pattern of discharge and storage amount will also affect the form of the error.

2.4 Setting the Observation Noise and System Noise

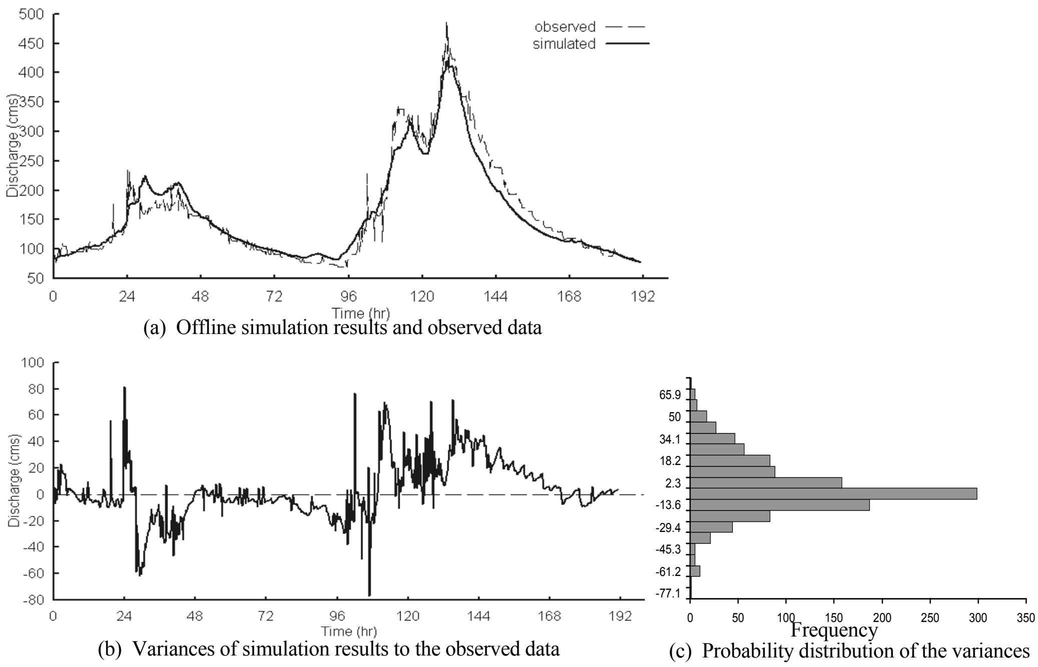
One of the difficulties in applying the Kalman filter is determining the error covariance of the system and the observation. Although the Kalman filter provides an algorithm for better forecasting by updating the state estimates, its success depends on determination of the error covariance which requires proper judgment by the hydrologist. Because the hydrological system is a natural system which varies in time and space, it is impossible to get the true value. This means that it is also impossible to get the error covariance which is based on noises to the true value of hydrological variables. For this reason, error covariance has been assumed when the Kalman filter is applied to hydrological models. The only action we can take is to try to get a reasonable error covariance with the least assumptions.

2.4.1 Observation Noise

The basic assumption of the Kalman filter is that system and observation noises are white and Gaussian. Thus, at any point in time, the noise value is not correlated in time and the probability density curve of noise takes on the shape of a normal bell-shape. This assumption can be justified physically by the fact that a system or observation noise is typically caused by a number of small sources (Maybeck, 1979). It is reasonable to see the observation noise from this point of view. Usually observed values are acquired through a conversion from stage to discharge by using a stage-discharge relation or rating curve. Observation noises are mainly caused by misreading of gauge stage, interpolation during conversion from stage to discharge, and improper stage-discharge relationship. Also it is not difficult to accept that the noise is normally distributed to the true value. The observed data used in this study are acquired from the Kamishiiba dam inflow data. The inflow data are calculated mainly by converting the dam reservoir stage into discharge with considering the release from the dam for various purposes. Since more research is needed to determine the observation noise reasonably, an assumed observation error variance is used in this paper. The assumed error variance is mentioned in each filtered result.

2.4.2 System Noise

If observed data are assumed to be true values, differences between simulation results and the observed data could be regarded as system noise. First, it needs to be examined whether the variances of simulation results to the observed data have a normal distribution. When the offline simulation results from the CDRMV3 are examined, the variances have a normal probability distribution to the observed data as sown in the Figure 6. Under this examination, the RMSE given by Equation 12 could be regarded as a standard deviation of system noise w_k. The mean value and RMSE from the offline simulation results of four different events are shown at Table 2. We can figure out that the mean values of variances are around zero and the RMSEs are around 30m^3/s. Following this analysis, the standard deviation of system noise in terms of discharge is set as 30m^3/s in this study.



(a) Offline simulation results and observed data

(b) Variances of simulation results to the observed data

(c) Probability distribution of the variances

Figure 6. Variances of system error and its probability distribution (Event 998)

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_{S,i} - Q_{O,i})^2}{n}} \tag{12}$$

where $RMSE$: Root Mean Square Error
Q_S : simulated discharge
Q_O : observed discharge
n : number of Q_S or Q_O values

Table 2. The RMSE and mean values of each event

EVENT	RMSE	MEAN
Event 979	33.64 m^3/s	-2.22 m^3/s
Event 996	21.15 m^3/s	-10.80 m^3/s
Event 998	20.85 m^3/s	3.32 m^3/s
Event 999	23.73 m^3/s	-1.02 m^3/s

Then, the discharge RMSE is converted to the error variance of the total storage amount. System noise in terms of discharge can be translated to the noise in terms of storage amount by using the Q-S curve as seen in the Figure 7. The term 'sd' in the figure means the standard deviation of discharge variances, which was previously determined to be 30m^3/s. Three discharges from simulation results at a specific time step will match with three different storage amounts on the Q-S curve. Because of nonlinearity of the relation, the differences between the upper value and lower value to the S_k will be different. Using those two differences of storage amount, $Sup_k – S_k$ and $S_k - Sdn_k$, the system error variance Q_k can be calculated as shown in Equation 13.

$$Q_k = \left|(Sup_k - S_k) \times (S_k - Sdn_k)\right| \tag{13}$$

where Q_k : system error variance at time step k
Sup, S, Sdn : converted storage amounts

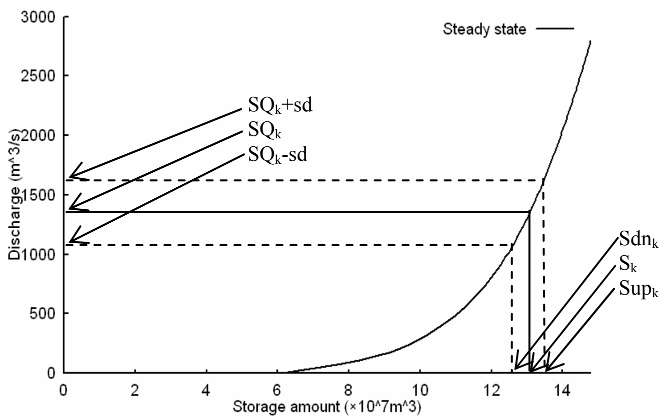


Figure 7. Conversion of noise from discharge to storage amount form

There is one important checking point about the Gaussian assumption of the Kalman filter for a nonlinear system. Because the relationship between discharge and storage amount is nonlinear, if the probability distribution of storage amount is a Gaussian distribution, the distribution of discharge will not follow the normal distribution, and vice versa. The distributions of the variables are no longer normal after undergoing their respective nonlinear transformations. However, this nonlinear effect on probability distribution is not significant in this study. As shown in Figure 5, the probability distribution of discharges, which is calculated from the simulation results according to each storage amount having normal distribution, can be roughly regarded as a normal distribution.

3. ANALYSIS OF RESULTS

3.1 Results from two different residual types

The Kalman filter, a recursive data processing algorithm, is successfully coupled to the distributed hydrological model, CDRMV3. To check the filtering results effectively, the observed data are assumed to have no error. If there is no observation error, which means that the observed data is the true value, the filtered results and observed data should match exactly. When the results are checked, there are discrepancies at the beginning and after the peak of the hydrograph compared to the other part as seen in Figure 8.

Three reasons are considered to explain these discrepancies. The first one is because of the steady state assumption of the Q-S curve. When the steady state assumption is made, discharge and storage is expressed as a single-valued function (see Figure 9 (a)). On the other hand, different storage values occurs in unsteady state conditions, even though the discharge is the same as shown in Figure 9 (b). The storage amount at the beginning of the hydrograph is different from the amount at the falling limb or after the peak of the hydrograph. The differences in storage amount, "Storage B-A" at the beginning of the runoff and "Storage C-A" after the peak, cause the under estimation of discharge at the rising limb and the over estimation at the falling limb of the hydrograph.

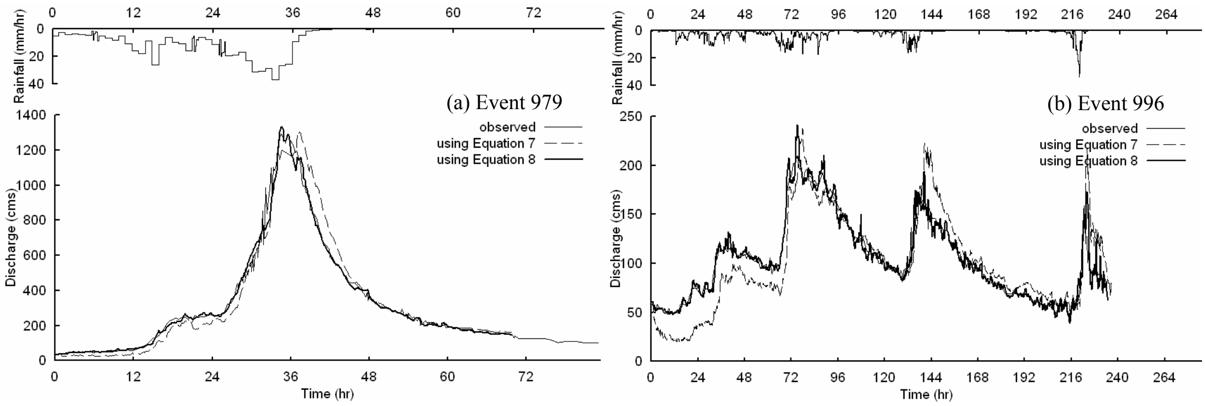


Figure 8. Updated results comparisons from two different residuals

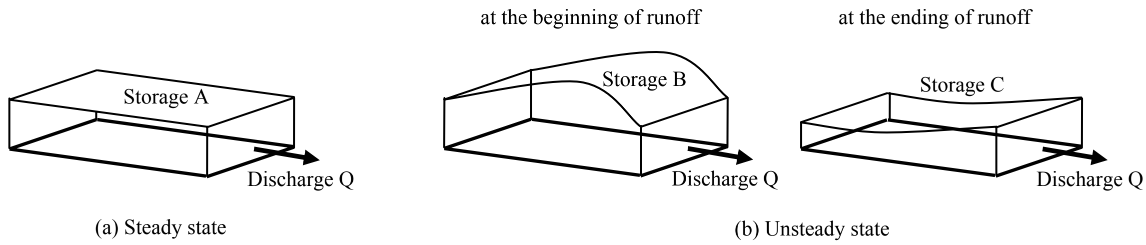


Figure 9. Conceptual storage amount distributions according to the state differences

Secondly, every observation update is given at every an hour while the calculation time step is ten minute. Five simulation results between nearest observation update would make its own hydrograph. When observation updates are carried at every calculation time step, the discrepancies are decreased a lot. The last reason is that the filtered results presented as the discharge hydrograph are one calculation time step ahead prediction results. The updated storage amounts are used to update each cells water depth, and then the next step's discharges are calculated using the updated water depths of the previous step. Even thought it is small, it results in some discrepancies in the hydrograph.

Two different updated results are presented in Figure 8; the hydrograph labeled "using equation 7" is the result when Equation 7 is used to calculate the residuals in the measurement update algorithm, and the other one labeled "using equation 8" is using Equation 8 in the algorithm. Those two kinds of results are obtained under the setting that the observation error variance is zero and the standard deviation of system noise is 30m^3/s. Direct use of the simulated discharge to obtain residuals gives much better filtered results than using the *H* value from the Q-S curve. When these two residuals at each updating step are compared, around 10 to 30m^3/s of difference is observed. In the case of “using Equation 8”, the discrepancies to the observed data are not recognizable as much as in the “using Equation 7” case. Every result from now on is obtained using Equation 8.

3.2 Comparison the updated results from different error variance

Figure 10 shows the various updated results by setting a different error variance. The label “oe30” means 30m^3/s of standard deviation as an observation noise and “se30” means 30m^3/s of standard deviation as a system noise, “se0” means no system noise, and “oe0” stands for no observation noise which means the observed data are regarded as true values.

When the hydrographs from the case “se0.0:oe30” are examined, the filtered results are exactly the same as the results from offline simulation because the system is regarded as a perfect one to do a simulation. On the other hand, the case “se30:oe0.0” shows that the filtered results trace the observed data. The case “se30:oe30” show that the filtered results are located between the offline simulation results and the observed data. Even though the noises are set to the same value as 30m^3/s, the filtered results are closer to the observed data than offline simulation results. Several reasons are considered to explain this phenomenon such as nonlinear observation equation, different form of error variance and initial error variance. While the observation error variance is in terms of discharge, the system error variance is in terms of storage amount which is transformed from the discharge noise. The initial error variance of the Kalman filter also affects the filtered results; the initial error variance is set as 20m^3/s of discharge noise in this study. Further research is needed to determine how these factors affect the filtered results.

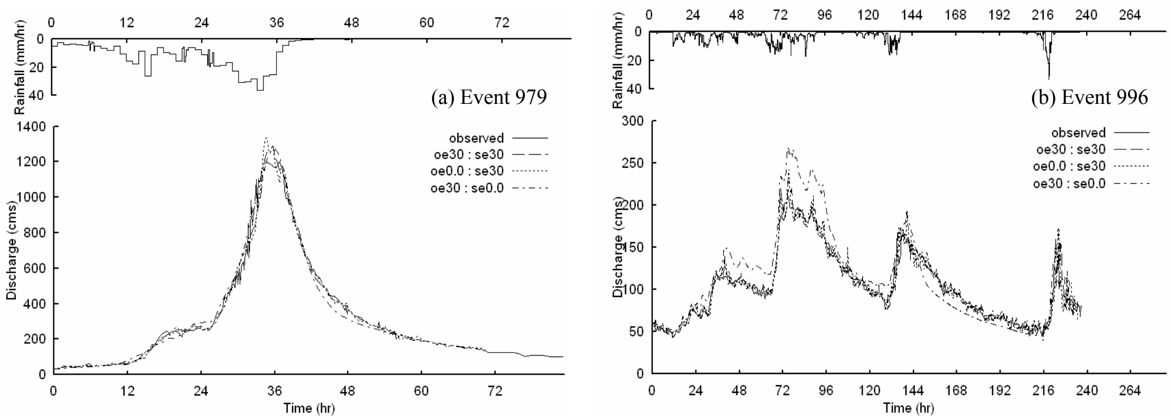


Figure 10. Updated results from variant error variance

3.3 Another noise form: Coefficient of Variation

It is sometimes unreasonable to set the standard deviation of noise as a constant value, no matter how much discharge there is. When discharge is comparably small, for example 50 or 100m^3/s, 30m^3/s of noise is a large value to set as a standard deviation. On the other hand, if discharge is over 1000m^3/s, that amount of the noise could be negligible. For this reason, it is more reasonable to set the noise not as a deterministic value but as a ratio to discharge, such as a form of Coefficient of Variation (C.V.) which is calculated as a ratio of standard deviation to mean value. The 30m^3/s of deterministic system noise can be converted to the C.V. by Equation 14. The converted C.V. value from the four flood events gives a 10% value for noise. Every result here after comes from the C.V. error form. The results shown in Figure 11 are obtained when both observation noise and system noise are set at 10% of observed discharge and simulated discharge. Figure 12 shows the variation of the standard deviation of discharge results using two different noise forms. It can be easily recognized that the variations from C.V. form show reasonable results in accordance with discharge variations.

$$\sqrt{\frac{\sum (y_i \times cv)^2}{n}} = \sqrt{\frac{\sum (x_i - y_i)^2}{n}} \tag{14}$$

where cv : coefficient of variation
y_i : simulated discharge
x_i : observed discharge
n : number of discharge values

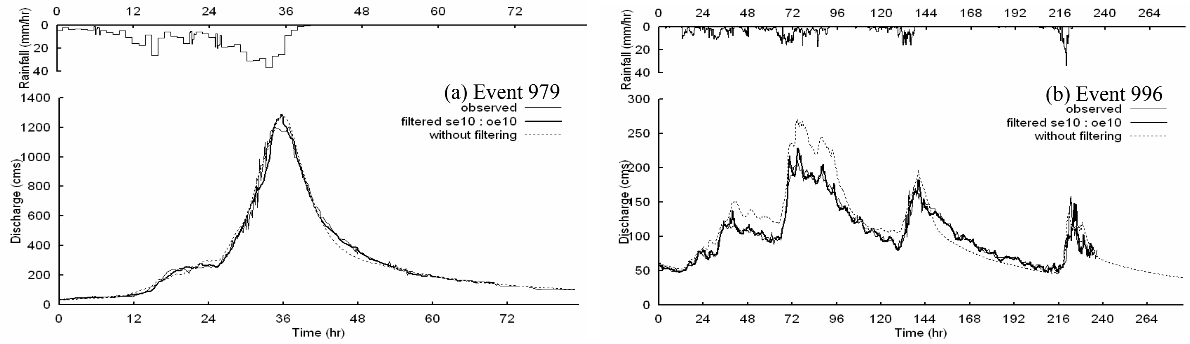


Figure 11 Filtered results when the noises are C.V. form

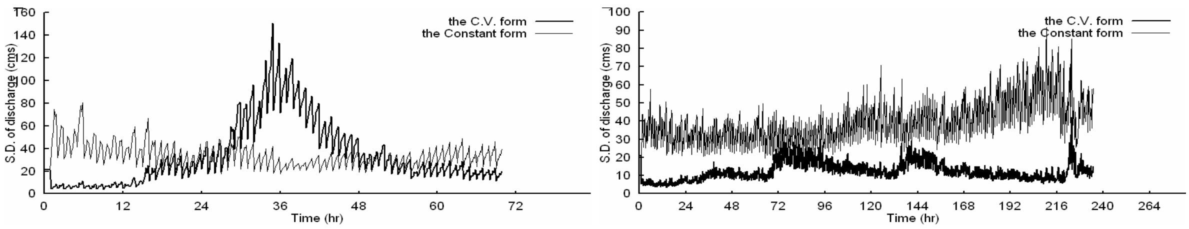


Figure 12. Standard deviation of discharge

3.4 Prediction Efficiency

To check the prediction accuracy after coupling with the Kalman filter, 1hr, 6hr and 12hr prediction results are compared. Figure 13 shows the prediction results when the system noise and observation noise are equally set as 10% of C.V. Table 3 shows the RMSE from the prediction results. The RMSE is calculated by Equation 12. As expected, prediction for short lead times shows higher accuracy. It is interesting that even prediction of 12hr ahead gives quite good accuracy compared to the short lead time forecasting. One main reason for this is the use of recorded rainfall data. Simulation and prediction are carried out under an assumption that we know exactly the expected rainfall. Because this paper focuses on coupling the Kalman filter to a distributed hydrological model, it needs to decrease the factors which impact on the filtered results. If the uncertainty of the rainfall forecast is provided, it should be added in the system error variance Q_k.

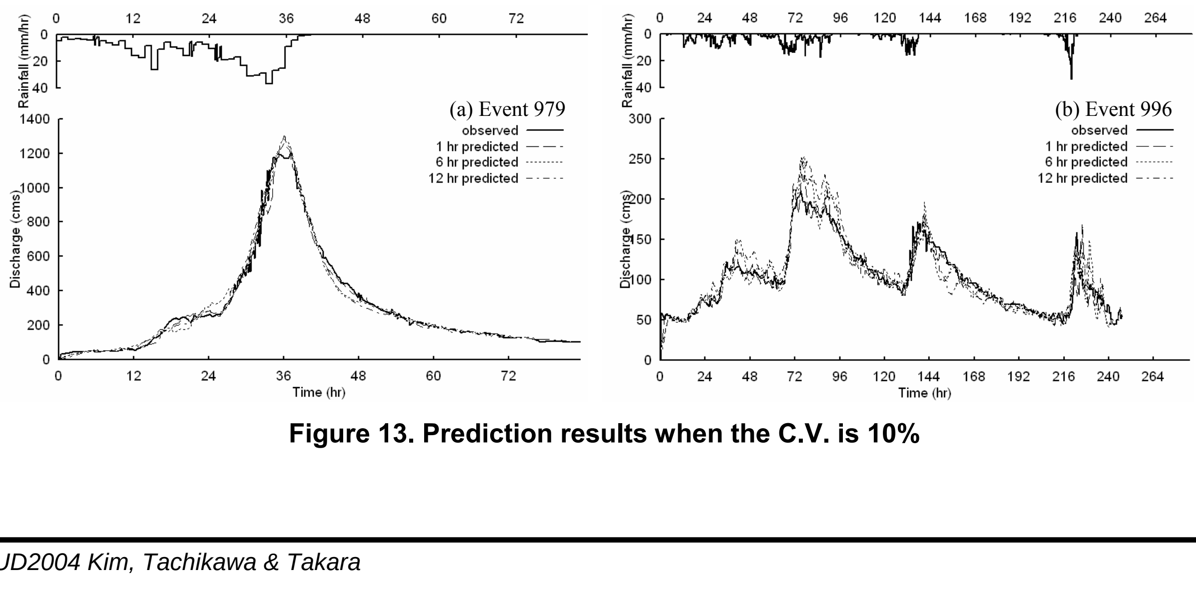


Figure 13. Prediction results when the C.V. is 10%

Table 3. The RMSE of predicted results (unit: m^3/s)

EVENT	1 hr predict	6 hr predict	12 hr predict
Event 979	37.18	39.64	37.28
Event 996	11.42	17.24	20.14
Event 998	16.45	21.93	22.17
Event 999	28.00	34.72	27.85

4. CONCLUSION

The Kalman filter was successfully coupled with the distributed hydrological model, CDRMV3, to update the state variables. Rather than formulate an impractical algorithm to apply the filter, several techniques, such as the Q-S curve, ratio method and Monte Carlo simulation are used. Total storage amount from the Q-S curve is used as an observed storage amount. The ratio method is used for updating each water stage of every cell in the model by multiplying by the ratio calculated from updated total storage amount and the simulated storage amount. For the prediction algorithm, Monte Carlo simulation is adopted to predict state variables and error covariance at the next step. Monte Carlo simulation is an effective technique to analyze the propagation of error covariance.

The CDRMV3 using Kalman filter yields better results than the CDRMV3 without the filter in terms of RMSE and computed hydrographs. Prediction efficiency after coupling with the Kalman filter, for 1hr, 6hr and 12hr prediction results, shows quite good accuracy when compared with the observed data when prediction are carried out under an assumption that the expected rainfall is exactly known.

Research to overcome the steady state assumption on the Q-S curve is needed to improve the filtered results. A methodology to include the uncertainty of rainfall forecasting in the system error covariance is a further research issue.

5. ACKNOWLEDGEMENTS

A part of this study is supported by "System Modelling Approaches for Assessment of Interaction Between Social Changes and Water Cycle (PI: Prof. Kaoru Takara, Disaster Prevention Research Institute, Kyoto University)", which is conducted under the CREST (Core Research for Evolutional Science and Technology) Program by the Japan Science and Technology Agency.

6. REFERENCES

Bras, R. L. and Rodriguex-Itulbe, I. (1985) *Random function and hydrology*. Chapter 8 and 9, Addison-Wesley, Reading, Mass., 425-520.

Hino, M. (1974) *Kalman fiter for the prediction of hydrologic runoff system - application of theory*, Journal of JSCE, Volume 221, 39-47.

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

Kalman, R. E. (1960) *A New Approach to Linear Filtering and Prediction Problems*, Transaction of the ASME – Journal of Basic Engineering, 83, 35-45.

Kim, S., Tachikawa, Y., and Takara, K., (2004) *Real-time Updating of State Variables in a Distributed Hydrological Model*, Annuals of Disaster Prevention Research Institute, Kyoto Univ., No. 47B.

Kojima, T. and Takara, K. (2003) *A grid-cell based distributed flood runoff model and its performance, Weather radar information and distributed hydrological modeling* (Proceedings of symposium HS03 held during IUG2003 at Sapporo, July 2003), IAHS Publ. No. 282, 234-240.

Lee, Y. H. and Singh, V. P. (1999) *Tank model using Kalman filter*, Journal of Hydrologic Engineering, ASCE, Volume 4, No. 4, 344-349.

Maybeck, Peter S. (1979) *Stochastic Models, Estimation, and Control*, Volume 1, Academic Press, New York.

Tachikawa, Y., Nagatani, G., and Takara, K. (2004) *Development of stage-discharge relationship equation incorporating saturated- unsaturated flow mechanism*, Annual journal of hydraulic engineering Vol. 48, 7-12.

Takasao, T., Shiiba, M., and Takara, K., (1989) *Stochastic state- space techniques for flood runoff forecasting*, Pacific international seminar on water resources systems, Tomamu, 117-132.