

REAL-TIME PREDICTION ALGORITHM WITH A DISTRIBUTED HYDROLOGICAL MODEL USING KALMAN FILTER

Sunmin KIM¹, Yasuto TACHIKAWA² and Kaoru TAKARA³

¹Student Member, Graduate student, Dept. of Urban and Environmental Eng., Kyoto University
(Kyoto 606-8501, Japan) E-mail:sunmin@rdp.dpri.kyoto-u.ac.jp

²Member of JSCE, Dr. Eng., Associate Professor, DPRI, Kyoto University (Gokasho, Uji 611-0011, Japan)

³Fellow of JSCE, Dr. Eng., Professor, DPRI, Kyoto University (Gokasho, Uji 611-0011, Japan)

The Kalman filter theory is 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. After updating total water storage with the Kalman filter, the ratio method is introduced to reset the distributed storage amount of a basin, maintaining the spatially distributed pattern. A Monte Carlo simulation is adopted to predict state variables and error variance propagations. A distributed model coupled with the Kalman filter theory gives updated simulation results with improved forecasting accuracy.

Key Words: kalman filter, distributed hydrological model, update state variables, real-time simulation

1. INTRODUCTION

R.E. Kalman¹⁾ 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 (e.g. Maybeck²⁾). A more detailed description about Kalman filter theory can be found at Jazwinski³⁾ and a good discussion of the filter with several application cases to the hydrological system is given by Bras and Rodriguez-Itulbe⁴⁾.

Since Hino⁵⁾ 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.*⁶⁾ described 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 Singh⁷⁾ 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, it is complicated to formulate the Kalman filter algorithm in the system structures of distributed models in most cases. A large number of state variables based on a fine grid cell hydrologic system also make 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 into a distributed hydrological model. Q-S curve which express the relationship between the discharge at the outlet and the total storage amount is determined under steady state assumption on a study basin. The Q-S curve is used for the observation equation of Kalman filter to update the simulated total storage amount of the basin with discharge observation. The spatially distributed storage amount in the model is reset by multiplying by ratio of the updated total storage amount to the simulated storage amount. This technique which is named as the ratio method is adopted to consider the spatial pattern of the storage amount in every grid-cell of a distributed hydrological system. For the prediction algorithm, Monte Carlo simulation is carried out to estimate state variable and error variance propagation of the next updating step. The methodology for deciding the 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 with the Cell based Distributed Runoff Model Version 3⁸⁾ (CDRMV3, <http://fmd.dpri.kyoto-u.ac.jp/~flood/product/cellModel/cellModel.html>). The state variable to be updated is the total storage amount in a basin, and its spatial distribution is calculated using water depth at all computational nodes in the model. The storage amount is easily calculated by multiplying the water depths by 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 variance 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.

(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 University⁸⁾. The model solves the Kinematic wave equation using Lax Wendroff scheme on every node in a cell. 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

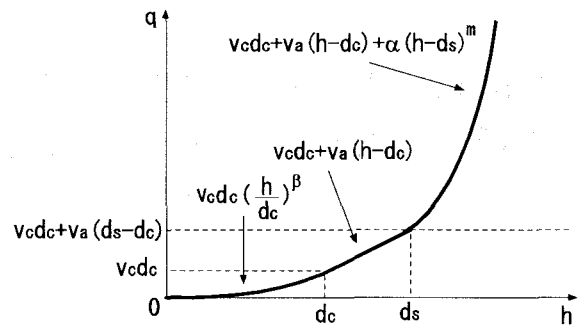


Fig.1 Relationship between unit width discharge and water depth in the CDRMV3.

Table 1 Flood event used in this study.

EVENT	Flood Term	MaxDischarge
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

characteristics of its own cell. Specified stage-discharge relationship, which incorporates saturated and unsaturated flow mechanism, is included in each cell⁹⁾. Because of the variable slope and roughness coefficient, each cell has its own relationship.

The stage-discharge relationship is expressed by three equations corresponding to the water levels divided into three layers (see Figure 1). When the water depth h is lower than the depth of unsaturated layer ($0 \leq h < d_c$), flow is described by Darcy's law with a degree of saturation, $(h/d_c)^\beta$ and velocity v_c . If the h increases ($d_c \leq h < d_s$), flow from the saturated layer is considered with a different velocity v_a of saturated layer d_c . The velocity of subsurface flow v_a and v_c are calculated by multiplying hydraulic conductivity k_a and k_c by slope i . After the water depth is greater than the soil layer ($d_s \leq h$), 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. More detail on the specified stage-discharge relationship and the model structure can be found in Tachikawa *et al.*⁹⁾.

The model is applied to the Kamishiiba basin (211km²) 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.

(2) Measurement Update Algorithm

In the measurement update algorithm of the Kalman filter, an observation equation (Eq. 1) which specifies a relationship 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 . The observations are affected by observation 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 $\hat{x}(k|k-1)$ 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_k . The matrix K_k which is called the Kalman gain is chosen to minimize the updated error covariance $P(k|k)$. In the algorithm, ‘ $\hat{\cdot}$ ’ indicates estimated value and ‘ T ’ indicates the transpose of a matrix.

Observation equation

$$y_k = H_k x_k + w_k; \quad w_k \sim N(0, R_k) \quad (1)$$

Measurement update algorithm

$$\hat{x}(k|k) = \hat{x}(k|k-1) + K_k (y_k - H_k \hat{x}(k|k-1)) \quad (2)$$

$$P(k|k) = P(k|k-1) - K_k H_k P(k|k-1) \quad (3)$$

$$K_k = P(k|k-1) H_k^T (H_k P(k|k-1) H_k^T + R_k)^{-1} \quad (4)$$

The difference, $y_k - H_k \hat{x}(k|k-1)$, which is called the residual or innovation reflects the discrepancy between the estimated observation $H_k \hat{x}(k|k-1)$ 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.

In the CDRMV3, the relationship between the discharge at the outlet and the total amount of storage has a loop shape as shown in Figure 2, whose shapes are different for each flood event. However, it is still possible to get a relationship 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 by water depths of each cell and sum up these entire amounts. The cell size in this study is 250m×250m. Assuming a steady state condition, 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 conversion matrix H in the measurement update algorithm. The matrix relates total storage amount and observed discharge. More specifically, it stands for the gradient of the Q-S curve in accordance with simulation results at

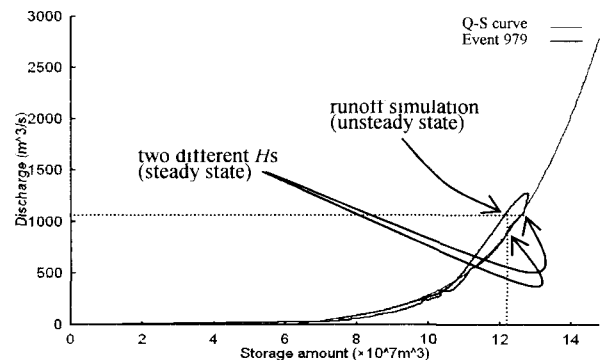


Fig.2 Two different curves between steady and unsteady state.

updating time step. As only one observation is available in the Kamishiiba basin, H is a scalar value in this study.

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; 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 residual in the measurement update algorithm. According to the conventional equation of the Kalman filter, the residual is calculated by the use of the observation equation. On the other hand, the residual also can be calculated directly from the difference between observed discharge and simulated discharge. When the filtered results are examined, the residual from directly using simulated discharge gives much better filtered results. Thus, the directly calculated residual is used rather than conventional form using the H value from the Q-S curve.

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. This method which is named as the ratio method, offers efficient and effective

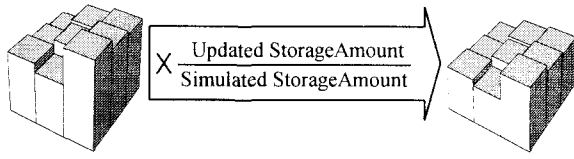


Fig.3 Concept of the ratio method

updating skill of state variables considering its spatial distribution pattern¹⁰.

(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 (Eq. 5) relates the state variables at current time step k to the state at next step $k+1$. The system is continuously affected by system noise, v_k , with covariance matrix Q_k . 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

$$x_{k+1} = F_k x_k + B_k + v_k; \quad v_k \sim N(0, Q_k) \quad (5)$$

Time update algorithm

$$\hat{x}(k+1|k) = F_k \hat{x}(k|k) + B_k \quad (6)$$

$$P(k+1|k) = F_k P(k|k) F_k^T + Q_k \quad (7)$$

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 inter-dependently 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 variance $P(k+1|k)$. 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 applied to solve the problem.

By the concept of the Monte Carlo simulation, lots of random variables are generated at time step k ; one hundred of total storage values were generated in this study. Each random variable is defined by any possible storage amount within a range of probability distribution, $N(\hat{x}(k|k), \sigma_k)$; where, $\sigma_k = P(k|k)^{0.5}$.

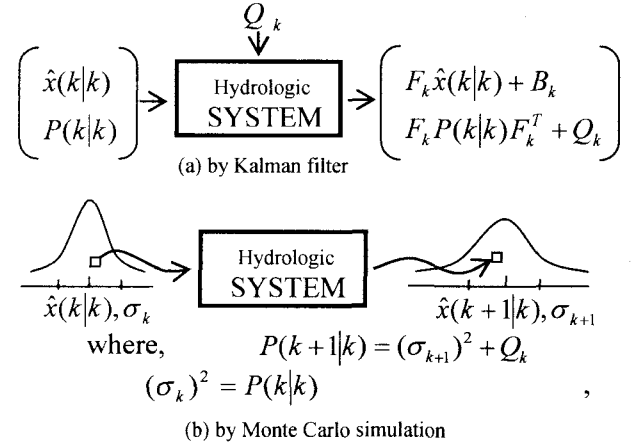


Fig.4 Two estimation concepts of time update algorithm.

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 $\hat{x}(k|k)$. After a simulation repeatedly calculates multiple input sets until the next update time step $k+1$, another probability distribution, $N(\hat{x}(k+1|k), \sigma_{k+1})$, is calculated from the simulated results. Now, the estimated state $\hat{x}(k+1|k)$ is the mean value of the probability distribution and the error variance $(\sigma_{k+1})^2$ is regarded as $F_k P(k|k) F_k^T$. Adding the system error variance Q_k completes the error variance $P(k+1|k)$ at time step $k+1$. Estimated error variance $F_k P(k|k) F_k^T$ means propagation of the error variance $P(k|k)$ through the simulation, and Q_k stands for a new generated or added system variance during simulation from time step k to the next time step $k+1$. The newly added system variance is caused by system structure or new input data such as rainfall. The methodology to determine the system error variance, Q_k , is discussed in the following section.

(4) Setting 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. However, it is impossible to get the true value of the hydrological system because the system is based on the nature which varies in time and space. The only action we can take is to try to get a reasonable error covariance with the least assumptions.

a) 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 sources²⁾. It is reasonable to see the observation noise from this point of view. 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, assumed observation error variance is used in this paper. The assumed error variance is mentioned in each filtered result.

b) System Noise

If observed data are assumed to be true values, biases between simulation results and the observed data could be regarded as system noise. When the first and second momentums of the biases are checked, the mean values are around zero and the standard deviations are around 30m³/s. Also it can be checked that the biases distribute as a normal probability distribution. Because the mean is around zero, the RMSE given by Equation 8 is almost same as the standard deviation of the biases as seen in the Table 2. Following this analysis, the RMSE is regarded as a standard deviation of the system noise w_k in terms of discharge.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{S,i} - Q_{O,i})^2}{n}} \quad (8)$$

where, $RMSE$: Root Mean Square Error
 Q_S : simulated discharge
 Q_O : observed discharge

Table 2 Statistical values of simulation results. (unit: m³/s)

EVENT	MEAN	StDev	RMSE
Event 979	-2.22	36.48	33.64
Event 996	-10.80	22.93	21.15
Event 998	3.32	22.11	20.85
Event 999	-1.02	25.70	23.73

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. Three discharges at a specific time step, Q_k , $Q_k + RMSE$ and $Q_k - RMSE$, will match on the Q-S curve with three different storage amounts, S_k , Sup_k and Sdn_k . Using the 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 9.

$$Q_k = |(Sup_k - S_k) \times (S_k - Sdn_k)| \quad (9)$$

where, Q_k : system error variance at time step k
 Sup_k , S_k , Sdn_k : converted storage amounts

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 has 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. The Q-S curve could be regarded as a linear line in local part.

3. ANALYSIS OF RESULTS

The Kalman filter is successfully coupled to the distributed hydrological model. Figure 5 and Figure 6 show two kinds of filtered results each by setting a different error variance to check the effectiveness of filtering results. The label “oe30” means 30m³/s of standard deviation as the observation noise and “se30” means 30m³/s of standard deviation as the 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:oe30” are examined, the filtered results are exactly the same as the results from the offline simulation. It is a reasonable result because the system is regarded as a perfect one to do a simulation. If there is no system error, which means that the system produces the true value, the filtered results and offline simulation results should match exactly.

On the other hand, the case “se30:oe0” shows that the filtered results trace the observed data. But, there are some discrepancies on the hydrograph rather than exactly match to the observed data. Two reasons are considered to explain these discrepancies. The first one is because of the steady state assumption of the Q-S curve. As shown already in the Figure 2, the difference of Q-S curve to the loop-shaped curve under unsteady condition make different gradient value H , and it affects to the observation update algorithm. The other reason concerns updating frequency. Every observation update is given at every an hour while the calculation time step is ten minutes. Five simulation results between nearest observation update make its own hydrograph. When observation updates are

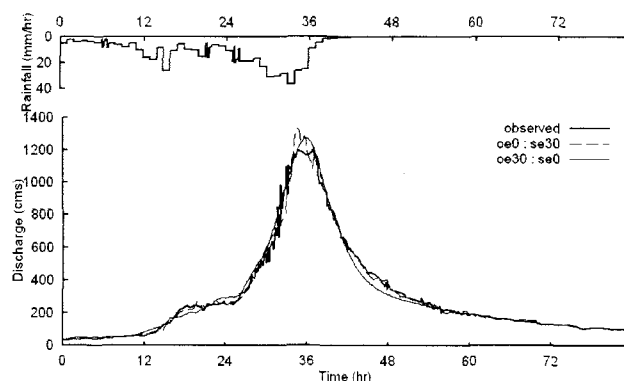


Fig.5 Filtered results of Event 979

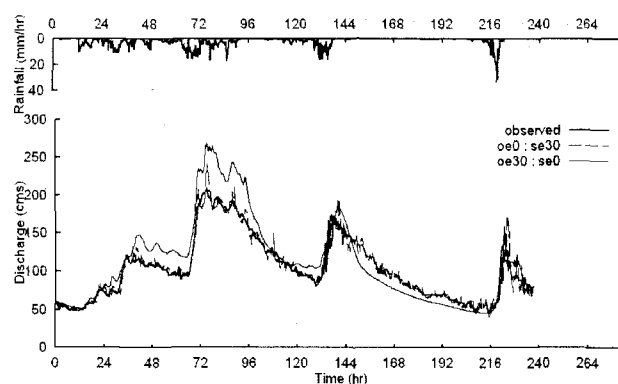


Fig.6 Filtered results of Event 996

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

EVENT	1hr ahead	6hr ahead	12hr ahead
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

carried at every calculation step, the discrepancies are decreased a lot.

To check the prediction accuracy after coupling with the Kalman filter, 1hr, 6hr and 12hr prediction results are compared. Table 3 shows the RMSE from the prediction results when the system noise and observation noise are equally set as $30m^3/s$. 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.

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 adopted. Total storage amount from the Q-S curve is used as the state variable. The ratio method is used for setting each water stage of every cell in the model by specific ratio. For the prediction algorithm, Monte Carlo simulation is adopted to propagate state variable and error variance to the next step. The CDRMV3 using Kalman filter yields better results than the CDRMV3 without the filter in terms of RMSE and computed hydrographs.

Research to overcome the steady state assumption on the Q-S curve is needed to improve the filtered results. Methodologies to include the uncertainty of rainfall forecasting in the system error variance and to use a multi-observation data set is a further research issue.

ACKNOWLEDGMENTS: The authors would like to express special thanks to Kyushu Electronic Co. for providing hydrological data, and CREST program of JST (PI: Prof. K. Takara, Kyoto Univ.).

REFERENCES

- 1) Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, *Transaction of the ASME – Journal of Basic Engineering*, 83, 35-45, 1960.
- 2) Maybeck, Peter S.: *Stochastic Models, Estimation, and Control, Volume I*, Academic Press, New York, 1979.
- 3) Jazwinski, A. H.: *Stochastic Processes and Filtering Theory*, Academic Press, New York, 1970.
- 4) Bras, R. L. and Rodriguez-Iturbide.: *Random function and hydrology. Chapter 8 and 9*, Addison-Wesley, Reading, Mass., 425-520, 1985.
- 5) Hino, M.: Kalman filter for the prediction of hydrologic runoff system - application of theory, *Journal of JSCE*, Volume 221, 39-47, 1974.
- 6) Takasao, T., Shiiba, M., and Takara, K.: Stochastic state-space techniques for flood runoff forecasting, *Pacific international seminar on water resources systems*, Tomamu, 117-132, 1989.
- 7) Lee, Y. H. and Singh, V. P.: Tank model using Kalman filter, *Journal of Hydrologic Engineering*, ASCE, Volume 4, No. 4, 344-349, 1999.
- 8) Kojima, T. and Takara, K.: 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, 2003.
- 9) Tachikawa, Y., Nagatani, G., and Takara, K.: Development of stage-discharge relationship equation incorporating saturated-unsaturated flow mechanism, *Annual journal of hydraulic engineering* Vol. 48, 7-12, 2004.
- 10) Kim, S., Tachikawa, Y., and Takara, K.: Real-time Updating of State Variables in a Distributed Hydrological Model, *Annals of Disaster Prevention Research Institute*, Kyoto Univ., No. 47B, 2004.

(Received September 30, 2004)