Proc. of Monitoring, Prediction and Mitigation of Water-Related Disasters (MPMD2005), pp. 245-250, 2005

New Methods for Applying Kalman Filter Concept into a Distributed Hydrological Model

Sunmin Kim[1], Yasuto Tachikawa[1] and Kaoru Takara[1]
[1]*Disaster Prevention Research Institute, Kyoto University, Uji, Kyoto 611-0011, Japan*

ABSTRACT: Two kinds of Kalman filter applications to a distributed hydrological model are introduced and discussed with their results. Both two applications are fundamentally for updating the spatially distributed storage amount, but the state value in the filter is different from the total storage amount and the discharge. To avoid the computational burden, the distributed storage amounts are updated at once by a specific ratio which is calculated from the state value. Monte Carlo simulation is adopted for the prediction algorithm of the filter.

1 INTRODUCTION

Flood forecasting with high accuracy is one of the challenges in the field of operational hydrology. Even though the computers and technologies become everyday more powerful, still they can not give a sufficient way to describe the complexity of the nature. Especially hydrological system which varies a lot with time and space makes it more difficult to build a model and to predict its phenomena. To strengthen the insufficient hydrological model structure, there is a need to correct the model behavior with observed data. Data assimilation technique stems from this demand to put the model in better compliance with the current observations prior to use its simulation results in rainfall-runoff forecasting.

The Kalman filter (Kalman, 1960) is the most well known sequential data assimilation scheme. The filter can be simply defined as an optimal recursive data processing algorithm to estimate the state variables for minimizing the error statistically. Because the filter needs only the system variables of the previous time step and the forcing terms and observations of the current time step, it can be the best method to handle large amount of data assimilation in the hydrological model. A good discussion of the filter with several application cases to the hydrological system is given by Bras and Rodriguez-Itulbe (1985). While the Kalman filter has been applied to many lumped models for better simulation or more accurate forecasting (Hino, 1974; Takasao *et al*, 1989; Lee and Singh, 1999), 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 make it harder to apply the filter.

In this research, two kinds of Kalman filter applications are introduced to update distributed state variables of a cell-grid based distributed hydrological model. Two applications are basically on the same concept for avoiding the computational burden of the measurement update algorithm; to calculate a ratio from the storage amount change (or a ratio from the outlet discharge change) and to multiply the ratio to the whole state variables. This simple method using a ratio gives efficient and effective update results for the distributed hydrological model with a limited observation in a basin. For the prediction algorithm of the filter, stochastic analysis with Monte Carlo simulation is adopted to predict state variables and error covariance of the next updating step. Monte Carlo simulation let avoid formulating a nonlinear and complex system matrix. The concepts and the results from the two applications are compared and discussed.

2 DISTRIBUTED HYDROLOGICAL MODEL, CDRMV3

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). 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 on each cell reflects the topographic and physical characteristics of its own cell. Specified stage-discharge relationship, which incorporates saturated-

contents page

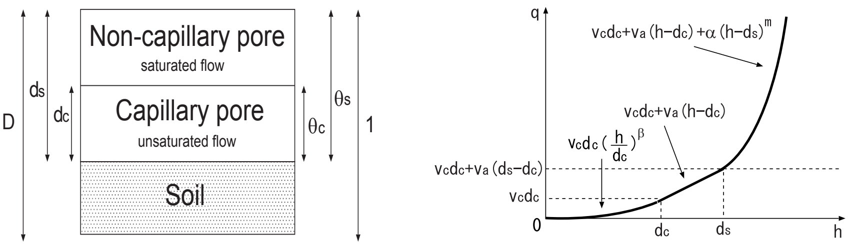


Figure 1. Conceptual soil layers and stage-discharge relationship in each layer

unsaturated flow mechanism, is included in each cell.

The stage-discharge relationship of each cell 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_s. 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 state-discharge relationship and the model structure can be found in Tachikawa *et al*. (2004).

The state variable to be updated in this research is the total storage amount and its spatial distribution; observation is available only at the outlet of a basin. The distributed storage amount is expressed in the form of water depth at all computational nodes in the model. The parameters of the model 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 error variances of the filter, though uncertainty caused by rainfall forecasting is not accounted for in this study.

3 THE RATIO METHOD

With the current technologies, it is impossible to observe storage amount or water depth of every point in a basin. Only observation we can get is discharge or water stage at several gauge stations. To update distributed storage amount using the limited observation, the ratio method is introduced here. First, specific ratio of observed value, which can be total storage amount or discharge, to simulated value is calculated. And then the calculated ratio applies to the simulation result of distributed state variables for maintaining its spatial distribution pattern (see Figure 2). Before getting into the Kalman filter application to the CDRMV3, it is needed to describe the ratio method, which takes a essential roll of the updating spatially distributed state variables.

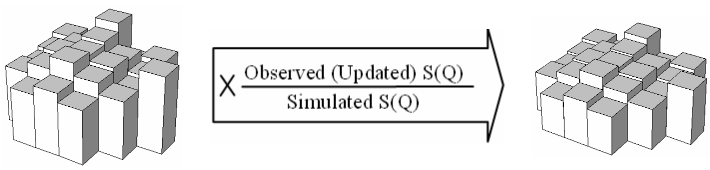


Figure 2. Reset of state variables by specific ratio of storage amount or discharge

3.1 S-ratio method

If there is a difference between observed discharge and simulated discharge at the outlet of the basin, we can assume that it is caused by wrong total storage amount in the model. To fix the simulated total storage amount to the observed one, both storage amounts should be acceptable at every procedure. Simulated total storage amount is easily calculated from water depths when the depth on each grid is multiplied by its area, which is $250 \times 250 \mathrm{m}^2$ in this research. On the other hand, the observed storage amount should be converted from the observed discharge because the total storage amount can not be measured directly. To relate discharge at the basin outlet, Q, and the total storage amount, S, the Q-S curve under steady state assumption is prepared. After reaching the steady state condition with a given constant rainfall intensity on the subject basin, one pair of total storage amount and discharge can be acquired from the CDRMV3. By applying variant rainfall intensity, Q-S curve as

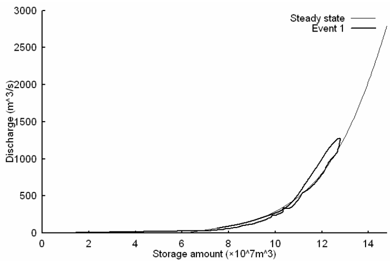


Figure 3. Discharge Storage relationship under steady state and unsteady state

shown in the figure 3 is prepared. Even though runoff simulation under unsteady state makes the loop shaped curve (Event 1 in the Fig. 3), which is different event by event, the difference between the two curves from the steady state and unsteady state seems acceptable.

The S-ratio of the observed storage amount, which is transformed from the observed discharge by the Q-S curve, to the simulated storage amount represent the ratio of basin average water depth. So, the calculated ratio is applied to the simulated water depth on every grid cell to reset the distributed storage amount. After this procedure, the updated water depths are equivalent to the observed storage amount value (from the Q-S curve) and the spatial distribute pattern would contain the pattern of simulation.

3.2 Q-ratio method

The other way to reset the distributed storage amount is applying a ratio of discharge difference at the outlet to the discharge of every point in a basin. This concept is under an assumption that every discharge in a basin would decrease or increase with the same ratio according to the discharge change at the outlet. The assumption is reasonable when we consider a steady state basin.

In the CDRMV3, water depth and discharge on every cell are simulated based on Kinematic wave equation, and those two values are compatible at each calculation time step by the specified stage-discharge relationship on each grid cell. If every discharge is reset by the ratio '$R=Q_s/Q_o$' as shown in the figure 4, every water depth is also reset by its own stage-discharge relationship. The discharge update means the water depth update which finally results in the storage amount update. Figure 4 shows the schematic drawing of discharge and water depth profile reformations at the ideal one dimension basin with the Q-ratio and the stage-discharge relationship respectively. While every updated discharge has the same ratio to the discharge before updating at every point, the ratio of water depth will be different to the point by the topographic and physical characteristic of each cell, such as slope and roughness coefficient.

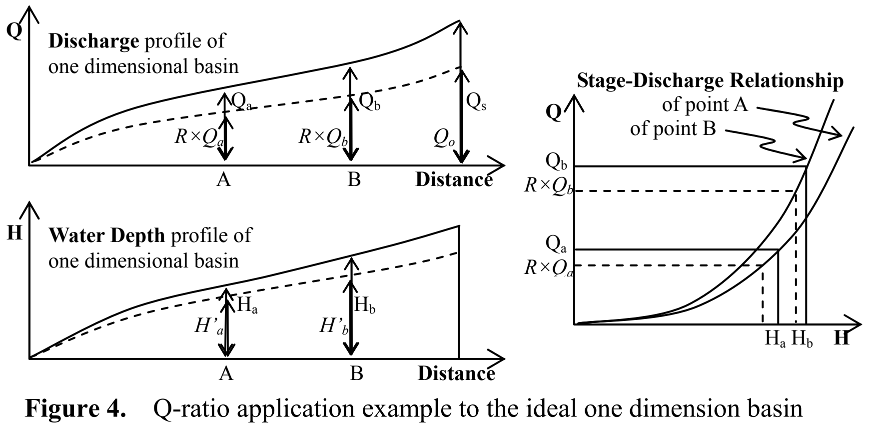


Figure 4. Q-ratio application example to the ideal one dimension basin

4 MEASUREMENT UPDATE ALGORITHM

In the measurement update algorithm of the Kalman filter, an observation equation specifies a relationship between observed data and state values. 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_k 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, '^' indicates estimated value and '^{T}' indicates the transpose of a matrix.

Observation equation

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

Measurement update algorithm

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

$$P(k|k) = P(k|k-1) - K_k H_k P(k|k-1) \tag{3}$$

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

4.1 KFSM (Kalman filter using S-ratio)

In the case of Klaman filter using S-ratio, hereafter KFSM, the observation is the basin outlet discharge and the state value is the total storage amount, and those two values are related by Q-S curve. The conversion matrix H_k is the gradient of the Q-S curve in accordance with simulation results at every updating time step. 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. According to the conventional concept of the Kalman filter, the residual is calculated by the use of the conversion matrix H_k. In this case, the residual is calculated directly from the difference between observed discharge and simulated discharge for simplicity and for avoiding unexpected calculation error.

After getting the updated total storage amount and its error variance from the measurement update algorithm, the ratio of updated storage amount to the simulated storage amount multiplies by every water depth on computation node. The KFSM follows the conventional Kalman filter application scheme for the storage function method except the use of Q-S curve for the observation equation and the use of S-ratio for the updating of distributed water depths. More detail description can be found in Kim *et al*. (2005).

4.2 KFQM (Kalman filter using Q-ratio)

In the Kalman filter using Q-ratio, hereafter KFQM, the observation and the state value of the Kalman filter have the same format as the outlet discharge (or discharges at several gauge station in a basin if they are available). Both KFSM and KFQM are basically for updating the distributed storage amount in the basin. But, each method has a different state value in the Kalman filter algorithm as the total storage amount and the discharge at the outlet, respectively. Because the state value is also the discharge, the measurement update algorithm only estimates the discharge at the basin outlet optimistically for minimizing its error statistically. The results from the update algorithm are simply the optimal discharge with its error variance. The actual update of spatially distributed state variables is carried by Q-ratio method. After getting the updated discharge, the ratio of updated discharge to the simulated discharge multiplies by every discharge on computation node. Then the discharges converted to the water depth through the stage-discharge relationship on each grid cell.

5 TIME UPDATE ALGORITHM

The time update algorithm is for projecting forward the current state and error covariance to obtain the estimation for the next time step. 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 V_k. The matrix B_k relates optional control input to the state x. The estimated error covariance P is a $n \times n$ matrix.

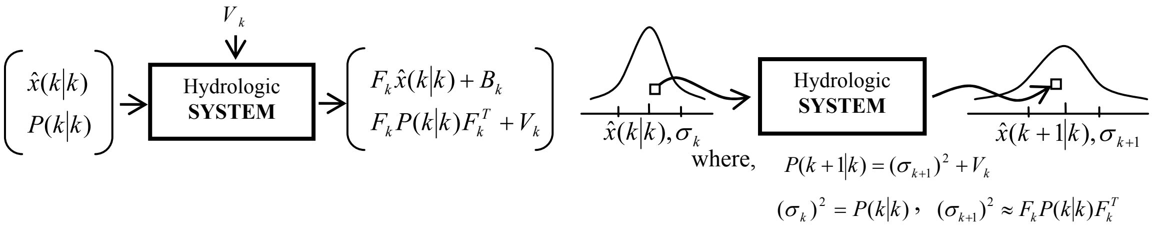
System equation

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

Time update algorithm

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

$$P(k+1|k) = F_k P(k|k) F_k^T + V_k \tag{7}$$



(a) by Kalman filter **(b) by Monte Carlo simulation**

Figure 5. Time update algorithm of the Kalman filter

In the CDRMV3, a complicate nonlinear relation is there between the present state variables and the next state variables which are expressed as the distributed storage amount on more than 3000 grid cells. 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. The complexity of the system makes it impractical to formulate the system matrix F_k, which is essential to update the error variance $P(k+1 \mid k)$. Furthermore computation effort would be doubled by the nonlinearity of the algorithm. For these reasons, rather than use the conventional concept of the Kalman filter as shown in the schematic drawing (a) in Figure 5, the Monte Carlo simulation technique (drawing (b) in Figure 5) is applied to solve the problem.

During the Monte Carlo simulation, lots of random values are generated at any time step k; one hundred total storage amounts for KFSM and one hundred discharges for KFQM. Each random variable is defined by any possible value within a range of probability distribution, $N(\hat{x}(k|k), \sigma_k)$; where, $\sigma_k = P(k \mid k)^{0.5}$. The ratio method is used to reset the distributed state variables by the ratio of each generated value to the updated value $\hat{x}(k|k)$ of the measurement update algorithm. 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 one hundred simulation 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 $F_k P(k \mid k) F_k^T$. Adding the system error variance V_k completes the error variance $P(k+1 \mid k)$ at time step $k+1$. Estimated error variance $F_k P(k \mid k) F_k^T$ means propagation of the error variance $P(k \mid k)$ through the simulation, and V_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.

Table 1. Comparison of the Monte Carlo Simulation (M.C.S.) Procedure

	KFSM	**KFQM**
1. Measurement Update Results	$S_{o,k}$, $\sigma_{s,k}$	$Q_{o,k}$, $\sigma_{q,k}$
2. Generated Value for M.C.S.	$\{S_{k,1}, \ldots S_{k,i}, \ldots S_{k,100}\} \sim N(S_{o,k}, \sigma_{s,k})$	$\{Q_{k,1}, \ldots Q_{k,i}, \ldots Q_{k,100}\} \sim N(Q_{o,k}, \sigma_{s,k})$
3. Ratio Calculation	$R_{s,i} = S_{k,i}/S_{o,k}$	$R_{q,i} = Q_{k,i}/Q_{o,k}$
4. Ratio Multiplies by	*Updated Water Depths of each cell*	*Updated Discharges of each cell*
5. M.C.S.(Time Update) Results	$N(S_{k+1}, \sigma_{s,k+1})$	$N(Q_{k+1}, \sigma_{s,k+1})$

Table 1 explains the input, output data and its procedure for the Monte Carlo simulation of KFSM and KFQM. In the case of KFSM, $S_{o,k}$ and $\sigma_{s,k}$ represent the optimized storage amount and its standard deviation at time step k. Following the probability distribution, $N(S_{o,k}, \sigma_{s,k})$, one hundred random storage amount are generated for the ratio calculation. When one hundred calculated ratio values are multiplied by updated water depths of each cell, one hundred input sets for the Monte Carlo simulation are ready. From the simulation results, the mean value of storage amount, S_{k+1} and its standard deviation, $\sigma_{s,k+1}$ at the next time step $k+1$ are acquired.

6 RESULTS AND ANALYSIS

The distributed hydrological model, CDRMV3, which is coupled with Kalman filter using S-ratio or Q-ratio method, is applied to the Kamishiiba basin (basin area: 211km^2) of Kyushu, Japan. 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 believed as the true value, the filtered results should match to the observed data. Figure 6 shows

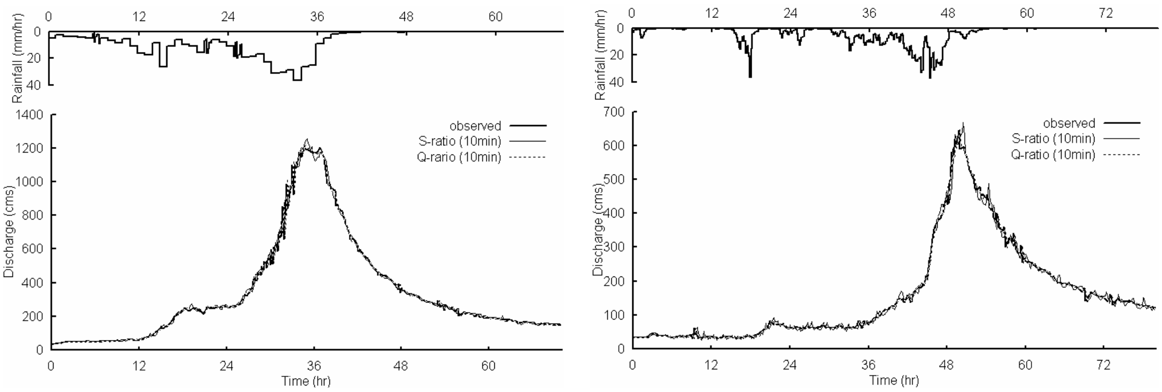


Figure 6. KFSM and KFQM filtered results comparison

the filtered results from the KFSM and KFQM when the system error is set as ±30m^3/s and the observation error is set as zero. The computation time step is ten minutes and also updates are fulfilled at every computation time. In each method, the filtered results are discharge estimations for the one computation time step ahead of the each cells water depth or discharge updating.

In the figure, filtered results from the KFSM shows some discrepancies to the observed data when it is compared to the KFQM results. The reason can be found at the steady state assumption of the Q-S curve. As shown already in the Figure 3, 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. Even though there is also steady state assumption on the Q-ratio method, we can find that the assumption for the discharge ratio is not critical as much as the assumption on the Q-S curve.

7 CONCLUSION

The Kalman filter was successfully coupled with the distributed hydrological model, CDRMV3, to update the state variables. Rather than formulate an impractical algorithm for the filter, several techniques such as the Q-S curve, S-ratio or Q-ratio method and Monte Carlo simulation are used. The CDRMV3 using Kalman filter shows good observed data assimilation results in the sense of reasonable match with the observed data after the updating. The ratio method can be applied to another physical distributed hydrological model such as Topmodel. State variables of the Topmodel, which are expressed as a saturation deficit of each point in a basin, can be reformed by the updated discharges with Q-ratio of the point. Until we can directly measure the distributed state variables, ratio method would be a useful updating method for distributed models with limited observation.

8 ACKNOWLEDGEMENTS

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

REFERENCES:

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

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

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

Kim, S., Tachikawa, Y., and Takara, K. (2005). Real-time prediction algorithm with a distributed hydrological model using Kalman filter, *Annual journal of hydraulic engineering*, JSCE, No. 49.

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, pp 234-240.

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

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

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

contents page