

# RECURSIVE UPDATING OF DISTRIBUTED STATE VARIABLES USING OBSERVED DISCHARGES THROUGH KALMAN FILTER

Sunmin KIM<sup>1</sup>, Yasuto TACHIKAWA<sup>2</sup> and Kaoru TAKARA<sup>3</sup>

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

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

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

By means of Kalman filter, this study examines recursive updating of a distributed model's state variables, which were the water depths on a fine grid cell, using observed discharges at the basin outlet. The measurement update algorithm of the filter decides the optimal discharge, and the ratio of the optimal discharge to the simulated discharges was applied to every discharge from each cell for a reformation of the distributed state variables with a specific stage-discharge relationship. For the time update algorithm of the filter, Monte Carlo simulation was adopted for the prediction algorithm of the filter. The proposed algorithm offers efficient state variables updating in a distributed hydrologic model.

*Key Words: Kalman Filter, Distributed Hydrologic Model, Updating of Distributed State Variables*

## 1. INTRODUCTION

The advances in knowledge and technology up to now have increased the performance of flood forecasting. Hydrologists have understood the detailed physics of water movement with various analytical and empirical equations. Data collecting as well as increased computational technology have harnessed more accurate flood forecasting on a real-time basis.

However, over the last several decades, there has been more progress in computing and developments of distributed models, than in field measurement techniques<sup>1)</sup>. Current hydrology suffers from a measurement problem especially with the distributed properties of catchment: such as water levels and soil moisture at every point in a basin. It is unrealistic to define the detailed spatial distribution of catchment properties through extensive field measurement, as such a process is too costly and time-consuming. Except for few specially designed catchments studied for a specific purpose, outlet discharge is the only commonly available observation in a given catchment. To reflect a low-resolution observation into a detailed spatial resolution for a distributed hydrologic model, an effective method is needed to estimate and update the distributed catchment properties.

To minimize the discrepancy between simulation and observation, correcting the model internal state variables is the most commonly used updating scheme in a real-time simulation. However, in updating the measurement for a distributed model, not only the magnitude of the state variable but also its spatial distribution pattern should be considered. During a rainfall-runoff simulation, inappropriate rearrangement of the spatial distribution of state variables produces obvious effects on the simulation results<sup>2),3)</sup>.

Kim *et al.*<sup>3)</sup> showed an efficient way of updating distributed state variables and successfully applied the Kalman filter into a distributed model. The study introduced an external relationship of observed data and the internal state variables of the hydrologic model, as an alternative to the linear observation function of the Kalman filter<sup>4)</sup>. In the relationship, which is named Q-S relationship, the observed data is defined as outlet discharge and the state value in the Kalman filter is the total amount in storage in the basin. In the time update algorithm, Monte Carlo simulation method made it possible to project a nonlinear variation of system states and their error covariance without the need for linearized system equations. The developed Kalman filter coupled with the physically based distributed hydrologic model can incorporate the uncertainty of input and

output measurement data as well as the uncertainty in the model's structure.

However, one considering point still remains on the use of the Q-S relationship. The relationship represents the overall link of the storage amount in a catchment to the outlet discharge under a steady state assumption, which would hardly happen in the natural phenomena. By this reason, using the Q-S relationship to calculate total storage amount causes additional uncertainty, which is not simple to estimate, in the measurement update algorithm.

This study introduces another efficient way of Kalman filter application into a distributed model, and its adaptability is examined. Newly introduced method uses a ratio of observed discharge (or updated discharge) to simulated discharge for the measurement update algorithm and updates the distributed state variables, water depths, efficiently. For the time update algorithm, the same as the previous study, Monte Carlo simulation offers the evolution of the state variables and its error variance until the next update time.

The organization of this paper is as follow: Section 2 briefly illustrates the previous update method (S-ratio method) of Kim *et al.*<sup>3)</sup> and introduced another ratio method using an outlet discharge (Q-ratio method). In Section 3, Kalman filter application using the Q-ratio method is given with a reviewing on the previous filter application. Discussions are on the efficiency of the new method through the feedback results comparison. Finally, Section 4 concludes the paper by summarizing this study.

## 2. UPDATING OF DISTRIBUTED STATE VARIABLES USING RATIO METHODS

### (1) Reviewing the Distributed Model, CDRMV3

The distributed hydrologic model used here is known as the Cell-based Distributed Runoff Model Version 3 (CDRMV3). CDRMV3 is a physically based model developed at the Disaster Prevention Research Institute of Kyoto University, Japan<sup>5)</sup>. The model solves the one-dimensional kinematic wave equations for both subsurface flow and surface flow using the Lax–Wendroff scheme on every computational node in a cell. Discharges and water depths diffuse to the steepest downward adjacent cell according to a flow direction map generated from DEM data. The CDRMV3 adopts specific stage–discharge relationship incorporating the saturated and unsaturated subsurface flows as well as the surface flow<sup>6)</sup>. This relationship is expressed by three equations as shown in Figure 1, one applying to each water level.

The Kalman filter coupled CDRMV3 is applied to Kamishiiba catchment in Kyushu, Japan, the same as the previous study<sup>3)</sup>. The identical flood

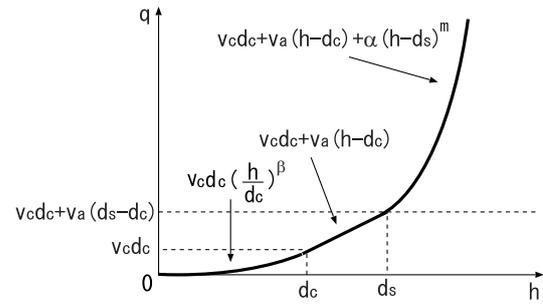


Fig. 1 Relationship between unit width discharge and water depth in the CDRMV3.

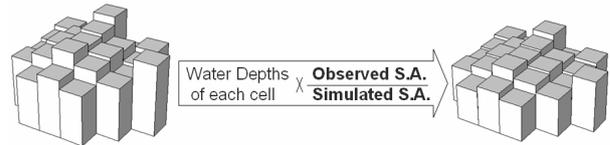


Fig. 2 Resetting of state variables using the ratio of total storage amounts.

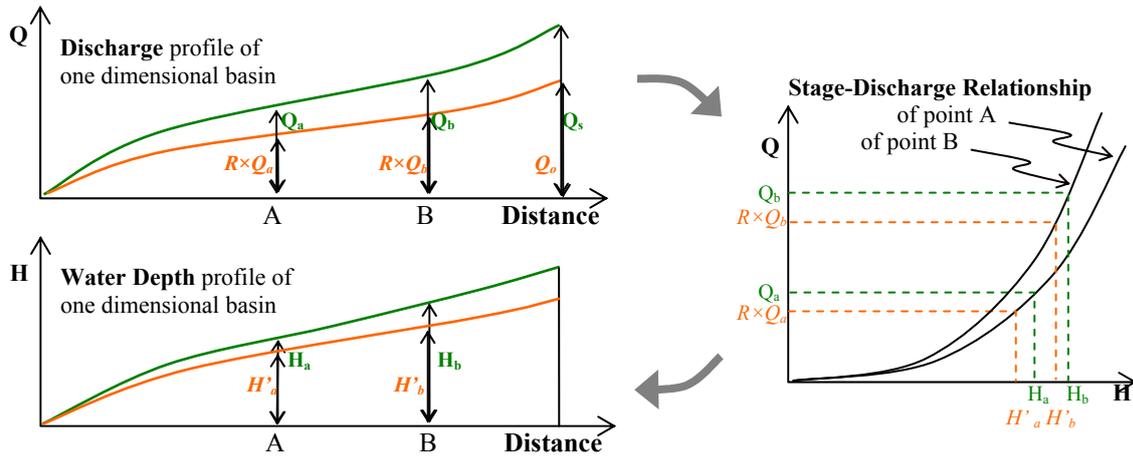
events and the parameter set are selected to compare the feedback results and efficiency with the previous study. More details for the CDRMV3, the subject basin, selected flood events and parameter set, please refer to Kim *et al.*<sup>3)</sup>.

### (2) S-ratio Method

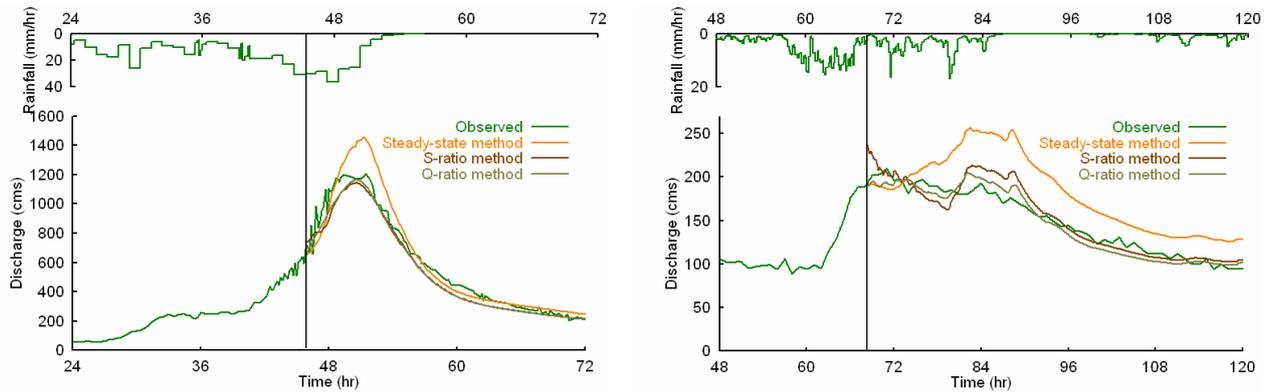
In the study of Kim *et al.*<sup>3)</sup>, to avoid an unpredictable collapse of the internal model state during a simulation, the update method maintains the spatial distribution pattern of the state variables before and after updating. As shown in Figure 2, by multiplying the variables by a specific factor, it was able to update the total storage amount while retaining the spatial distribution pattern of the internal state variables. This factor was calculated from the ratio of the total storage amount, estimated from observed discharge, and the simulated total storage amount.

Even though the S-ratio method offers effective updating results, it is not easy to estimate the observed storage amount. Since the total storage amount in any given basin cannot be measured directly, the corresponding total storage amount must be estimated from the observed discharge, assuming there is a specific discharge–storage relationship. To relate discharge at the basin outlet and the total storage amount, the Q–S relationship under a steady-state assumption was established by applying constant rainfall intensities over the study basin, and variant pairs of total storage amount and discharge was acquired from the CDRMV3.

The critical point on the S-ratio method is that 1) the Q-S relationship is produced under steady state assumption and 2) it should be replaced for every different catchment and even for every different parameter set<sup>3)</sup>.



**Fig. 3** Schematic illustration of the Q-ratio application into the ideal one-dimensional basin.



**Fig. 4** Updated results from the ratio-methods for Event 979 (left) and Event 996 (right), The vertical line stands for the updating time.

### (3) Q-ratio Method

Newly adopted method in this study to reset the distributed storage amount is by applying a ratio of discharge difference at the outlet to the discharge of every point in a basin. This concept is under a steady state assumption such that every discharge in a basin would increase or decrease with the same ratio according to the discharge change in the outlet.

In the CDRMV3, water depth and discharge on every cell are simulated based on the kinematic wave equation, and those two values are convertible at each calculation time step by the specified stage-discharge relationship on each grid cell. If every discharge is reset by the ratio of simulated and observed discharge at the basin outlet ( $R=Q_s/Q_o$ ), each water depth would also be reset by its own stage discharge relationship. The updating of the discharge means the water depth is also updated, which finally results in the storage amount reformation.

Figure 3 shows the schematic drawing of discharge and water depth profile at the ideal one dimensional basin, and an updating procedure with the Q-ratio and the stage discharge relationship. While every updated discharge has the same ratio of discharge before and after the updating, the ratio of water depth will be different by each point based on

the variant topographic and physical characteristics on that cell, such as slope and roughness coefficient.

The Q-ratio method has several merits compared to S-ratio method: 1) a characteristic stage discharge relationship of each cell can reflect topographic and physical characteristics of each individual cell, 2) steady state assumptions of Q-S relationship are not needed to acquire an observed storage amount and 3) can efficiently get the ratio and update.

Figure 4 shows the updated results from the two ratio methods and the steady-state method. The steady-state method is resetting the state variables uniformly as the initial condition based on the outlet discharge at the updating time. The reset of the spatial variables causes the poorest simulation result as discussed in the previous study<sup>2),3)</sup>. The Q-ratio method gives positive updating results as good as the results from the S-ratio method. To check the quantitative updating efficiency, root mean square error (RMSE) is calculated and presented in Table 1.

**Table 1.** RMSE after the updating point (unit: m<sup>3</sup>/sec).

	Steady-state	S-Ratio	Q-Ratio
<b>Event 979</b>	63.046	41.740	39.036
<b>Event 996</b>	36.142	29.152	28.677

### 3. KALMAN FILTER APPLICATION INTO THE DISTRIBUTED MODEL

#### (1) Measurement Update Algorithm

##### a) Using the S-ratio Method

In the measurement update algorithm of the Kalman filter, an observation vector  $y_k$  at time step  $k$  is described as a linear vector function of a state vector  $x_k$ , and observation noise vector  $w_k$  assuming white noise is included. In the previous study of Kim *et al.*<sup>3)</sup>, the observation equation was the Q-S relationship; thus, the gradient of the relationship represents the conversion matrix  $H_k$  at the updating time step. In the Kalman filter using S-ratio method (hereafter KFSM), the observation is the basin outlet discharge and the state value is total storage amount, and those two values are related by the Q-S relationship.

As illustrated in Figure 5 (a), the measurement update algorithm was used to update the total storage amount and its error variance of the subject basin. With the updated storage amount, the S-ratio method is applied to update the spatially distributed water depth in the distributed hydrologic model. Then the updated state variables are given to the model simulation for the time update algorithm. The procedure of the measurement update algorithm using the Q-S curve and S-ratio method is presented in Figure 5 (a).

##### b) Using the Q-ratio Method

In the Kalman filter using Q-ratio method (hereafter KFQM), the observation and the state value of the filter have the same form as the outlet discharge (or discharges at several points in a basin if the basin has multiple gauge station). Both KFSM and KFQM are basically for updating the distributed storage amount in the basin. However, each method has a different state value in the Kalman filter algorithm: total storage amount for the KFSM, and discharge at the outlet for the KSQM.

Since the state value in the KFQM is the same as the observation (discharges at the basin outlet), the measurement update algorithm directly estimates the best discharges minimizing its error statistically. The results from the update algorithm are the updated discharge with its error variance as shown in Figure 5 (b).

The updating of spatially distributed state variables is carried by Q-ratio method; the ratio of updated discharge to the simulated discharge is applied to every discharge on computation node, then the discharges converted to the water depth through the specific stage-discharge relationship on each grid cell. The procedure of the measurement update algorithm using the Q-ratio method is presented in Figure 5 (b).

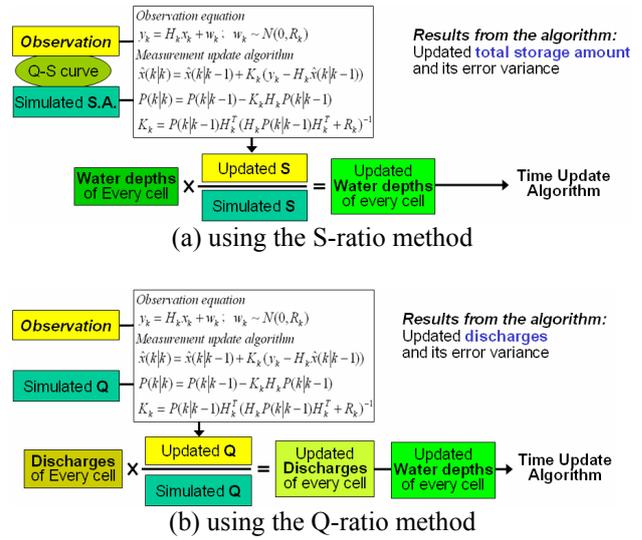


Fig. 5 Illustration of the measurement update algorithm.

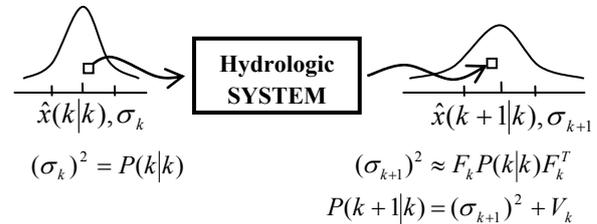


Fig. 6 Schematic drawing of the time update algorithm using Monte Carlo simulation technique.

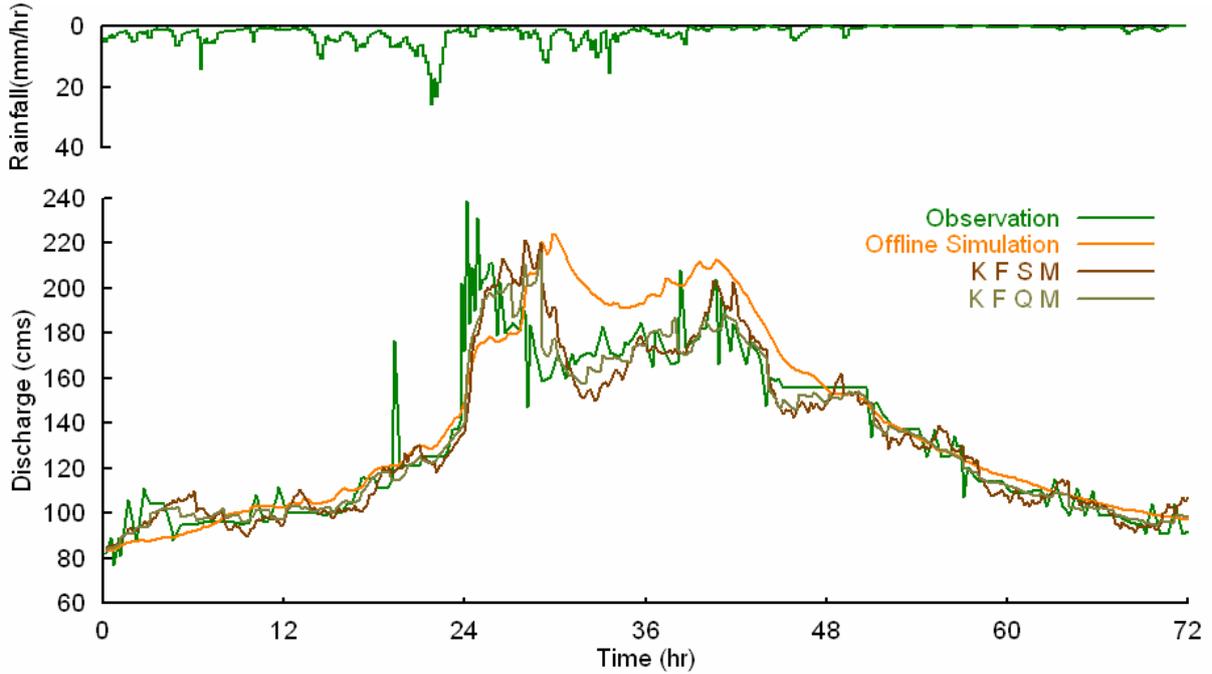
#### (2) Time Update Algorithm

In the CDRMV3, a complicated relationship exists between the present and the next time-step state variable. The current water depth at each cell responds interdependently to the next step's water depth according to the current spatial distribution of water depth and rainfall input. It is impractical to define the system matrix  $F_k$  to express the process formally from the hydrologic system equations; instead, a Monte Carlo simulation (Figure 6) was applied to solve this problem in the previous study<sup>3)</sup>. Many random variables (one hundred total storage amounts in the previous study, KFSM) values were generated within a range of the normal distribution,  $N(\hat{x}(k|k), \sigma_k)$ , where  $\hat{x}(k|k)$  is the updated state variables and  $\sigma_k$  is the square root of the error variance  $P(k|k)$  at time step  $k$ .

For KFQM in this study, one hundred discharge values are generated for the Monte Carlo simulation. Then the Q-ratio method is applied to reset the distributed state variables using the ratio of each generated value to the updated value  $\hat{x}(k|k)$ . Then the multiple input sets are repeatedly given to the runoff simulation to calculate its evolution through the system until the next update time step  $k+1$ . Finally, the propagated probability distribution,  $N(\hat{x}(k+1|k), \sigma_{k+1})$ , is acquired from the one hundred simulation results assuming that the results follow normal distribution.

**Table 2.** Comparison of the Monte Carlo Simulation (M.C.S.) Procedure for KFSM and KFQM

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

**Fig. 7** Observed and feedback values through the Kalman filter using the S-ratio and Q-ratio method (Event 998).

Now, the estimated state  $\hat{x}(k+1|k)$  is mean value of the one hundred state variables evolved through the model, and the variance  $\sigma_{k+1}^2$  is regarded as  $F_k P(k|k) F_k^T$ . Adding the system error variance  $V_k$  completes the error variance  $P(k+1|k)$  at time step  $k+1$  (see Figure 6). The estimated error variance  $F_k P(k|k) F_k^T$  stands for an evolution of the error variance  $P(k|k)$  until time step  $k+1$  through the simulation, and  $V_k$  stands for new generated or added system variance during that simulation. The newly added system variance is caused by system structure or input data such as rainfall.

Table 2 explains the input, output data and its procedure for the Monte Carlo simulation for both KFSM and KFQM. In the procedure of the KFQM, firstly,  $Q_k$  and  $\sigma_{q,k}$  represent the updated discharge and its standard deviation at time step  $k$ , which are the output from the measurement update algorithm of the filter. Following the probability distribution,  $N(Q_k, \sigma_{q,k})$ , one hundred possible discharges ( $Q_{k,1}, \dots, Q_{k,i}, \dots, Q_{k,100}$ ) are generated and each Q-ratio is calculated ( $R_{q,1}, \dots, R_{q,i}, \dots, R_{q,100}$ ) with an updated discharge  $Q_{k,o}$ . After one hundred ratio values are

applied to the updated discharges on each cell, one hundred input sets for the Monte Carlo simulation are ready. From the Monte Carlo simulation results, which are assumed following a normal distribution, the mean value of discharge,  $Q_{k+1}$  and its standard deviation,  $\sigma_{q,k+1}$  are acquired for the time step  $k+1$ . Adding the system error variance  $V_k$  completes the time update algorithm.

In the case of the KFSM, the difference is in the generation of storage amount values and applying the storage amount ratio to the updated water depths on each cell. For the details of KFSM, please refer to Kim *et al.*<sup>3)</sup>

### (3) Application Results and Discussions

The Kalman filter coupled distributed hydrologic model CDRMV3 was tested on the Kamishiiba basin and the feedback results from the both application are compared. Both system noise and observation noise in the filter are assumed time-invariant and having 30m<sup>3</sup>/sec amount of standard deviation following the previous study<sup>3)</sup>. For the KFSM, the system noise in discharge form is

converted to the form of storage amount using the Q-S relationship before it is applied in the filter (refer to Kim *et al.*<sup>3)</sup> for more details). Utilizing the Q-S relationship to convert the noise term, even though it is not significant, offers another additional error source. However, the system noise of discharge form is directly utilized in the KFQM, since the state value in the filter is the outlet discharge itself, which is another benefit of this newly introduced method.

Figure 7 shows the comparison of the feedback results from the both Kalman filter applications in the form of discharge hydrograph. The feedback from the KFSM and KFQM both show the improved simulation results compared to the offline simulation result. When RMSE to the observation was examined, it was found that the Kalman filter using Q-ratio method gives more improved results as shown in Table 3.

**Table 3.** RMSE of each simulation (unit: m<sup>3</sup>/sec).

EVENT	Offline	KFSM	KFQM
Event 979	36.59	31.04	30.32
Event 996	23.84	8.43	6.55
Event 998	21.73	12.97	11.55
Event 999	25.21	22.11	16.52

The computation time for KFQM requires about the same for the KFSM calculation. The most time consumption of the both method is on the Monte Carlo simulation. In this study, one hundred simulations were carried out to assess the evolution of the state variable and its error variance at every time increment: every one hour. Time elapsing will be variable under different model running conditions, such as catchment size, updating frequency and simulation number for the Monte Carlo simulation. Reasonable model condition should be carefully designed based on operating environments.

#### 4. CONCLUSION

As shown already in the previous research, the S-Ratio method showed improved simulation results upon updating. However, the Q-S relationship under a steady state assumption was found necessary in order to produce a storage amount corresponding to an observed discharge, which may also lead to unexpected error in the S-Ratio method. On the other hand, with the Q-ratio method, a successful updating of the state variables becomes simple and effective, as well as the stage-discharge relationship of each cell reflects topographic and physical

characteristics of each cell during the update.

The Kalman filter was successfully coupled to the distributed hydrological model, CDRMV3, to update internal distributed state variables using the Q-ratio method. In the Kalman filter application, both the observation and state variable to be updated are outlet discharge, therefore there is no necessary to formulate the observation equation. The measurement update algorithm decide the optimal outlet discharge based on the given noise terms, and the ratio of updated to simulated discharge is applied for setting the discharges for each cell in the distributed hydrologic model. In the CDRMV3, the reset of discharges is directly connected to the reset of water depths on each grid cell by the stage-discharge relationship.

For the prediction algorithm, a Monte Carlo simulation was adopted to diffuse the state variable and its error covariance. The CDRMV3 using the Kalman filter yielded more effective results than off-line simulations and can thus be used as a probabilistic forecast algorithm. The developed algorithm can incorporate the uncertainty of input and output measurement data as well as the uncertainty of the model itself.

Further work will include developing a method to apply several observations for the updating method to fully utilize the properties of a distributed hydrologic model. One way would be to define sub basins according to the location of observation stations, and to apply different factors to each sub basin for updating the spatial pattern of water depth. For nested river basins with different observation stations, sophisticated methods to define the ratio would be required.

#### REFERENCES

- 1) Beven, K.: Towards an alternative blueprint for a physically based digitally simulated hydrologic response modeling system, *Hydrological Processes*, Vol. 16, pp. 189-206, 2002.
- 2) 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, pp. 273-282, 2004.
- 3) Kim, S., Tachikawa, Y. and Takara, K.: Real-time prediction algorithm with a distributed hydrologic model using Kalman filter, *Annual of Hydraulic Engineering, JSCE*, Vol. 49, pp. 163-168, 2005.
- 4) Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems, *Transaction of the ASME – Journal of Basic Engineering*, 83, pp. 35-45, 1960.
- 5) 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, pp. 234-240, 2003.
- 6) 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, pp. 7-12, 2004.

(Received September 30, 2006)