Annual Journal of Hydraulic Engineering, JSCE, Vol.51, 2007, February

EVALUATION OF THE INFLUENCE OF UNCERTAINTY IN RAINFALL AND DISCHARGE DATA ON HYDROLOGICAL MODELING

Khada Nanda DULAL[1], Kuniyoshi TAKEUCHI[2] and Hiroshi ISHIDAIRA[3]

[1]Student Member of JSCE, PhD student, Dept. of Civil and Environmental Engineering, University of Yamanashi (Takeda 4-3-11, Kofu, Yamanashi 4008511, Japan)

[2]Member of JSCE, PhD, Professor, Dept. of Civil and Environmental Engineering, University of Yamanashi (Takeda 4-3-11, Kofu, Yamanashi 4008511, Japan)

[3]Member of JSCE, PhD, Associate Professor, Dept. of Civil and Environmental Engineering, University of Yamanashi (Takeda 4-3-11, Kofu, Yamanashi 4008511, Japan)

The objectives of this study are to analyze the influence of systematic and random error in rainfall data, discharge data, and the spatial representation of rainfall data on the performance of a distributed hydrological model, BTOPMC. A framework for uncertainty analysis was developed using a Monte Carlo approach, which was applied to the Kalu River basin in Sri Lanka. Findings show that a systematic error exceeding +/-10% in rainfall or discharge data is detrimental to model results. A random error with standard deviation is equal to 10% of rainfall or discharge is not substantial. Calibration of parameters can compensate for some error. The impact of systematic error in rainfall in terms of Nash-Sutcliffe Efficiency (NSE) is higher than that in discharge. The impact of random error in discharge in terms of NSE is higher than that in rainfall. The impact of the random error is the lowest for the gauge network of the highest density, but the impact of systematic error is not the lowest for this case.

Key Words: *Uncertainty, Rainfall, Discharge, Kalu River, Monte Carlo*

1. INTRODUCTION

The nature of hydrological model is inherently imperfect, which leads to the uncertainty in the prediction of model results. The sources of uncertainty in hydrological modeling are: data (input data and data for calibration), parameters, and model structure. In uncertainty analysis researches, most of the studies focus on parameter uncertainty. Since any model starts with data and most of the uncertainty analysis researches in hydrology do not give much attention to data uncertainty, the uncertainty in data is the primary focus of this study.

Rainfall and discharge data are very important data in hydrology. In most of the hydrological modeling, observed rainfall is assumed to be exact and the uncertainty in rainfall data is neglected. However, rainfall data is very uncertain due to the error in measurement and due to the high spatial and temporal variability of rainfall. Although discharge data is considered to be accurate, it is subjected to uncertainty due to error in measurement and uncertainty in rating curve.

The impact of systematic error (bias) and random error in rainfall on hydrological model results has been analyzed by few researchers using Monte Carlo approach [1), 2)]. The influence of spatial variability of rainfall on the output of model has been studied by either generating rainfall using stochastic rainfall generator [3)] or by using the observed rain gauge network [4)]. A few studies [5)] have analyzed the effect of discharge uncertainty on the performance of hydrological model. There is a lack of study on how the systematic and random error in precipitation data, discharge data and precipitation with different sets of gauge network affect the model performance. Therefore, the objectives of this study are: (I) to analyze the influence of error in point rainfall on hydrological modeling results, (II) to analyze the influence of error in discharge on the model results, and (III) to analyze the influence of error with different gauge density on the prediction of discharge. Both systematic and random errors are considered.

2. STUDY AREA

The study area for the research is the Kalu River basin (**Fig. 1**). The Kalu River basin is chosen for the research because it has high density of gauge network. The drainage area of the basin is about 603 km^2 and it is located in the southwest and south of the central highlands in Sri Lanka and lies between 80.4^{0}N-80.6^{0}N longitude and 6.53^{0}E-6.8^{0}E latitude. Elevation varies from 100m to 1700m above mean sea level. Due to the geographical location, the Kalu River basin receives rain during both of the monsoons from May to June and from September to October. There seven precipitation gauging stations in and around the basin, and there is one discharge gauging station. The mean annual values of precipitation and discharge (1987-1996) of the basin are 3497mm and 2570mm respectively. Since the basin is entirely situated in the wet zone, it has a high rainfall to runoff response.

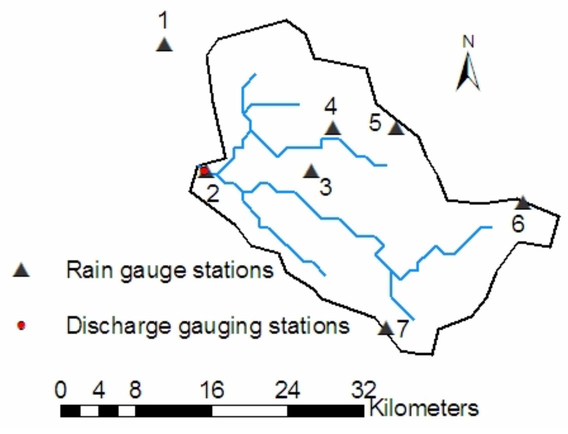

Fig.1 Map of the Kalu River basin

3. HYDROLOGICAL MODEL

The hydrological model used in this study is BTOPMC, "Blockwise use of TOPMODEL with Muskingum-Cunge routing". This is a distributed hydrological model developed at the University of Yamanashi, Japan (refer to Takeuchi et al. [6] for detailed description). It is a grid based model, in which the study basin can be divided into sub-basins to take care of spatial heterogeneity of parameters. The runoff generation of BTOPMC is based on the TOPMODEL concepts and the routing of flow is based on the Muskingum-Cunge routing approach.

4. METHODOLOGY

(1) True set of rainfall, parameters and discharge

A set of observed rainfall and observed discharge together with other inputs is selected and the best set of parameters is identified by calibrating the BTOPMC model. The observed rainfall is assumed as true rainfall (observed rainfall assumed to be free of measurement error); the optimized parameters are considered as true parameters (error free parameters); and the simulated discharge is taken as true discharge (error free discharge).

(2) Monte Carlo approach for uncertainty assessment

a) Assessment of the impact of rainfall uncertainty

True parameter case: Error is applied to the true rainfall and the BTOMPC model is run with perturbed rainfall and true parameters. This procedure is repeated for each set of perturbed rainfall, and the performance of the model with different levels of error is assessed.

Optimized parameter case: Error is applied to true rainfall and for each set of perturbed rainfall, parameters of BTOMPC model are calibrated. This procedure is repeated for each set of perturbed rainfall, and the performance of the model with different levels of error is assessed. The purpose of this experiment is to examine to what extent the parameters can absorb the uncertainty in rainfall.

In this study, the parameters of the model are calibrated manually. If the parameters are physically meaningful, they can be measured in principle. However, the hydrological process is highly variable in space and time, and it is practically impossible to measure the parameters at the spatial and temporal resolutions required by the model, particularly, the distributed model. As a result, they have to be estimated by calibration.

b) Assessment of the impact of discharge uncertainty

Error is applied to the true discharge and the BTOPMC model is calibrated keeping all other inputs at the reference value. This approach is repeated for each set of erroneous discharge, and the impact of discharge uncertainty on the performance of the model is evaluated.

c) Assessment of the impact of error in spatial representation

Different sets of gauge networks are selected, ranging from one gauge to the maximum number of available gauges. Then for each set of gauge network, the impact of error on rainfall is assessed.

(3) Error models

For systematic error (SE), the following error model is used:

$$X_p = (1+K)X_m \qquad (1)$$

where X_p is perturbed variable (rainfall or discharge), X_m is true variable (rainfall or discharge), and K is coefficient.

For random error (RE), the following error model is used:

$$X_p = X_m + \sigma e \quad ; \quad \sigma = rX_m \tag{2}$$

where X_p is perturbed variable (rainfall or discharge), X_m is true variable (rainfall or discharge), σ is standard deviation of random error, e is normally distributed random error with mean equals to zero and standard deviation equals to one, and r is coefficient.

(4) Performance indicators

To analyze the performance of the model due to the systematic and random error in rainfall or discharge (for objectives I and II), three indices are used: Nash-Sutcliffe Efficiency (NSE), Volume Bias (VB), and error in mean annual maximum discharge (MaxE). To check the performance for overall time series, Nash-Sutcliffe Efficiency (NSE) is computed, which is given by the following expression:

$$NSE = 1 - \left(\sum (Q_t - Q_{sim})^2 / \sum (Q_t - Q_m)^2 \right) \tag{3}$$

where Q_t = reference discharge, Q_{sim} = simulated discharge, and Q_m = mean of reference discharge. To check the mass balance, Volume Bias (VB) is computed, which is given by the following expression:

$$VB = \left(\sum Q_{sim} - \sum Q_t \right) / \sum Q_t \tag{4}$$

where Q_t = reference discharge and Q_{sim} = simulated discharge. To check the error in peak discharge estimation, error in mean annual maximum discharge (MaxE) is computed, which is given by the following expression:

$$MaxE = (Q_{sm} - Q_{tm}) / Q_{tm} \tag{5}$$

where Q_{sm} = simulated mean annual maximum discharge and Q_{tm} = reference mean annual maximum discharge.

To compute the indices, the reference discharge in case of rainfall uncertainty is the true discharge, and the reference discharge in case of discharge uncertainty is the perturbed discharge.

The analysis of the impact of error in rainfall with different density (objective III) is the extension of objective I. So, the comparison of performance for objective III is done in terms of the NSE only.

5. RESULTS AND DISCUSSIONS

(1) True/Reference set of rainfall and discharge

To establish true set of rainfall and discharge, time series data of 1987 to 1992 was used for calibrating

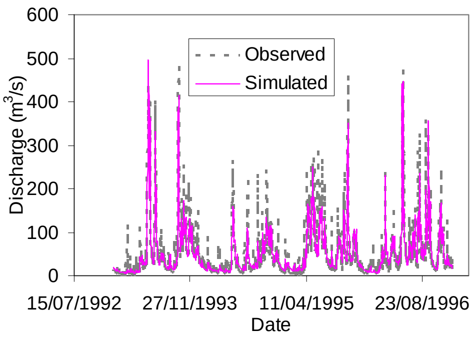

Fig.2 Validation result

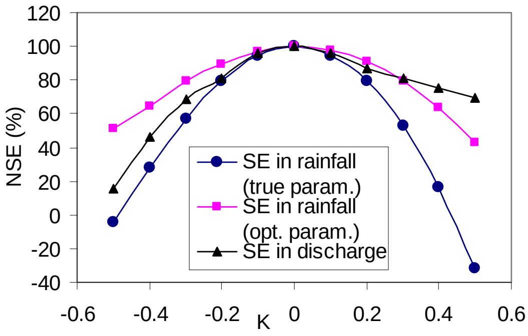

Fig.3 NSE variation due to SE

the BTOPMC model and the data from 1993 to 1996 was used for validating the model. The NSE of the model for calibration is 84.7%, whereas the NSE for validation is 83.5%. VB for calibration is 3.8%, and it is -8.5% for validation. Validation result (**Fig. 2**) shows that although some peaks are underestimated by the model, it has reproduced most parts of the hydrograph well.

For further analysis, observed rainfall data of 1987 to 1992 was considered as true rainfall; simulated discharge of the same period was taken as true discharge; and the optimized parameters were considered as true parameters.

(2) Impact of systematic error in rainfall and discharge

The range of error in rainfall data is [7]: wind: 2%-10%, evaporation: 0%-4%, wetting: 2%-10%, splashing: 1%-2%. As the basin is in wet zone, evaporation error is not significant. Wetting error is also not a major error for automatic gauge. Splashing error is minor as due care is taken for setting the gauge to prevent splash in and out. Wind is the dominant error in rainfall compared to all other errors. So, the usual range of error in rainfall is considered to be within 10%. The error in observed discharge data is around 10% [8, 9]. In the discharge data, measurement error is around 5% [10] and the rating curve error can be taken as 5% [9]. In this study, systematic error of -50% to +50% in step of

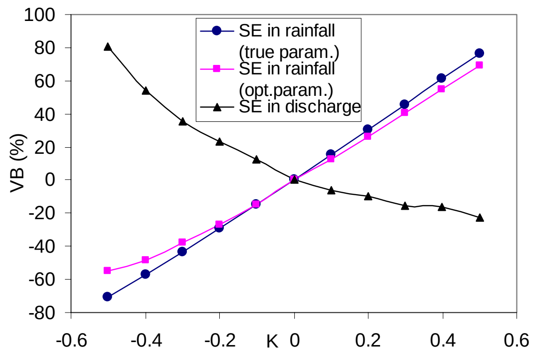

Fig.4 VB variation due to SE

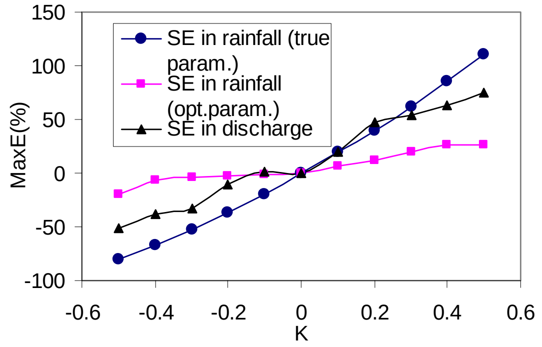

Fig.5 MaxE variation due to SE

of 10% is applied to get a trend of error with different levels of errors. As the systematic error introduces a consistent bias, the higher the error, the worse the model performance. Therefore, this study focuses on the impact of error within the usual range, i.e. in 10% range for both rainfall and discharge.

The variation of the NSE due to the systematic error in rainfall and discharge is shown in **Fig. 3**. For -/+10% systematic error, the decrease in NSE from true model in case of rainfall uncertainty (with true parameters) is 6.1% and 5.4%; the increase in NSE in case of rainfall uncertainty (with optimized parameters) is 3.3% and 2.7%; and the decrease in NSE from true model in case of discharge uncertainty is 4.1% and 4%. The trend of the NSE for all three cases is similar. Although the calibration improves the model performance, the NSE decreases with the increase of error in either direction. The impact of systematic error in rainfall with true parameter is higher than that in discharge. Since the rainfall is the driving variable, the impact is higher. The impact of systematic error in rainfall with optimized parameters is lower than that for discharge from K = -0.5 to 0.2, and higher than that for discharge from K = 0.2 to 0.5. The reason for this is that although optimization can absorb some error, the impact depends on the method of optimization, types of objective function, and the length and quality of data set.

The variation of the VB due to the systematic error in rainfall and discharge is shown in **Fig. 4**. For -/+10% systematic error, the VB in case of rainfall uncertainty (with true parameters) is -14.6% and 15.3%; the VB in case of rainfall uncertainty (with optimized parameters) is -14.7% and 12.3%; and the VB in case of discharge uncertainty is 12.3% and -6.4%. The trend of VB for two cases of rainfall is similar, but for discharge it is opposite. For systematic error in rainfall, the VB increases in the positive direction with the increase of K in the positive direction and vice-versa because increase in rainfall will increase the volume of runoff and vice versa. In case of discharge, VB increases in positive direction with the increase of K in negative direction and vice versa. As the rainfall is fixed, the water balance can not be maintained even after optimization if discharge has very high error. This is one of the reasons for the trend of VB in case of discharge uncertainty.

The variation of the MaxE due to the systematic error in rainfall and discharge is shown in **Fig. 5**. For -/+10% systematic error, the MaxE in case of rainfall uncertainty (with true parameters) is -19.7% and 19.4%; the MaxE in case of rainfall uncertainty (with optimized parameters) is -1.2% and 6.4%; and the MaxE in case of discharge uncertainty is 1% and 19.7%. The trend of MaxE for all three cases is almost similar, with gradual decrease of the MaxE in the negative direction of K and the gradual increase in the positive direction of K. Error in measurement of heavy rainfall obviously amplifies the peak error. In case of discharge, although the calibration of parameters follows the trend of observed hydrograph, the high discharge is not usually captured. So, with the increase of error in discharge, the MaxE increases.

For -/+10% systematic error for both rainfall and discharge, the reduction of the NSE is not very high, within 10% from true model, but the VB and MaxE are in the range of -/+20%. Considering +/-10% VB and MaxE as good indicators of performance, it is concluded that the systematic error exceeding -/+10% degrades the performance of model substantially. Calibration of parameters can compensate for some systematic error due to curve fitting procedure, but the optimized parameters will be biased in presence of error in either rainfall or discharge.

(3) Impact of random error in rainfall and discharge

The variation of the NSE due to the random error in rainfall and discharge is shown in **Fig. 6**. For r = 0.1, the decrease in the NSE from true

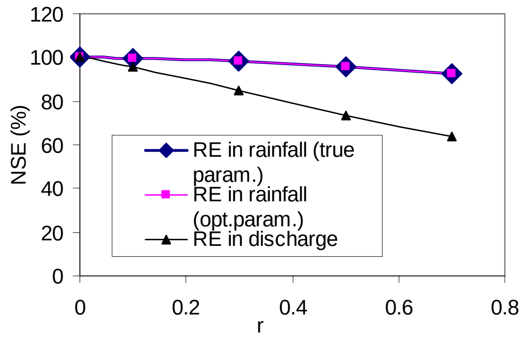

Fig.6 NSE variation due to RE

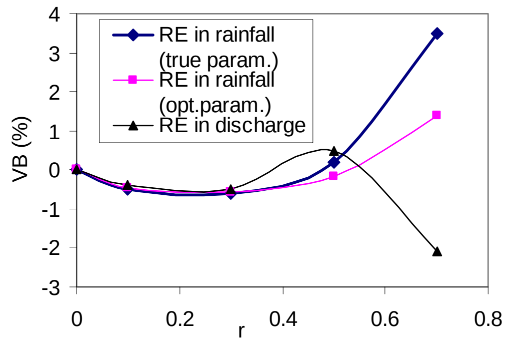

Fig.7 VB variation due to RE

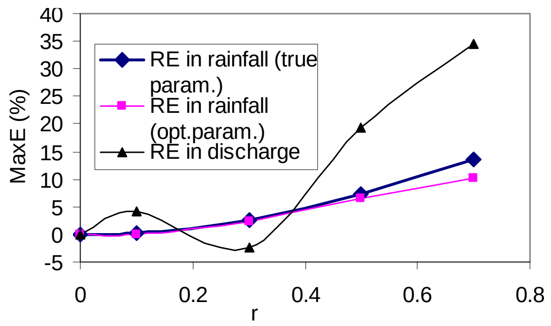

Fig.8 MaxE variation due to RE

model in case of rainfall uncertainty (with true parameters) is 0.2%; the increase in the NSE in case of rainfall uncertainty (with optimized parameters) is 0.01%; and the decrease in the NSE from true model in case of discharge uncertainty is 4%. The trend of the NSE due to the random error is similar for all three cases, with decreasing trend with the increase of r. Calibration of parameters with the random error in rainfall does not improve the performance significantly. The impact of random error in discharge is higher than that in rainfall. One of the possible reasons for this trend is that in case of random error in discharge, there is large error in timing or magnitude of high flow. But in case of random error in rainfall, as the long term expected value of random error is equivalent to zero; and the discharge is fixed at true value to compute the indices, the effect of random error is less than that in discharge.

The variation of the VB due to the random error in rainfall and discharge is shown in **Fig. 7**. For r = 0.1, the VB in case of rainfall uncertainty (with true parameters) is -0.5%; the VB in case of rainfall uncertainty (with optimized parameters) is 0.47%; and the VB in case of discharge uncertainty is-0.4%. The trend of the VB with increase of random error is similar for the cases of rainfall showing gradual increase with the increase of error, while for discharge, it is fluctuating. As rainfall is the driving variable, the increased random error makes the performance deteriorate gradually. Hence, the deteriorating trend of the VB. The random errors fluctuate around the measured value, which is the reason for fluctuating trend of the VB in case of discharge uncertainty.

The variation of the MaxE due to the random error in rainfall and discharge is shown in **Fig. 8**. For r = 0.1, the MaxE in case of rainfall uncertainty (with true parameters) is 0.3%; the MaxE in case of rainfall uncertainty (with optimized parameters) is -0.1%; and the MaxE in case of discharge uncertainty is 4.1%. The trend of the MaxE with increase of random error is similar for the cases of rainfall, but it is fluctuating in case of discharge. The reason for this trend is similar to that of the VB.

Findings show that the impact of random error for r = 0.1 is not significant. The random errors fluctuate around the measured value and therefore the deviations on average show less spread. This is the reason for the insignificant improvement in the performance of model after calibrating with erroneous rainfall.

(4) Impact of error in rainfall with different gauge density

The gauge network with 7 rain gauge stations (see **Fig. 1** for location of rainfall stations with gauge number) is considered as a reference gauge network (Sref). The following 6 sets of gauge networks were considered (based on representative of locations): S1 (gauge no. 3), S2 (gauge no. 3, 7), S3 (gauge no. 1, 5, 7), S4 (gauge no. 1, 3, 6, 7), S5 (gauge no. 1, 2, 4, 6, 7), and S6 (gauge no. 1, 2, 3, 5, 6, 7). The impact of rainfall uncertainty is analyzed using true parameters.

a) Impact of systematic error

The variation of the NSE due to the systematic error in rainfall for different gauge density is shown in **Fig. 9**. The trend of the NSE for all cases shows that the NSE becomes worse with the increase of systematic error in either direction. As higher is the systematic error, the worse the performance of the

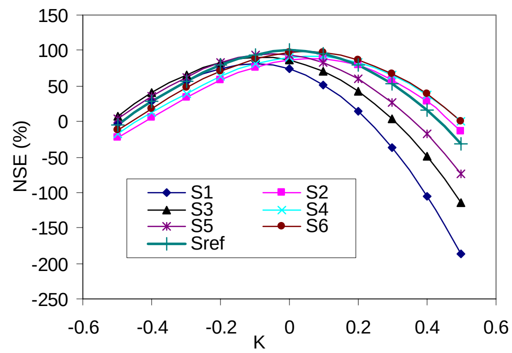

Fig.9 NSE variation due to SE in rainfall with different number of gauges

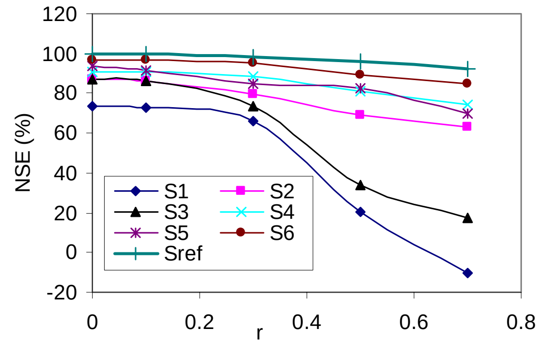

Fig.10 NSE variation due to RE in rainfall with different number of gauges

model for gauge network of any density. Findings show that the impact of the systematic error is not the lowest for the case of the gauge network with the highest density. This reason for this trend is not only related to the errors imposed in the rainfall, but also related mainly to the number of gauges, their location and the representativeness of the location to capture the spatial variability. Even if there is rainfall in one station, there might be no rainfall in other stations. No rainfall means no error. This is one of the reasons for the better performance of the low density gauge compared to the high density gauge in presence of error. The trend further shows that for negative *K*, the NSE is decreasing slowly, but for positive *K*, the NSE is deteriorating rapidly. This asymmetry is due to the non-linear nature of the hydrological model.

b) Impact of random error

Fig. 10 shows the variation in the NSE due to the random error in rainfall with different gauge density. With the increase of the random error, the NSE gradually decreases for all cases. The performance of the model due to the random error in terms of the NSE is the worst for S1 and the best for the gauge network with the highest density. In between S1 and Sref, the NSE is not necessarily better for high density network for some cases, e.g. NSE for S2 is better than S3. As the long term mean of random error is assumed to be zero, the performance of model with high density is high. However, similar to the reasoning of systematic error case, the performance in some cases for low density is better due to the number of gauges, their location and spatial coverage of the gauges.

6. CONCLUSIONS

The main conclusions of the research for the Kalu river basin are as follows: The systematic error exceeding +/-10% of rainfall or discharge is detrimental to model results. The impact of systematic error in rainfall in terms of the NSE with true parameter is higher than that in discharge. Random error with r = 0.1 is not influential to model results. The impact of random error in discharge in terms of the NSE is higher than that in rainfall. Calibration can compensate for some error, but the higher errors will make the parameters biased. The impact of the random error is the lowest for the gauge network of the highest density, but the impact of systematic error is not the lowest.

REFERENCES

1) Oudin, L., Perrin, C., Mathevet, T., Andreassian, V. and Michel, C.: Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *Journal of Hydrology*, Vol. 320, pp.62-83, 2006.
2) Dulal, K.N., Takeuchi, K. and Ishidaira, H.: An assessment of uncertainty in precipitation on runoff imitation, *Annual Journal of Hydraulic Engineering*, JSCE, Vol.50, pp.91-96, 2006.
3) Krajewskei, W.J., Lakshmi, V., Georgakakos, K.P. and Jain, S.C.: A Monte Carlo study of rainfall sampling effect on a distributed catchment model, *Water Resources Research*, Vol.27, pp.119-128, 1991.
4) Dong, X., Dohmen-Janssen, C.M. and Booij, M.J.: Appropriate spatial sampling of rainfall for flow simulation, *Hydrological Sciences Journal*, Vol.50, pp.279-298, 2005.
5) Montanari, A.: An attempt to quantify uncertainty in observed river flows: effect on parameterization and performance evaluation of rainfall-runoff models, 2004 AGU Fall Meeting, San Francisco, USA, H11G-03, 2004.
6) Takeuchi, K., Ao, T. and Ishidaira, H.: Introduction of block-wise use of TOPMODEL and Muskingum-Cunge method for the hydro-environmental simulation of a large ungauged basin, *Hydrological Sciences Journal*, Vol.44, pp.633-646, 1999.
7) Sevruk, B.: Methods of correction for systematic error in point precipitation measurement, WMO Tech. Doc. 589, 91pp, Geneva, Switzerland, 1982.
8) Sorooshian, S. and Gupta, V.: Model Calibration, *Computer Models of Watershed Hydrology*, Singh, V.P. eds., Water Resources publications, pp.23-68, 1995.
9) Butts, M.B., Payne, J.T., Kristensen, M. and Madsen, H.: An evaluation of the impact of model structure on hydrological modeling uncertainty for streamflow simulation, *Journal of Hydrology*, Vol.298, pp.242-266, 2004.
10) WMO: Guide to Hydrological Practices, WMO publication No. 168, Geneva, Switzerland, 1994.

(Received September 30, 2006)