

PARAMETER UNCERTAINTY ANALYSIS OF A STORAGE FUNCTION MODEL USING BOOTSTRAP METHOD FOR AN URBAN WATERSHED

Saritha Gopalan PADIYEDATH¹, Akira KAWAMURA², Hideo AMAGUCHI³
and Gubash AZHIKODAN⁴

¹Member of JSCE, Ph.D. Student, Dept. of Civil and Environmental Eng., Tokyo Metropolitan University
(1-1 Minami Osawa, Hachioji, Tokyo 192-0397, Japan)

E-mail:charu666@gmail.com

²Member of JSCE, Professor, Dept. of Civil and Environmental Eng., Tokyo Metropolitan University
(1-1 Minami Osawa, Hachioji, Tokyo 192-0397, Japan)

E-mail:kawamura@tmu.ac.jp

³Member of JSCE, Asst. Prof., Dept. of Civil and Environmental Eng., Tokyo Metropolitan University
(1-1 Minami Osawa, Hachioji, Tokyo 192-0397, Japan)

E-mail:amaguchi@tmu.ac.jp

⁴Member of JSCE, Asst. Prof., Dept. of Civil and Environmental Eng., Tokyo Metropolitan University
(1-1 Minami Osawa, Hachioji, Tokyo 192-0397, Japan)

E-mail:gubash@tmu.ac.jp

Parameter uncertainty analysis of rainfall-runoff models is very important especially in urban watersheds due to the high flood risk in these areas. Among the different methods available for uncertainty analysis, bootstrap method gained popularity in view of its flexibility. Hence, this study aims to conduct the parameter uncertainty analysis of the urban storage function (USF) model, a storage function model specifically developed for the urban watersheds, using the model-based bootstrap method. We successfully evaluated the uncertainty of USF model parameters and the results exhibited that the 95% confidence interval of all parameters is wide compared with the search range during parameter estimation except for two parameters. Moreover, the parameters with the highest and least uncertainties were identified. Further, model simulation efficiency using the estimated parameters was found to be high with a Nash-Sutcliffe Efficiency value of 97%. Lastly, the effect of parameter uncertainty on model simulation uncertainty was analysed and found that the SCE-UA method along with the model-based bootstrap method can predict, on an average, 68% of observed data within the simulation uncertainty range of USF model.

Key Words: *urban storage function model, model-based bootstrap resampling method, parameter estimation, uncertainty analysis, model simulation uncertainty*

1. INTRODUCTION

Urban areas are characterized by the presence of sewer systems and impervious surfaces which will accelerate the rainfall-runoff (R-R) transformation process, and flood flows are therefore higher and more rapid when compared to that in rural areas. Therefore, it is very important to detect urban floods because of the increased risks and costs associated with them. For this purpose, the R-R models are important tools and they play a central role in urban

watersheds¹). These models contain various components in the form of different equations and parameters, which further describes the hydrological processes and the physical watershed characteristics. The ability of a model to truly reflect the hydrological processes mainly depends upon the precision of its parameter values. Generally, the model parameters are quantified from the watershed properties or theoretically analysed by comparing with similar models²). However, after the introduction of computer-intensive statistics, the parameters are

estimated by calibrating the models against the available observations even though it will increase the efforts and cost taken for the data measurements required for calibration. The so-obtained calibrated parameters are influenced by several factors such as quantity and quality of input data, model error, correlation between the parameters, etc.³⁾ and cannot fully characterize the actual processes, thus leading to a great level of uncertainty of parameters. This parameter uncertainty will further contribute to model simulation uncertainties and hence its quantitative evaluation is critical in reducing the uncertainty of these simulations.

Appropriate uncertainty consideration of the model parameters is necessary although it has been much ignored until recently. Recent researches have paid much more attention to the parameter uncertainty and its associated effects on the model performance. Among the different methods proposed for the assessment of parameter uncertainty, most of the techniques rely on either parametric methods or Bayesian methods⁴⁾. However, in the parametric method, the structure of the model is specified a priori and the number and nature of the parameters are generally fixed in advance, and there is a little flexibility. Also, the Bayesian techniques necessitate the assumption of the prior distribution of model parameters⁴⁾. Hence, the nonparametric method has gained popularity over the above methods because they make no prior assumptions on the model structure and is more flexible.

The bootstrap method, a nonparametric technique, has been developed by Efron⁵⁾ for random resampling of the original data set to develop replicate data sets from which the underlying distribution of the statistics of interest such as mean, variation, correlation, etc. can be estimated. This resampling technique has applications in diverse fields like hydrology, groundwater and air pollution modelling, ecological indices, etc.⁶⁾ in which it has been successfully used in hydrological modelling to estimate the sampling variability of reconstructed runoff⁷⁾, to develop artificial neural network (ANN) models⁸⁾, etc. by utilising the non-time series data. However, the use of bootstrap for time series data was limited because the classical bootstrap assumes that the data set is independent and identically distributed (iid)⁵⁾ which means each data of the data set will be mutually independent and selected from the same population as the others at each index of time and space. Therefore, the direct resampling is not feasible for a time series data which exhibits strong temporal correlation, and the dependence cannot be preserved²⁾.

In order to overcome the problem of dependence of time series, the bootstrap method can be extended

either as a model based bootstrap⁹⁾, which involves the resampling of model residuals or as a block bootstrap¹⁰⁾ in which sequences of observations are resampled rather than the individual data points. Thereafter, the use of bootstrap methods in time series analysis has received considerable attention especially for the parameter uncertainty analysis by the estimation of confidence intervals (CI). For instance, parameter uncertainty of conceptual salt load model, trend analysis of temperature time series, etc.^{4), 11)} was conducted using the block bootstrap. However, very limited studies have been conducted to quantify the parameter uncertainty of R-R models using the model-based bootstrap approach even though it is the topic of current research. Specifically, to the best of our knowledge, no studies have been reported in the urban watersheds for the parameter uncertainty analysis of R-R models. Hence, this study aims to analyse the parameter uncertainty of urban storage function (USF) model¹²⁾, a relatively new storage function (SF) model specially developed for the urban watersheds where combined sewer systems are in use, using the model-based bootstrap method. The authors have already evaluated the performance of different SF models and found that the USF model has higher performance compared with conventional SF models for the Kanda river basin, a typical small-to medium-sized urban watershed in Tokyo¹³⁾ and having a future role in urban hydrology.

2. MATERIALS AND METHODS

(1) Model-based bootstrap method

The classical idea behind the bootstrap method is the resampling and extraction of N samples from the original sample with an unknown distribution having a size of n . The time series data sets are highly dependent in nature and it is quite unreasonable to perform the classical bootstrapping which destroy the original dependency structure of the time series. Therefore, the model-based bootstrap method is utilised for this study and is described as follows⁹⁾.

Consider the original data set $\{X(t), Y(t)\}$; where $t = 1, \dots, n$, $X(t)$: input data set, $Y(t)$: observed discharge data set, n : length of the sample data. The model discharge can be written as $Y(t) = F(X(t), \theta)$; where θ is the parameter vector $\theta_1, \dots, \theta_p$ with p being the number of model parameters²⁾. Initially, we calibrate the model to obtain the estimated parameter vector $\hat{\theta}$ and this was further used along with input data set to compute the simulated discharge data set, $\hat{Y}(t)$ and it can be demonstrated as $\hat{Y}(t) = F(X(t), \hat{\theta})$. Then the model residuals were estimated using the following Eq. (1).

$$\varepsilon(t) = Y(t) - \hat{Y}(t) = Y(t) - F(X(t), \hat{\theta}) \quad (1)$$

The model residuals, $\varepsilon(t)$ were assumed to be iid for $t = 1, \dots, n$ which is the only assumption made for the bootstrapping. The assumption was checked by constructing the distribution plot of the residuals and it demonstrated a very small bias with a mean value close to zero. The detailed model-based bootstrapping procedure is outlined as follows:

- (1) Bootstrap resampling of the residual time series $\varepsilon(t)$ with replacement to form the new bootstrap residual sample series, $\varepsilon^*(t)$.
- (2) Add the new series $\varepsilon^*(t)$ to the simulated discharge data set $\hat{Y}(t)$ to form the bootstrapped discharge data set as $Y^*(t) = \hat{Y}(t) + \varepsilon^*(t)$.
- (3) Calibrate the bootstrapped discharge data set with input data set as $\{X(t), Y^*(t)\}$ to obtain the bootstrapped parameter vector $\hat{\theta}^*$ and associated computed discharge data $\hat{Y}^*(t) = F(X(t), \hat{\theta}^*)$.
- (4) Repeat the steps 1-3 N times to obtain bootstrap samples. In the present study, N is set as 1000.
- (5) Derive the ordered bootstrap estimates $\{\theta_{i1}^*, \dots, \theta_{iN}^*\}$; ($i = 1, \dots, p$), for each θ_i^* and obtain the 95% CI for θ_i^* from the ordered bootstrap samples.

(2) USF model

The USF model is characterized by the relationship between the storage and discharge and is given as¹²⁾:

$$s = k_1(Q + q_R)^{p_1} + k_2 \frac{d}{dt}(Q + q_R)^{p_2} \quad (2)$$

where s : storage (mm), Q : observed river discharge (mm/min), q_R : storm drainage from the basin through the combined sewer system (mm/min), t : time (min), k_1, k_2, p_1, p_2 : model parameters. Combining the above Eq. (2) with the following continuity equation yields the nonlinear expression of USF model.

$$\frac{ds}{dt} = R + I - E - O - Q - q_R - q_l \quad (3)$$

where R : observed rainfall (mm/min), I : inflow from other basins (mm/min), E : evapotranspiration (mm/min), O : water intake from the basin (mm/min), q_l : loss to the groundwater (mm/min). Further, the loss to groundwater (q_l) was defined by considering the infiltration hole height (z) and is given by¹²⁾:

$$q_l = \begin{cases} k_3(s - z) & (s \geq z) \\ 0 & (s < z) \end{cases} \quad (4)$$

where k_3 and z are the parameters. The storm drainage q_R is defined as¹²⁾,

$$q_R = \begin{cases} \alpha(Q + q_R - Q_0) & \alpha(Q + q_R - Q_0) < q_{R \max} \\ q_{R \max} & \alpha(Q + q_R - Q_0) \geq q_{R \max} \end{cases} \quad (5)$$

where α is the slope of the linear relationship between total discharge ($Q + q_R$) and the storm drainage (q_R), and Q_0 is the initial river discharge just before the

rain starts¹²⁾. Substituting Eqs. (2) and (4) into (3) will lead to a second-order Ordinary Differential Equation (ODE) and can be numerically solved after transforming into a first-order ODE. The river discharge Q will obtain from the solution.

The USF is a seven-parameter model with parameters $k_1, k_2, k_3, p_1, p_2, z, \alpha$ used in the R-R modelling. The model parameters are estimated using the SCE-UA method proposed by Duan³⁾ with root mean square error (RMSE) as the objective function and it was minimized to identify the optimal parameters of the USF model. Further, the coefficient of determination, R^2 and Nash-Sutcliffe Efficiency (NSE) were used to assess the USF model simulation efficiency with the identified parameters. The RMSE was used as an indicator for calibration and R^2 and NSE for validation because these are the most basic and widely used evaluation criteria which can reflect the performance in a better way. The authors have already identified the optimal parameters of the USF model and evaluated its efficiency in hydrograph reproducibility with optimal parameters¹⁾.

(3) Study area and data description

The study was conducted in the upper Kanda river basin having an area of 7.7 km² at Koyo Bridge and is shown in Fig.1. The rainfall and water level data during 2003-2006 at one-minute interval were collected from the Bureau of Construction, Tokyo Metropolitan Government for the present study. The basin average rainfall was determined using the Thiessen polygon method from the eight rain gauges scattered over the basin. Five target events were selected from the data, whose 60-minute maximum rainfall (R_{60}) is greater than 30 mm, for the present study and are shown in Table 1.

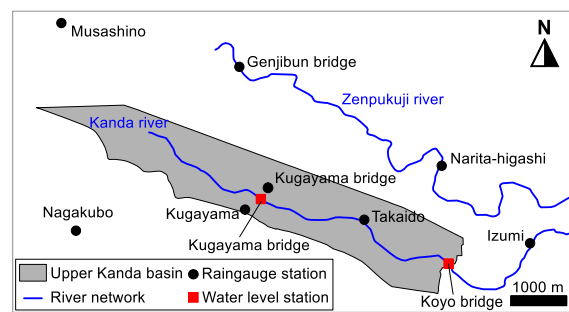


Fig.1 Location map of study area.

Table 1. Characteristics of target events.

Event No.	Event date	R_{60} (mm)	Total R (mm)	Climatic factors
1	13/10/2003	53.9	57.5	Intensive localized storm
2	24~25/6/2003	42.8	55.5	Frontal event
3	8~9/10/2004	42.0	261.1	Typhoon
4	11/09/2006	32.7	37.9	Frontal event
5	15/07/2006	31.5	31.5	Frontal event

Table 2. Description and search range of USF model parameters.

Parameter	Definition	Search range ¹²⁾
k_1	Physical watershed characteristics ¹⁴⁾	[10, 500]
k_2	Loop relationship between the storage and discharge ¹⁵⁾	[100, 5000]
k_3	Groundwater related loss	[0.001, 0.05]
p_1	Flow regime ¹⁴⁾	[0.1, 1]
p_2	Non-linear unsteady flow effects ¹⁵⁾	[0.1, 1]
z	Infiltration hole height	[1, 50]
α	Effect of storm drainage diverted to the treatment plant	[0.1, 1]

3. RESULTS AND DISCUSSIONS

The seven parameters of USF model are shown in **Table 2** with their descriptions and search range¹²⁾ used in the SCE-UA parameter estimation method.

(1) Parameter Uncertainty analysis

The estimated parameter vector, $\hat{\theta}$ was used to perform the model-based bootstrap method and generated 1000 bootstrapped parameter vector of the USF model. **Fig.2** shows the scatter plots of 1000 bootstrapped parameter samples in log scale with their 95% CI. Since we use the SCE-UA global optimisation method for parameter estimation, we represented the search range of SCE-UA method as the lower and upper scales of the y-axis in **Fig.2** and the position of 95% CI within the search range by grey shading with their percentage contribution to the search range. It is apparent from **Fig. 2(a)** that the parameter values of k_1 lies close to the lower search range and converged to a reduced range between 20 and 50. The 95% CI of k_1 was very narrow compared to the search range and it constitutes only 4% of the search range. The scatter plot of parameter k_2 was narrow as shown in **Fig. 2(b)** and the values were clustered around 500 to 1000 which was close to the lower search range. Likewise parameter k_1 , the 95% CI of k_2 was around 4% of the search range. The parameter k_3 has a widespread pattern compared with parameters k_1 and k_2 , and most of the points were concentrated between 0.005 and 0.02 as depicted in **Fig. 2(c)**. The 95% CI of k_3 was quite wide and it represents 23% of the search range. There is a well spread pattern for parameter p_1 from 0.2 to 1 within the search range whereas most of the p_2 values are accumulated near the lower search range in between 0.2 and 0.4 as shown in **Figs. 2(d)** and **(e)** respectively. The 95% CI of these parameters are wide compared with k_1 and k_2 , and they comprise 29% and 45% of search ranges respectively. It is evident from **Fig. 2(f)** that most of the z values are

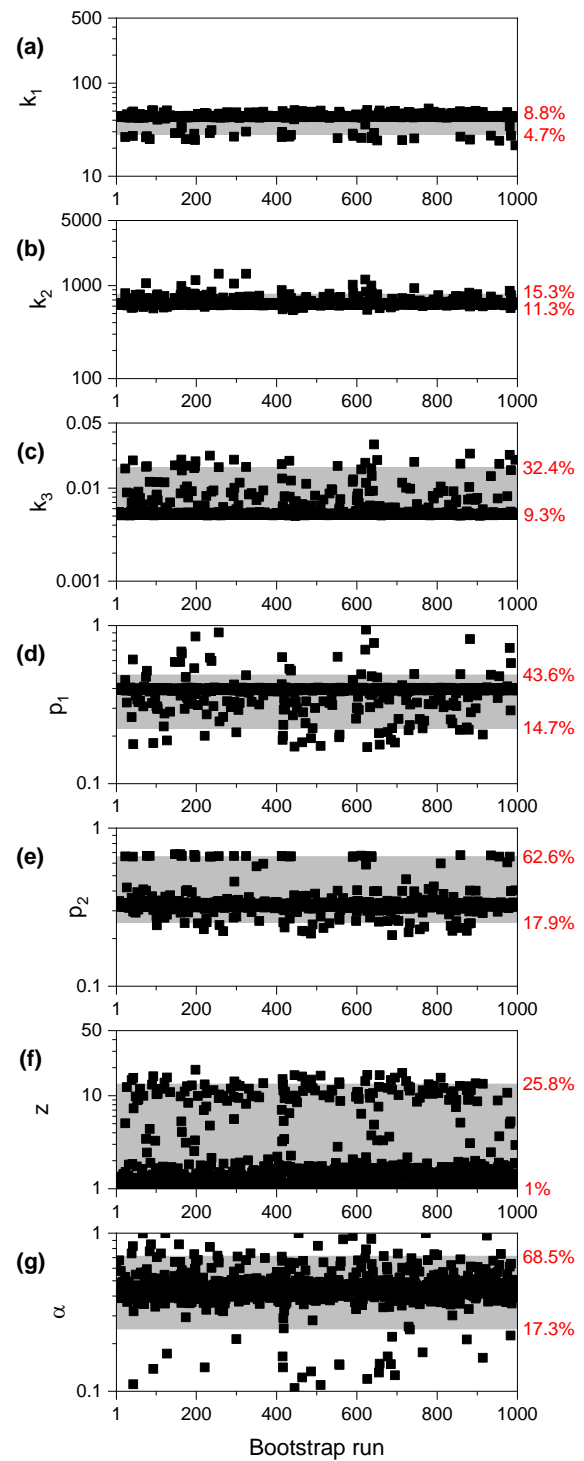


Fig.2 Scatter plots 1000 bootstrapped parameter samples of USF model with their 95% CI.

gathered near one, which is the lower limit of the search range, with a wide range of variation in values from 1 to 20. Similar to k_3 and p_1 , the 95% CI of z was 25% of the search range. Parameter α demonstrated a very widespread pattern beginning from the lower limit of the search range and ends at the upper limit as depicted in **Fig. 2(g)**. It exhibited a very wide 95% CI and it was 51% of the search range.

After obtaining the 95% CI of each parameter, we computed the mean ($\bar{\theta}$), standard deviation (σ_{θ}),

Table 3. Different statistics computed for the bootstrapped parameter sets along with their initial estimated parameter vector $\hat{\theta}$.

Parameter	$\hat{\theta}$	$\bar{\theta}$	σ_{θ}	CV (%)	O _{CV}
k_1	43.78	43.57	3.46	7.96	7
k_2	627.19	647.22	62.28	9.62	6
k_3	0.005	0.006	0.002	43.20	2
p_1	0.398	0.390	0.061	15.71	5
p_2	0.328	0.336	0.065	19.44	4
z	1	2.591	3.49	134.88	1
α	0.424	0.454	0.099	21.92	3

coefficient of variation (CV), and order of CV (O_{CV}) of 1000 bootstrapped parameter vector and is shown in **Table 3**. It is clear from **Table 3** that the bootstrap estimate of mean ($\bar{\theta}$) was close to the estimated parameter values ($\hat{\theta}$) except for parameter z . The highest CV value was observed for parameter z which was around 135%. It was followed by the parameters k_3 and α whose CV values were about 43% and 22% respectively. The remaining parameters exhibited CV values of less than 20% and the least CV value was noted for parameter k_1 . Further, the parameters were ordered based on their CV values and the order of parameters is as follows: $z > k_3 > \alpha > p_2 > p_1 > k_2 > k_1$. The highest and least uncertainty was demonstrated by the parameters z and k_1 respectively based on their CV values.

The parameter z represents the infiltration hole height in USF model¹²⁾ which highly depends upon the height of river storage, rainfall intensity, etc. and will be highly varying from event to event. Also, most of the time, the optimum value of z estimated from the 1000 bootstrapped discharge data sets by SCE-UA method is the lower limit of the search range which is one as shown in **Fig. 2(f)**. This further indicates that the search range recommended by Takasaki et al.¹²⁾, which was set based on the physical minimum and maximum values of parameters, is not sufficient and reconsideration of the search range is necessary. Parameter k_3 is associated with z to depict the groundwater related loss as shown in Eq. (4) and there is a chance of high correlation between these two parameters which will lead to high uncertainty in k_3 values after z . Additionally, parameter α in Eq. (5) has a direct impact on the outflow from the basin through the combined sewer system and is vulnerable to changes from time to time and got a higher uncertainty after z and k_3 . The parameters p_2 , p_1 , and k_2 are influenced by the changes in flow conditions^{14), 15)} and are prone to variations in different degrees. The parameter k_1 describes the watershed features¹⁴⁾ and the chance to change this watershed features within a short span is low. Hence, this could be a reason for the least uncertainty exhibited by parameter k_1 .

(2) Model simulation uncertainty

The R^2 and NSE were used to assess the USF model simulation efficiency and **Fig.3** shows the linear regression of the observed and simulated data. For easy understanding, the $x = y$ line is also plotted as a reference in blue colour. From **Fig.3**, it can be observed that there is a lower prediction of discharge values greater than 0.3 mm/min, whereas the low flows exhibit almost close fit except for very few data points in a loop shape. The R^2 and NSE values were found to be 0.97 and 97.6% respectively which indicate that the SCE-UA method has successfully identified the optimal parameters for the model. The reproduced hydrographs for events 1 and 3 are shown in **Fig.4** (in red colour as simulated) and the model was capable to reproduce the shape as well as the peak discharge which makes the model parsimonious.

Furthermore, the model simulation uncertainty range due to the parameter uncertainty was computed by estimating the 95% CI of the 1000 bootstrapped discharge series samples. **Fig.4** shows the simulated uncertainty range due to the parameter uncertainty of the model for two selected events (events 1 and 3) out of five due to page limitation. It is desirable to have a narrow range and can be seen from **Fig.4** that the range is wide at the peak flows compared to the range at the low flows, and hence it can be envisaged that the model simulates peak discharge with high uncertainty compared with low flows. This can be attributed to the uncertainties involved in the rainfall during high flows. The greater the number of observations fall within the uncertainty interval, the greater the capability of the model to reasonably capture the observed discharge. The percentage of observed discharge falling inside the ranges are 35% and 63.2% for events 1 and 3 respectively. Likewise, the observations falling inside the range during remaining events 2, 4, and 5 are 76.7%, 87.5%, and 78.1% respectively. Therefore, the model is good in predicting the observations within this range except for event 1. During event 1, the model uncertainty range cannot cover most of the observed values and this can be ascribed to the highly fluctuating rainfall

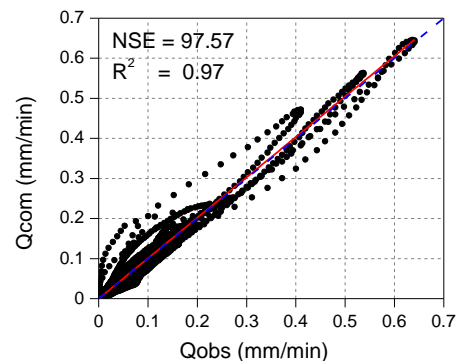


Fig.3 Linear regressions of observed and simulated discharge for USF model.

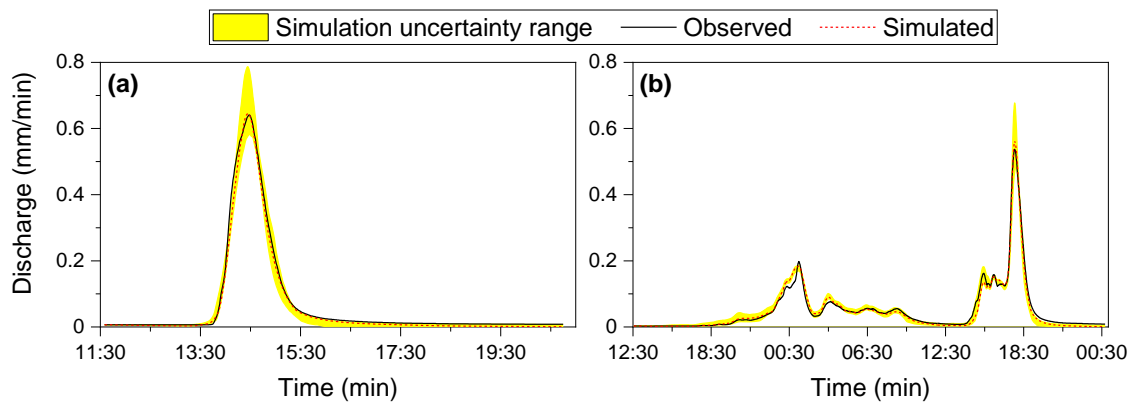


Fig.4 Simulation uncertainty range due to parameter uncertainty for (a) event 1, and (b) event 3.

pattern which is an intensive localised storm. On average, 68% of the observations lay within this prediction interval of the model. This indicates that the SCE-UA method along with the model-based bootstrap reasonably well predicted the uncertainty range of USF model, and captured most of the observed discharge within this range.

4. CONCLUSIONS

The model-based bootstrap technique was utilised to analyse the parameter uncertainty of USF model in order to assess their impact on the model simulation in the upper Kanda River basin, an urban watershed in Tokyo. The 95% CI obtained from 1000 bootstrapped parameter samples showed a wider confidence band for most of the parameters. Further, parameter z was identified with the highest uncertainty by comparing the bootstrap estimate of CV, and the uncertainty of other parameters was relatively low. Additionally, the effect of parameter uncertainty on the model simulation uncertainty was investigated by computing the 95% CI of 1000 bootstrapped discharge series. The number of observations falling inside the 95% CI was found to be 68% which further disclosed that the model-based bootstrap can reasonably predict the observed discharge within the uncertainty range of USF model.

This study primarily focuses on the model-based bootstrap approach for the parameter uncertainty analysis. However, it is also necessary to compare this technique with other resampling techniques, with redefined SCE-UA search ranges, to check the reliability of the assumptions made.

ACKNOWLEDGEMENT: This study was carried out as a part of the research project entitled “Study on guerrilla rainstorm, flood, and water pollution in megacity urban watersheds - Countermeasures against megacity urban water-related disasters bipolarized by climate change” supported by Tokyo Metropolitan Government, Japan (Represented by Prof. Akira Kawamura).

REFERENCES

- 1) Padiyedath, G. S., Kawamura, A., Takasaki, T., Amaguchi, H. and Azhikodan, G.: An effective storage function model for an urban watershed in terms of hydrograph reproducibility and Akaike information criterion, *J. Hydrol.*, Vol. 563, pp. 657-668, 2018. doi:10.1016/j.jhydrol.2018.06.035.
- 2) Li, Z., Shao, Q., Xu, Z., and Cai, X.: Analysis of parameter uncertainty analysis in semi-distributed hydrological models using Bootstrap method: a case study of SWAT model applied to Yingluoxia watershed in Northwest China. *J. Hydrol.*, Vol. 385, pp. 76–83, 2010.
- 3) Duan, Q., Sorooshian, S., and Gupta, V.: Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, *Water Resour. Res.*, Vol. 28 (4), pp. 1015-1031, 1992.
- 4) Selle, B., and Hannah, M.: A bootstrap approach to assess parameter uncertainty in simple catchment models. *Environ. Model. Softw.*, Vol. 25, pp. 919–926, 2010.
- 5) Efron, B.: Bootstrap methods: another look at the jackknife. *Ann. Statist.* Vol. 7, pp. 1-26, 1979.
- 6) Dixon, P. M.: Bootstrap Resampling, in: Encyclopedia of Environmetrics. John Wiley & Sons, Chichester, UK, 2006.
- 7) Sun, J., Liu, Y., Wang, Y., Bao, G., and Sun, B.: Tree-ring based runoff reconstruction of the upper Fenhe River basin, North China, since 1799 AD. *Quat. Int.* Vol. 283, pp. 117–124, 2013.
- 8) Jeong, D., and Kim, Y. O.: Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. *Hydrol. Processes*, Vol. 19 (19), pp. 3819–3835, 2005.
- 9) Stine, R. A.: Bootstrap prediction intervals for regression. *J. Am. Statist. Assoc.* Vol. 80 (392), pp. 1026–1031, 1985.
- 10) Davison, A. C., and Hinkley, D. V.: Bootstrap methods and their application. Cambridge University Press, 1997.
- 11) Sonali, P., and Nagesh Kumar, D.: Review of trend detection methods and their application to detect temperature changes in India. *J. Hydrol.* Vol. 476, pp. 212–227, 2013.
- 12) Takasaki, T., Kawamura, A., Amaguchi, H. and Araki, K.: New storage function model considering urban runoff process, *Journal of JSCE, B*, Vol. 65 (3), pp. 217-230, 2009.
- 13) Padiyedath, G. S., Kawamura, A., Takasaki, T., Amaguchi, H. and Azhikodan, G.: Performance evaluation of urban storage function (USF) model compared with various conventional storage function models for an urban watershed, *J. JSCE, Ser. B1*, Vol. 74 (4), pp. 973–978, 2018.
- 14) Sugiyama, H., Kadoya, M., Nagai, A. and Lansley, K.: Evaluation of the storage function model parameter characteristics, *J. Hydrol.*, Vol. 191, pp. 332-348, 1997.
- 15) Hoshi, K. and Yamaoka, H.: A relationship between kinematic wave and storage routing models. *Proc. 26th Japanese Conf. on Hydraulics, JSCE*, pp. 273-278, 1982.

(Received May 31, 2018)