

# An Extended Model Evaluation Method under Uncertainty in Hydrologic Modeling

Giha Lee<sup>1)</sup> · Sangkuk Youn<sup>†</sup> · Yeonsu Kim<sup>1)</sup>

Received: February 4<sup>th</sup>, 2015; Revised: February 11<sup>th</sup>, 2015; Accepted: April 20<sup>th</sup>, 2015

**ABSTRACT :** This paper proposes an extended model evaluation method that considers not only the model performance but also the model structure and parameter uncertainties in hydrologic modeling. A simple reservoir model (SFM) and distributed kinematic wave models (KWMSS1 and KWMSS2 using topography from 250-m, 500-m, and 1-km digital elevation models) were developed and assessed by three evaluative criteria for model performance, model structural stability, and parameter identifiability. All the models provided acceptable performance in terms of a global response, but the simpler SFM and KWMSS1 could not accurately represent the local behaviors of hydrographs. Moreover, SFM and KWMSS1 were structurally unstable; their performance was sensitive to the applied objective functions. On the other hand, the most sophisticated model, KWMSS2, performed well, satisfying both global and local behaviors. KWMSS2 also showed good structural stability, reproducing hydrographs regardless of the applied objective functions; however, superior parameter identifiability was not guaranteed. A number of parameter sets could result in indistinguishable hydrographs. This result indicates that while making hydrologic models complex increases its performance accuracy and reduces its structural uncertainty, the model is likely to suffer from parameter uncertainty.

**Keywords :** Extended model evaluation, Model performance, Model structural stability, Parameter identifiability

## 1. Introduction

A primary objective of hydrologic modelers is to identify an appropriate hydrologic model and optimal parameter set that are suitable for the modeling purpose, catchment characteristics, and available data (Wagener, 2003; Wagener et al., 2004). Beven (2001) reviewed the basic criteria for model choice, summarized as follows: 1) model availability; 2) model predictability for hydrological variables; 3) model reliability (*i.e.*, whether the assumptions underlying the model structures are understandable according to the modeler's expertise/experience); and 4) model suitability within the time and cost constraints of the modeling objectives. However, he warned that all available models will be easily rejected by these evaluative criteria because of inadequate model conceptualization and insufficient field data to fully support the model parameters. Therefore, a more effective and practical guideline is necessary to enable modelers to identify the best model providing more accurate and less uncertain prediction results.

Conventional model evaluation usually judges models in a one-dimensional manner based only on model performance

and does not consider various uncertainty sources that propagate into the prediction results. As a result, many different model structures are often judged as equally good representations of catchment hydrological responses even though some models are overly simplified (Refsgaard & Knudsen, 1996; Uhlenbrook et al., 1999; Beven & Freer, 2001). Moreover, numerous plausible parameter combinations can exist within a feasible parameter space and provide identically good model performance measures or indistinguishable hydrographs in rainfall-runoff modeling (Beven & Binley, 1992). This phenomenon, called "equifinality" (Beven & Binley, 1992; Beven, 2006), has been recognized as a modeling issue by hydrological modeling communities such as the International Working Group on Uncertainty Analysis in Hydrologic Modeling, a part of the Predictions in Ungauged Basins (PUB) initiative of the International Water Management Institute. Beven (2001, 2002) outlined an alternative strategy for model identification considering uncertainty, which Wagener et al. (2001, 2003a, 2004) then developed by incorporating a parsimonious model with uncertainty analysis tools such as the multi-objective complex evolution method (MOCOM, Yapo et al., 1998) and dynamic

1) Department of Construction and Disaster Prevention Engineering, Kyungpook National University

† Department of Construction and Disaster Prevention Engineering, Kyungpook National University (Corresponding Author : ysk9771@lycos.co.kr)

identifiability analysis (DYNIA, Wagener et al., 2003b). Wagner and colleagues abandoned the notion of “uniqueness”, which aims to obtain a true representation of a hydrological system by calibration and validation steps. Instead, they emphasized the need for enhanced model identification to provide more information for either confirming possible predictor(s) or rejecting inadequate one(s).

In this regard, this paper proposes an extended model evaluation framework under uncertainty in rainfall-runoff modeling for identifying a more reliable model. The new framework follows the basic concepts of uncertainty proposed by Beven (2002) and Wagener & Gupta (2005). It admits numerous plausible representations providing identically good model performance measures, while newly developed criteria are used to assess other inherent model characteristics related to structural and parameter uncertainties. We prepared seven different rainfall-runoff models ranging from a simple lumped model to sophisticated distributed models and then evaluated the models with respect to model performance, model structural stability, and parameter identifiability. A highly ranked model by these criteria is structurally stable, shows less parameter uncertainty, and ensures accurate prediction results. This evaluation process may provide a more useful guideline for selecting a suitable model for various rainfall-runoff model applications. Section 2 introduces the concept underlying the new method of model identification under uncertainty, and Section 3 describes the models used in this study. Section 4 introduces the new evaluative criteria in detail and addresses the comparative results. Finally, we summarize our major conclusions in Section 5.

## 2. Concept of Extended Model Evaluation under Uncertainty

Fig. 1 illustrates the extended model evaluation under uncertainty. Initially, a set of rainfall-runoff models, with different representations of rainfall-runoff processes and spatial topography, is prepared for model evaluation. Here, all the models are assumed to be potentially available simulators, unless obvious evidence indicates that a model should be rejected.

Three different evaluative criteria are then applied to the

competing models. The first (or the most fundamental) measure of model evaluation is the model performance index (MPI), which assesses whether the models are capable of accurately simulating the observed streamflow in terms of local response modes such as low and high flows. The second criterion is the model structural stability index (MSSI) for assessing how precisely the models can represent the local response modes regardless of given objective functions. More stable models can provide more constant and accurate simulation results with respect to various local behaviors irrespective of the applied objective functions. The last measure is the model parameter identifiability index (MPII) for evaluating whether the model parameters are well identified within a predefined feasible parameter space. The models showing higher parameter identifiability indicate less parameter uncertainty during model calibration and guarantee increased prediction accuracy. The resulting criteria values from the extended model evaluation

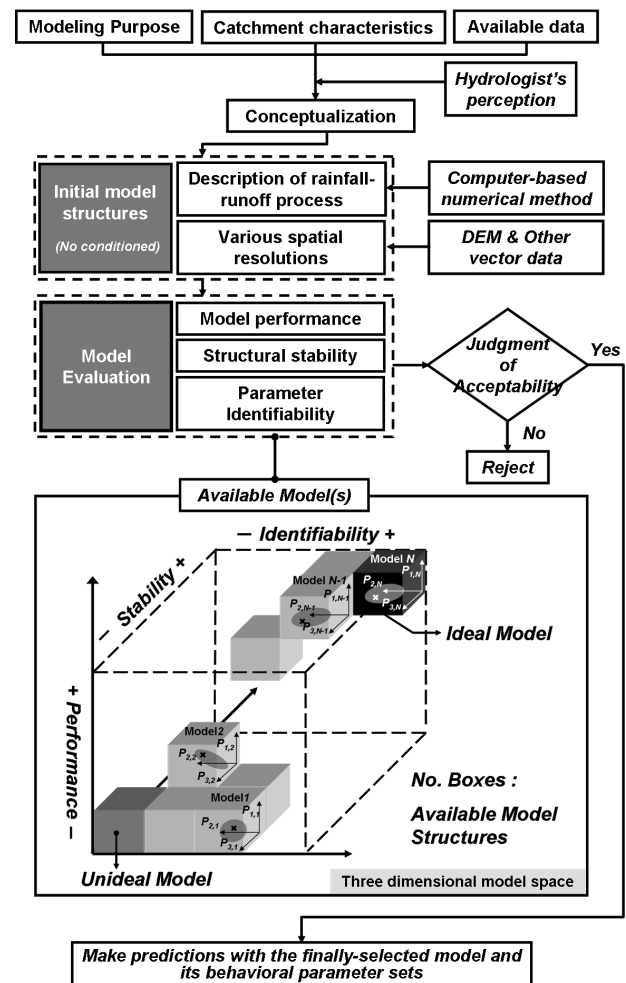


Fig. 1. Schematic diagram of the extended model evaluation under uncertainty in rainfall-runoff modeling

should give some objective basis by which to search for a model that balances prediction accuracy, structural stability, and parameter uncertainty.

Fig. 1 also presents the schematic model space with respect to the three proposed evaluative criteria; the plus marks indicate “good” results while the minus marks indicate “bad” results. In this model space, each box indicates candidate of acceptable rainfall-runoff models. The black box indicates the best model leading to good model performance, stable model structure, and high parameter identifiability, while the dark gray box on the bottom left in the three-dimensional model space is the “nonideal” model. As shown in Fig. 1, the new model evaluation process emphasizes that one-dimensional model evaluation, which is based only on a single criterion, is likely to result in many possible predictors. For example, both Model  $N$  (the ideal model) and Model  $N-1$  can provide equally good MPI values, but Model  $N-1$  is worse than Model  $N$  in terms of both MSSI and MPII values. Additional criteria in the proposed model evaluation procedure can provide richer information, enabling modelers to identify a balanced model, which can lead to a more accurate and less uncertain prediction result, among a number of possible simulators.

A model structure with the best balance (or the final selected model) may provide an assurance of good parameter identifiability. However, it is not completely free from parameter uncertainty. This means that such a model has plausible (or behavioral) parameter sets that yield similarly good outcomes. In the three-dimensional model space of Fig. 1, the light gray elliptical regions of each box (*i.e.*,  $P_{ij}$  is a model parameter, where  $i$  is the number of parameters to be calibrated and

$j$  is the number of possible model candidates) schematically indicate the constrained parameter spaces after model calibration. All parameter sets within these regions provide objective function values as good as the global optimal sets marked by the symbol  $x$ . Note that each model perhaps has different parameter dimensions, but for visualization purposes, the parameter spaces of all models are represented in three dimensions. The behavioral parameters (inside the constrained parameter space) sometimes lose physical meaning numerically with respect to their value, but they are meaningful in terms of model predictions. As a consequence, the surviving model with its behavior parameter sets should be retained until parameter sets that violate new evaluative criteria are found. Then, the retained parameter sets are used for runoff prediction. The prediction results of the selected model are therefore not a single output sequence but a set of hydrographs.

### 3. Rainfall–runoff Models used in this Study

In this study, three different types of rainfall-runoff models, from a simple lumped model to distributed kinematic wave models, were developed under an object-oriented hydrological modeling system (Takasao et al., 1996; Ichikawa et al., 2000). Moreover, three different spatial resolutions of a digital elevation model (DEM) were used to investigate the scale effect on both model performance and uncertainty assessment in distributed rainfall-runoff modeling. The following subsections provide more details of the models. All the models were

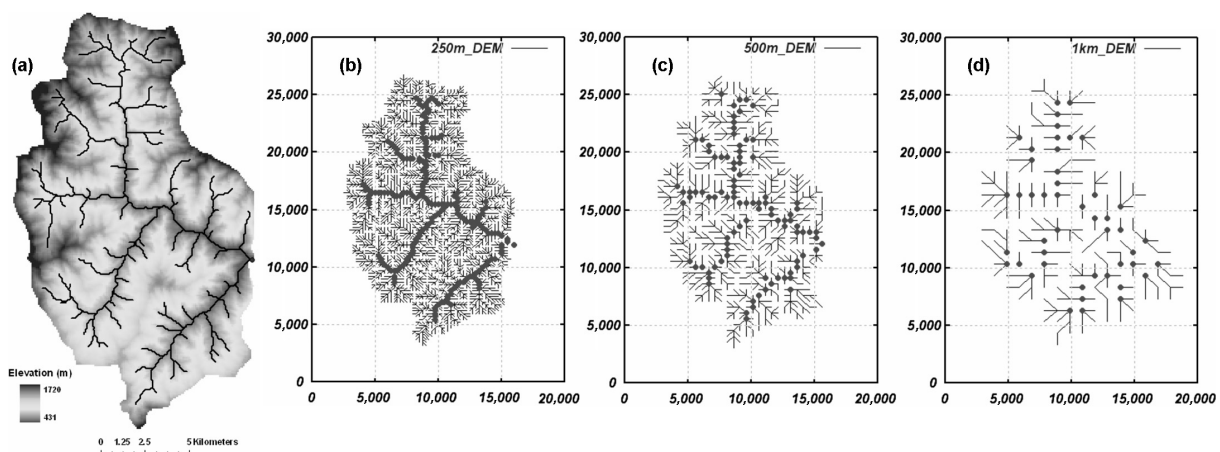


Fig. 2. (a) The Kamishiiba catchment and modeled drainage networks of (b) 250–m, (c) 500–m, and (d) 1–km DEMs for distributed rainfall–runoff modeling

applied to the Kamishiiba catchment in Kyushu, Japan, as illustrated in Fig. 2 (a). The drainage networks of the study site were represented by DEMs with 250-m, 500-m, and 1-km resolutions for distributed rainfall-runoff modeling. The study area is upstream from Kamishiiba Dam and covers an area of 211 km<sup>2</sup>. The catchment has hilly topography, with elevations varying from 431 to 1,720 m. Most of the land is forested. Observed discharge data, converted from the water level of the dam inflow with 10-min temporal resolution, and radar rainfall data observed from the Ejiroyama X-band radar, covering a radius of 128 km, are available for this area. Spatially distributed rainfall data for the historical flood event caused by Typhoon No. 9 (15-19 September 1997) were used for both the rainfall-runoff simulations and a multifarious model evaluation; the radar rainfall data has 1 km × 1 km spatial and 10-min temporal resolutions.

### 3.1 Storage Function Method (SFM)

The SFM is a simple nonlinear reservoir model that is widely used for practical engineering applications in Japan despite its simplicity. The form of SFM is expressed as

$$\frac{dS}{dt} = r_e(t - T_l) - q, \quad S = kq^p \quad (1)$$

$$r_e = \begin{cases} f \times r, & \text{if } \sum r \leq R_{SA} \\ r, & \text{if } \sum r > R_{SA} \end{cases} \quad (2)$$

where  $S$  [m<sup>3</sup>] is the water storage,  $r_e$  [mm/hr] is the effective rainfall intensity,  $r$  [mm/hr] is the rainfall intensity,  $q$  [m<sup>3</sup>/hr] is the runoff,  $t$  is time,  $k$  [-] is the storage coefficient,  $p$  [-] is the coefficient of nonlinearity,  $f$  [-] is the primary runoff ratio,  $T_l$  [hr] is the lag time, and  $R_{SA}$  [mm] is the accumulated saturated rainfall. Four parameters ( $k$ ,  $p$ ,  $f$ , and  $R_{SA}$ ) need to be optimized in SFM.

### 3.2 Kinematic Wave Method for Subsurface and Surface Runoff with a Single Threshold (KWMSS1)

In this model, the catchment surface is assumed to be covered with a highly permeable stratum called the ‘‘A-layer’’ with uniform thickness  $D$  [m]. The depth,  $d$  [m], is referred

to as a threshold to account for the surface and subsurface flow, and is defined as  $d = D\gamma$ , where  $\gamma$  is the porosity of the A-layer (see Fig. 3). Takasao & Shiiba (1988) proposed the following piecewise relationship between water depth  $h$  [m] and discharge per unit width  $q$  [m<sup>2</sup>/s]:

$$q = \begin{cases} vh, & 0 \leq h \leq d \\ vh + \alpha(h-d)^m, & d < h \end{cases} \quad (3)$$

where  $v = k_d i$ ,  $\alpha = \sqrt{i}/n$  [m<sup>1/3</sup>s<sup>-1</sup>],  $v$  [m/s] is the velocity through the A-layer,  $i$  is the slope gradient,  $k_d$  [m/s] is the saturated hydraulic conductivity, and  $n$  [m<sup>-1/3</sup>s] is Manning’s roughness coefficient; if the overland flow follows Manning’s resistance law, then  $m = 5/3$ . Three model parameters ( $n$ ,  $k_d$ , and  $d$ ) must be adjusted to the observed data.

### 3.3 Kinematic Wave Method for Subsurface and Surface Runoff with Double Thresholds (KWMSS2)

Like the preceding KWMSS1, this model also assumes that a catchment is covered with a permeable soil layer. However, instead of the single threshold,  $d$  is partitioned into three flow components by two thresholds, the depth  $d_c$  [m] corresponding to the water content in the capillary pores, and the depth  $d_s$  [m] corresponding to the maximum water content,

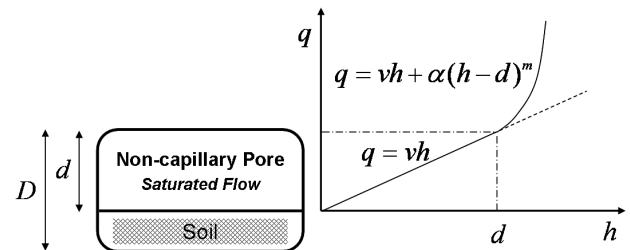


Fig. 3. Schematic model structure and the stage-discharge relationship of KWMSS1

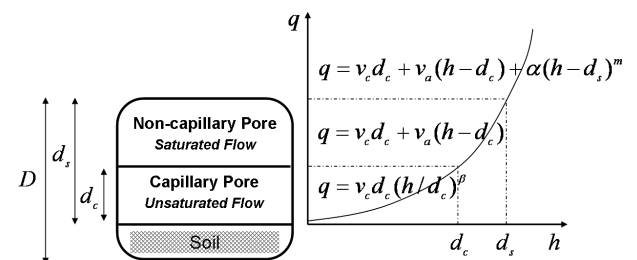


Fig. 4. Schematic model structure and the extended stage-discharge relationship of KWMSS2

as illustrated in Fig. 4. The depth of the capillary pore layer is referred to as the domain for simulation of unsaturated flow. After the water depth reaches it, the capillary pore layer is assumed to be saturated and gravity flow occurs in the non-capillary pore layer. Finally, when the water depth begins to exceed total subsurface water capacity, surface flow occurs. Tachikawa et al. (2004) modified the relationship between depth and flow to account for surface and subsurface runoff systems and then proposed the following extended stage-discharge relationship:

$$q = \begin{cases} v_c d_c (h/d_c)^\beta, & 0 \leq h \leq d_c \\ v_c d_c + v_a (h - d_c), & d_c < h \leq d_s \\ v_c d_c + v_a (h - d_c) + \alpha (h - d_s)^m, & d_s < h \end{cases} \quad (4)$$

where  $v_c = k_c i$  [m/s],  $v_a = k_a i$  [m/s],  $k_c = k_a / \beta$  [m/s],  $\alpha = \sqrt{i} / n$  [ $m^{1/3} s^{-1}$ ],  $k_c$  [m/s] is the hydraulic conductivity of the capillary soil layer, and  $k_a$  [m/s] is the hydraulic conductivity of the non-capillary soil layer. There are five parameters ( $n$ ,  $k_a$ ,  $d_s$ ,  $d_c$ , and  $\beta$ ) that need to be calibrated in KWMSS2.

The flow rate of KWMSS1 and KWMSS2 is calculated by the stage-discharge relationships written as Eqs. (3) and (4), combined with the continuity equation, Eq. (5), assuming that the hydraulic gradient is parallel to the slope in a hilly area:

$$\frac{\partial h}{\partial t} + \frac{\partial q}{\partial x} = r(t) \quad (5)$$

where  $r(t)$  [mm/s] indicates the lateral inflow, given here as rainfall over the catchment.

In both kinematic wave models, rainfall over the catchment is directly added to both the subsurface flow and surface flow, so that it is represented by the water depth,  $h$ . The stage-discharge relationship stated above effectively produces a lag time when modeling the subsurface flow without adding a vertical infiltration model to the rainfall-runoff models.

As discussed above, one simple nonlinear reservoir model (SFM) and six distributed models (KWMSS1 with 250-m, 500-m, and 1-km DEMs and KWMSS2 with 250-m, 500-m, and 1-km DEMs) were evaluated under the new model evaluation framework.

## 4. Model Evaluation with Three Different Types of Criteria

### 4.1 Model Evaluation with the Model Performance Index (MPI)

Model performance is a principal benchmark not only for selecting a model but also for presenting convincing model results to other hydrologists or stakeholders. Performance is typically judged using an objective function that is minimized or maximized according to modeling purposes. A wide range of statistical and hydrological objective functions is available. Boyle et al. (2000, 2001) pointed out that overall measures such as the root mean-square error (RMSE) could capture global behavior, but such aggregation of error would likely decrease the amount of information in the data such that various local behaviors involved in hydrological responses could be overlooked. Therefore, they recommended that the runoff time series be partitioned into specific response periods to investigate the influence of individual model parameters on both global and local behaviors. Wagener et al. (2003a) also demonstrated that while some models could reproduce specific local behaviors (*e.g.*, peak or rising/recession flows), their global behaviors (*i.e.*, overall model performances) were not acceptable.

To assess model performance by the new model evaluation framework, we partitioned hydrographs into two components: high flow and low flow periods divided by the threshold, 315 m<sup>3</sup>/s, defined as the mean value of the observed discharge data (see Fig. 5 (a)). The performances of each model structure were assessed using the Nash-Sutcliffe coefficient (NSC) for the two periods; the two measures were then averaged to obtain the MPI, defined as

$$NSC_{High} = 1 - \frac{\sum_{t=1}^{N_{High}} (q_{obs}(t) - q_{sim}(t))^2}{\sum_{t=1}^{N_{High}} (q_{obs}(t) - \bar{q}_{obs}^{High})^2} \quad (6)$$

$$NSC_{Low} = 1 - \frac{\sum_{t=1}^{N_{Low}} (q_{obs}(t) - q_{sim}(t))^2}{\sum_{t=1}^{N_{Low}} (q_{obs}(t) - \bar{q}_{obs}^{Low})^2} \quad (7)$$

$$MPI = 0.5(NSC_{High} + NSC_{Low}) \quad (8)$$

where  $q_{obs}(t)$  is the observed discharge at time step  $t$ ,  $q_{sim}(t)$  is the simulated discharge, and  $\bar{q}_{obs}^{High}$  and  $\bar{q}_{obs}^{Low}$  are the mean observed discharge over simulation periods of lengths  $N_{High}$

and  $N_{Low}$ , respectively.

Here, all models were calibrated using the shuffled complex evolution algorithm developed at the University of Arizona (SCE-UA; Duan et al., 1992; 1993; 1994) with the objective function of simple least-squares (SLS), expressed as

$$SLS = \sum_{t=1}^N (q_t^{obs} - q_t(\theta))^2 \quad (9)$$

where  $q_t^{obs}$  is the observed streamflow value at time  $t$ ,  $q_t(\theta)$  is the simulated streamflow value at time  $t$  using parameter set  $\theta$ , and  $N$  is the number of flow values available.

The SCE-UA is a single-objective optimization method designed to handle the high parameter dimensionality encountered in the calibration of a nonlinear hydrologic simulation model. Numerous researchers have applied this evolutionary method to a variety of hydrologic models and have proven the method to be efficient and potent for automatic optimization (e.g., Gan et al., 1997; Yu et al., 2001).

The calibrated parameter values were used for hydrograph simulations of each model, and the model performances were then evaluated by comparisons of both the simulated hydrographs and the MPI values. Fig. 5 (a) shows the simulation results for SFM, and Figs. 5 (b) and (c) present KWMSS1 and KWMSS2 results for the various DEM scales. Table 1 summarizes the results statistically.

The table shows that all models produced quantitatively acceptable MPI values larger than 0.93. KWMSS1 provided a good fit during the high flow period but could not accurately reproduce the rising and recession limbs of the hydrograph. All the MPI values for the low flow period were underestimated when compared to the results of the high flow period. In addition, the peak discharge of KWMSS1 was sensitive to the applied DEM sizes; it gradually increased as the spatial resolution of the DEM became coarser. This result means that the underlying assumption (or conceptualization) of the model structure was not appropriate for simulating low flow and thus

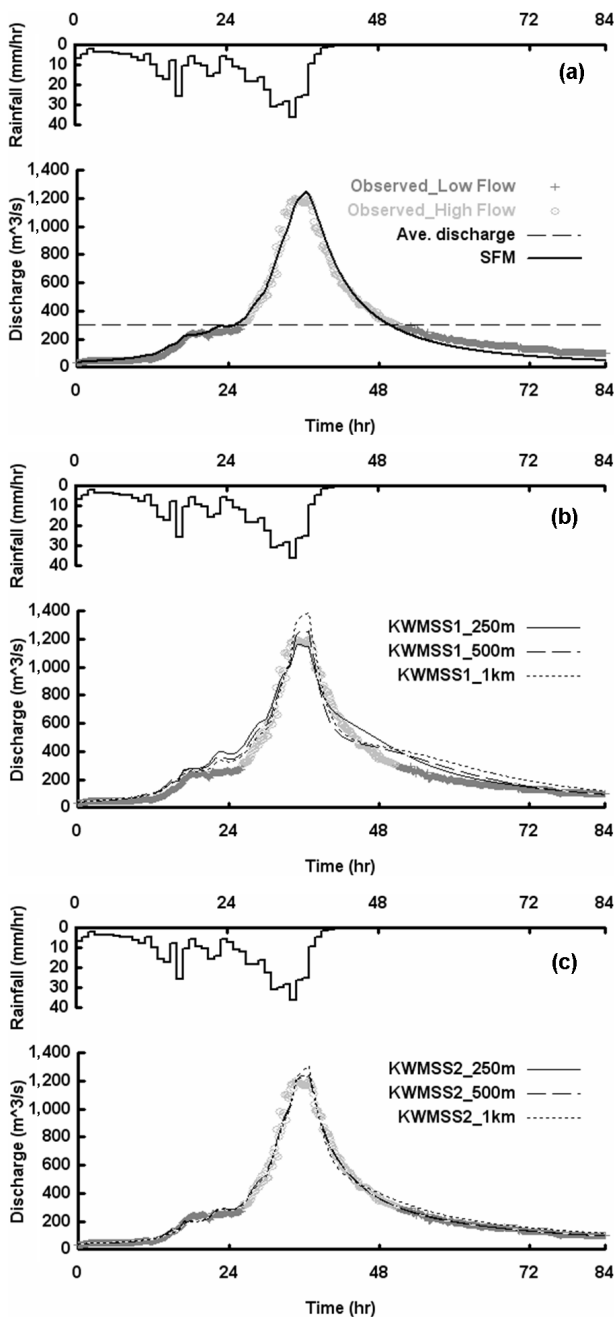


Fig. 5. Model performances of (a) SFM, (b) KWMSS1, and (c) KWMSS2 based on different DEMs with resolutions ranging from 250-m to 1-km

Table 1. MPI values for each model structure: evaluation of global and local behaviors

No. parameter	Model	SFM	KWMSS1			KWMSS2		
			250 m	500 m	1 km	250 m	500 m	1 km
	$NSC_{Low}$	0.944	0.885	0.885	0.812	0.993	0.993	0.974
	$NSC_{High}$	0.987	0.969	0.969	0.967	0.993	0.992	0.983
	MPI	0.977	0.95	0.95	0.931	0.993	0.993	0.981

needs further modification. On the other hand, even though SFM was the simplest model, its simulation results were more accurate than those of the distributed model KWMSS1 for both periods. This finding supports the idea that if hydrologists have data of sufficient quality and quantity and a well-calibrated model, they do not need to choose a complicated hydrologic model for streamflow estimation because it will not necessarily produce an outstanding improvement (Jakeman & Hornberger, 1993). KWMSS2 reproduced more balanced hydrographs than the other models used here, and its results were equally good regardless of the applied DEM sizes; in particular, the peak discharge changed little with the different DEMs.

#### 4.2 Model Evaluation with the Model Structural Stability Index (MSSI)

Structural error is an unavoidable problem in hydrological modeling since a hydrologic model is only a conversion and simplification of reality; thus, models merely represent conceptual or empirical aspects, as designed by modelers. Consequently, the output time series of hydrologic models are as reliable as the hypotheses, model structure, quantity and quality of available data, and parameter estimates. Gupta et al. (1998, 2003) showed that a major consequence of structural inadequacy is the inability of a rainfall-runoff model to reproduce the entire (or global) hydrologic behavior with a single optimal parameter set estimated by traditional single-objective optimization algorithms. In other words, the subjective selection of an objective function for calibration of structurally imperfect models may result in overemphasis of different local response modes in the estimated hydrographs. This finding implies that different parameter combinations are required to represent specific local behaviors of the real rainfall-runoff system (Wagener et al., 2004). Yapo et al. (1998) and Vrugt et al. (2003) developed effective and efficient algorithms called multi-objective complex (MOCOM-UA) and multi-objective shuffled complex evolution metropolis (MOSCEM-UA) algorithms, respectively, for assessing the structural uncertainty.

As documented by Gupta et al. (2003) and Yapo et al. (1998), a better (or more stable) model structure results in both smaller Pareto sets and more improved values with respect to the given objective functions during calibration trials. This implies that a structurally stable model can be regarded as a model representing constant and accurate hydrologic behaviors

regardless of the objective function. Therefore, the level of performance consistency for various objective functions can serve as an indicator for assessing the model's structural stability. In this study, two objective functions having different characteristics were used to evaluate the structural stability of models. The first was SLS, as described in the preceding section. In this objective function, residuals between the observed and simulated discharge are evenly weighted across an event; thus a parameter set, which matches well around the peak discharge, can be obtained. The second objective function was the heteroscedastic maximum likelihood estimator (HMLE); this is the most successful form of the maximum likelihood criteria, which properly accounts for non-stationary variance in streamflow measurement errors (Sorooshian & Dracup, 1980). This measure incorporating weight provides a more balanced performance across the entire flow range and is calculated as

$$HMLE = \frac{1}{N} \frac{\sum_{t=1}^N w_t \varepsilon_t}{\prod_{t=1}^N w_t} \quad (10)$$

where  $\varepsilon_t = q_t^{obs} - q_t(\theta)$  is the model residual at time  $t$ ,  $w_t$  is the weight assigned to time  $t$  computed as  $w_t = f_t^{2(\lambda-1)}$ ,  $f_t$  is the expected true flow at time  $t$ , and  $\lambda$  is the transformation parameter which stabilizes the variance. Yapo et al. (1998) recommended the use of  $f_t$  as observed flow for a more stable estimation.

Finally, the MSSI is formulated in the form of the RMSE between both simulated discharges based on the optimal parameter sets estimated by the SCE-UA with SLS and HMLE, expressed as

$$MSSI = \sqrt{\frac{1}{N} \sum_{t=1}^N (q_{SLS}(t) - q_{HMLE}(t))^2} \quad (11)$$

where  $N$  is the total number of simulation time steps, and  $q_{SLS}(t)$  and  $q_{HMLE}(t)$  are the simulated discharges using the optimal parameters of SLS and HMLE, respectively. Here, a lower value of the MSSI indicates a more stable model structure.

Figs. 6 (a) and (b) show that the parameters based on SLS produce better hydrographs than the cases based on HMLE

in SFM and KWMSS1; SLS-based optimal parameter sets reproduce the peak discharge well, while HMLE-based optimal parameter sets provide improved results during the low-flow period. In contrast, the objective functions have only very small or no influence on the performances of KWMSS2 (see Fig. 6 (c)).

Notably, as summarized in Table 2, the models with finer DEM resolution led to smaller differences between the reproduced hydrographs in distributed rainfall-runoff modeling using KWMSS1 and KWMSS2; the MSSI values decreased

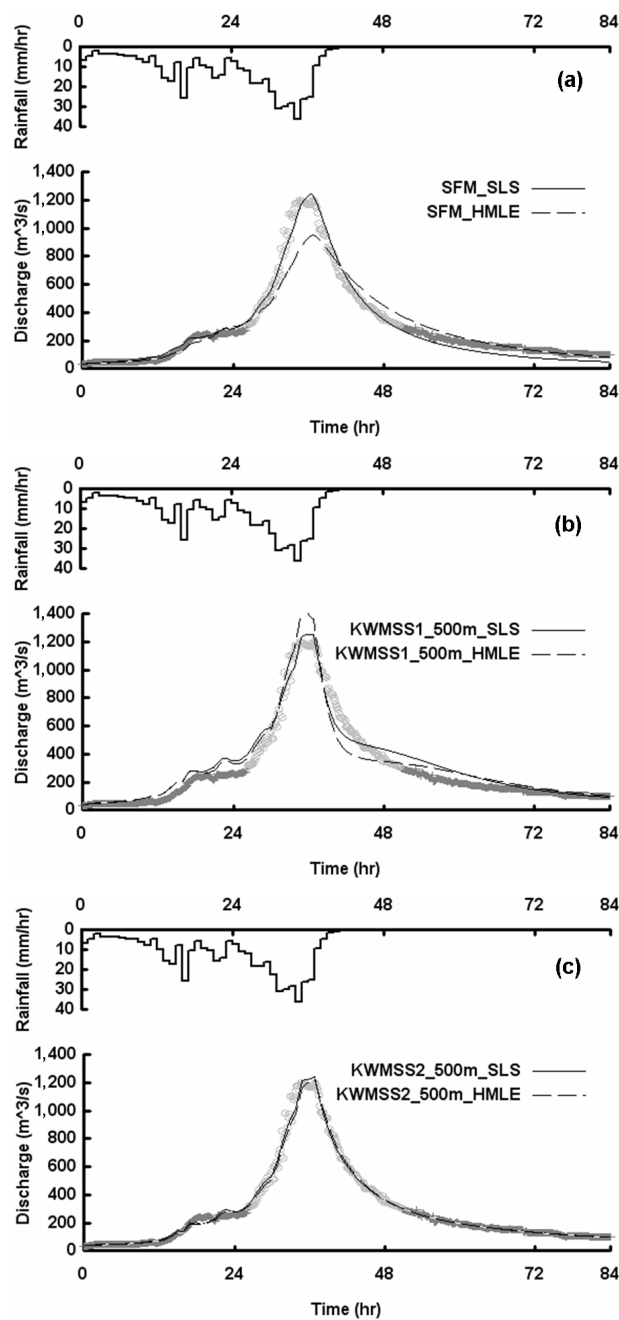


Fig. 6. Assessment of the model structural stability of (a) SFM, (b) KWMSS1, and (c) KWMSS2 based on the 500-m DEM

from 86.49 to 3.63 in KWMSS1 and from 31.47 to 9.97 in KWMSS2 as the DEM size decreased from 1 km to 250 m.

The results given from the structural stability assessment demonstrate that the parsimonious models used in this study, such as SFM and KWMSS1 with coarse DEM resolutions, were structurally unstable in terms of model output consistency with the objective functions; in turn, the model parameter set would have to be changed according to the modeling purpose. On the other hand, KWMSS2 led to comparatively stable model performances for the two objective functions. Moreover, in distributed rainfall-runoff modeling, the extent of the spatial aggregation of topography could be a dominant factor determining model structural stability as well as the model conceptualization in representing rainfall-runoff processes. Another interesting finding is that identical parameter sets were not obtained for both objective functions, even though KWMSS2 had the best structural stability. Instead, the optimal parameter combinations for SLS and HMLE resulted in indistinguishable hydrographs, as shown in Fig. 6 (c). This result implies that increased model complexity guarantees increased structural stability, while the identifiability of model parameters decreases; various parameter sets may lead to equally good simulation results. The subsequent section examines a method to assess parameter identifiability.

### 4.3 Model Evaluation with the Model Parameter Identifiability Index (MPII)

Parameter identifiability suggests the level of “uniqueness” of parameters. A well-identified model has a certain (or global optimal) parameter value, while a poorly identified model accepts many behavioral parameter values, which can provide model performance measures as good as the value given the best-performing parameter. Wagener et al. (2001, 2003a) proposed a simple measure of parameter identifiability based on a regional sensitivity analysis (RSA, Spear & Hornberger, 1980; Freer et al., 1996). Their uniform random sampling method can be used to extrapolate the parameter space easily, but it is computationally inefficient because a large number of samples are needed to protect against misleading results (Feyen et al., 2007). On the other hand, Markov chain Monte Carlo (MCMC) methods generate samples from Markov chains in an attempt to estimate the stationary posterior parameter distribution and can be useful for high-dimensional optimization



Table 2. MSSl values for each model structure: evaluation of the influence of the objective function on the model performance

Model	MSSl	Objective Function	Best-performing Parameter				
			$k$	$p$	$f$	$R_{SD}$	-
SFM	92.21	SLS	$k$	$p$	$f$	$R_{SD}$	-
			49.61	0.52	0.63	201.21	-
			49.64	0.64	0.57	118.94	-
KWMSS1 (250 m)	3.63	SLS	$N$	$k_d$	$d$	-	-
			0.50	0.010	0.389	-	-
			0.49	0.014	0.399	-	-
KWMSS1 (500 m)	49.46	SLS	0.50	0.016	0.313	-	-
		HMLE	0.50	0.013	0.282	-	-
KWMSS1 (1 km)	86.59	SLS	0.50	0.031	0.322	-	-
		HMLE	0.49	0.029	0.259	-	-
KWMSS2 (250 m)	9.97	SLS	$N$	$k_a$	$d_s$	$d_c$	$\beta$
			0.498	0.034	0.601	0.478	5.528
KWMSS2 (500 m)	12.97	SLS	0.322	0.05	0.626	0.481	5.361
		HMLE	0.453	0.05	0.755	0.600	7.602
KWMSS2 (1 km)	31.47	SLS	0.497	0.05	0.738	0.6	7.596
		HMLE	0.500	0.05	0.696	0.6	6.980

problems (Kuczera & Parent, 1998).

The shuffled complex evolution metropolis (SCEM-UA) algorithm is an effective and efficient evolutionary MCMC sampler that has an enhanced search capability and operates by merging the strengths of the metropolis algorithm, controlled random search, competitive evolution, and complex shuffling (Vrugt et al., 2003). This algorithm can reveal not only the optimal parameter set but also its underlying stationary posterior distribution within a single optimization run. In individual posterior parameter distributions, the parameter value corresponding to the highest density (*i.e.*, the values marked by  $x$  within the parameter spaces of each model in Fig. 1) represents the optimal parameter value, while other parameter values within these distributions (*i.e.*, the values inside the elliptical regions of each model in Fig. 1) are referred to as behavioral parameters. For parameter identifiability assessment, we applied the SCEM-UA to estimate individual posterior parameter distributions, and then investigated the uniqueness of the calibrated parameters from the probability density functions of each model. Here, the highest density values of each distribution were used as the individual indicators of parameter identifiability, and the mean value of each maximum identifiability indicator was used for the MPII.

First, we checked the convergence of the MCMC sampler using a special criterion, the scale-reduction score ( $\sqrt{SR}$ )

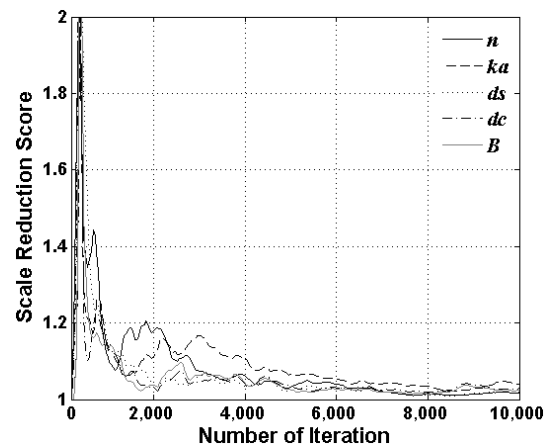


Fig. 7. Evaluation of the Gelman & Rubin (1992) scale reduction score for KWMSS2 based on the 500-m DEM

developed by Gelman & Rubin (1992). If  $\sqrt{SR}$  is less than 1.2, the Markov chain is considered to have converged into the target posterior distribution; otherwise, the evaluation steps are repeated until these sequences become stable. As shown in Fig. 7, the parameter sets corresponding to the first 4,000 simulations out of 10,000 iterations of the SCEM-UA are “non-behavioral” ones because the MCMC sampler had not yet converged into the stationary posterior distribution. Therefore, these sets were discarded from the estimation of the marginal posterior parameter distributions. For the remaining 6,000 parameter sets, the initial (or feasible) ranges of each parameter were split into 100 containers, and the samples

within each bin were counted to calculate the frequencies. The resulting frequencies were transformed into probability density functions; the best performing parameters were then assigned the highest value, with all measures summing up to 1. The highest density function value was used as the index for individual parameter identifiability, and the mean value of the highest densities of the marginal posterior distributions was used as the MPII. Fig. 8 shows the estimated marginal posterior parameter distributions of parameter  $p$  of SFM and  $\beta$  of KWMSS2 with the three DEMs, and Table 3 presents the estimated MPII values for each model structure.

The SFM led to the lowest MPII value even though it has

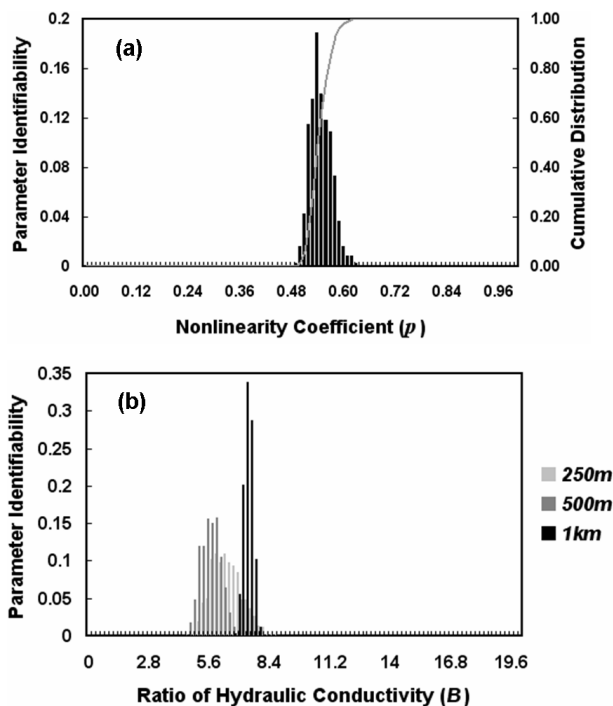


Fig. 8. Estimation of parameter identifiability for (a)  $p$  of SFM and (b)  $\beta$  of KWMSS2 based on the 250-m, 500-m, and 1-km DEMs

the most conceptualized model structure without consideration of the catchment topography. Moreover, KWMSS1 provided a generally higher MPII value than KWMSS2 in distributed rainfall-runoff modeling. It can be interpreted that the parameter interaction of the two additional parameters of KWMSS2,  $d_c$  and  $\beta$ , with the other parameters resulted in poor identifiability. In other words, adding model parameters increases the degree of freedom and eventually decreases parameter identifiability. However, the parameter number does not solely affect the parameter identifiability in distributed rainfall-runoff modeling using KWMSS2. The topographic representation is also one of the fundamental factors determining model parameter identifiability; for example, the MPII values of models with 250-m and 500-m topographic resolution were much lower than the value for the coarsest resolution application in KWMSS2.

In addition to parameter identifiability quantification, the uncertainty associated with behavioral parameter sets from the estimated posterior distributions was visualized by making probabilistic predictions. Probabilistic predictions were implemented using 5400 parameter combinations at the 90% confidence level for poorly identified models (SFM and KWMSS2 based on the 500-m DEM) and a well-identified model (KWMSS1 based on the 500-m DEM). Fig. 9 illustrates how the parameter uncertainty propagates into estimates of hydrograph simulation uncertainty. In this figure, the black circles indicate the observed streamflow data, and the gray shaded region shows hydrograph simulation uncertainty, which is associated with the posterior distribution of the parameter estimates. The uncertainty boundaries estimated by the posterior distributions of SFM and KWMSS1 fail to bracket the observations over the entire period, particularly for the peak flow part of SFM and the low flow part of KWMSS1. Thus, improvements in the model structures

Table 3. MPII values for each model structure: average maximum values of the marginal posterior parameter distributions of each parameter

Model No. / parameter	SFM		KWMSS1			KWMSS2				
			250 m	500 m	1 km	250 m	500 m	1 km		
1	$k$	0.076	$n$	0.205	0.45	0.251	$n$	0.122	0.059	0.24
2	$p$	0.188	$k_d$	0.679	0.36	0.155	$k_a$	0.343	0.115	0.883
3	$f$	0.068	$d$	0.532	0.804	0.8	$d_s$	0.094	0.121	0.359
4	$R_{St}$	0.078	-	-	-	-	$d_c$	0.085	0.107	0.324
5	-	-	-	-	-	-	$\beta$	0.11	0.157	0.339
MPII		0.103		0.472	0.538	0.402		0.151	0.112	0.429
Ave. MPII		0.103		0.471				0.231		

and calibration data may result in more accurate predictions. Moreover, the parameter uncertainty of SFM resulted in a very wide uncertainty boundary, while the behavioral parameter sets of KWMSS2 provided a narrow range of hydrographs. This result supports the idea that complex hydrological modeling is likely to encounter the equifinality problem, making it difficult to discriminate quantitatively and qualitatively between reliable and unreliable parameter sets.

The overall results of the model evaluation demonstrate that the ideal model structure, which guarantees the best values in terms of the three criteria, was not found in this study. The distributed model, KWMSS2, was much better than the simple models, SFM and KWMSS1, in terms of two evaluative

criteria, MPI and MSS1, but KWMSS2 did not ensure the best parameter identifiability. Therefore, additional constraints that are able to reject unreliable parameter set(s) and provide reliable prediction results need to be combined in the proposed modeling framework for further model identification.

## 5. Concluding Remarks

The traditional method of model evaluation usually relies only on model performance by comparing simulated variables to corresponding observations based on different periods and catchments. However, this classic type of model evaluation has been criticized because of its insufficient consideration of the various uncertainty sources involved in modeling processes. Despite the significant effects of such uncertainties on prediction results, the current modeling framework failed to fully incorporate the uncertainty components into the model evaluation. This paper discussed the use of an extended model evaluation method under uncertainty with the aim of identifying a reliable model structure. A set of rainfall-runoff models was developed and then evaluated three dimensionally with respect to model performance, model structural stability, and parameter identifiability. From the specific types of evaluation criteria used in this study, we note the following findings.

- (1) The results of the model performance evaluation indicated that all the models provided acceptable overall numerical outcomes as well as simulated hydrographs. However, the models showed different performances with regard to local behaviors, such as the rise and fall fractions of the hydrographs. KWMSS2 gave a balanced fit for all observed streamflow data regardless of the DEM size, while SFM and KWMSS1 provided marginally poor model performances during low flow periods. Moreover, despite the structural and systemic complexity of KWMSS1, it did not provide significantly improved model performance compared to SFM. Despite its simplicity, the very simple lumped conceptual model with sufficient available data and well-identified parameters can be expected to successfully represent the rainfall-runoff process in the mesoscale mountainous catchment (211 km<sup>2</sup>) studied here.
- (2) The discrepancy between hydrographs simulated for the two objective functions, SLS and HMLE, was used to assess

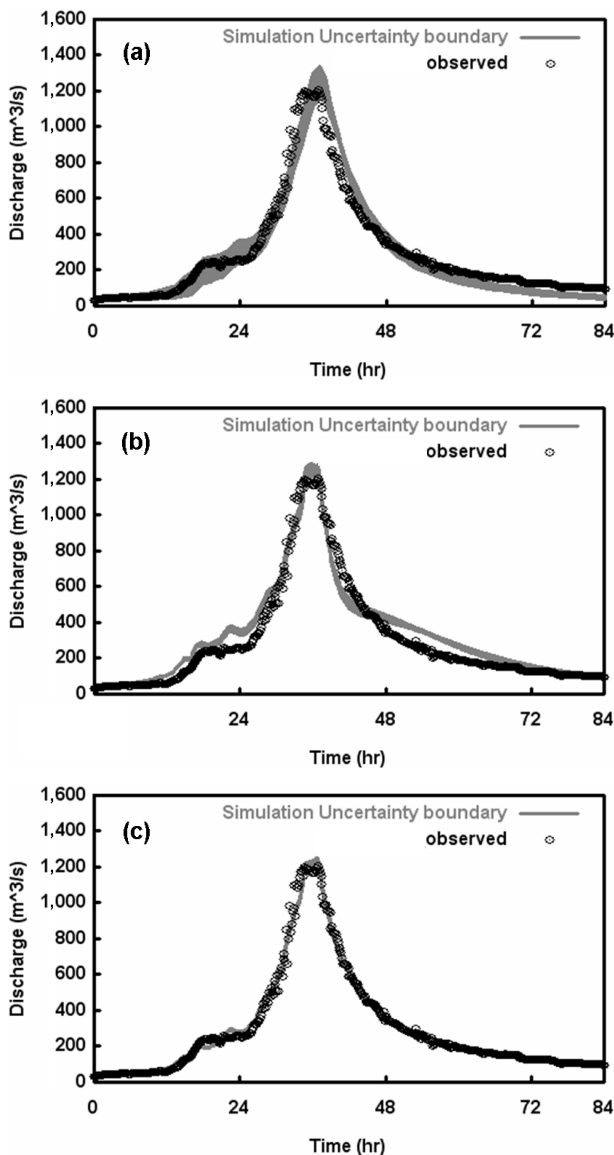


Fig. 9. Simulation uncertainty associated with the behavioral parameter sets for (a) SFM, (b) KWMSS1 based on the 500-m DEM, and (c) KWMSS2 based on the 500-m DEM

the model structural stability. There was no significant difference between the two hydrographs of KWMSS2; however, the simpler conceptual models, SFM and KWMSS1, showed poor structural stability. Moreover, in distributed rainfall-runoff modeling with coarser DEM sizes, model performances greatly depended on the objective functions. An interesting finding was that despite having identical hydrographs and constant numerical statistics, an identical parameter set satisfying both objective functions was not given for KWMSS2.

- (3) A new index, MPII, was proposed for evaluating parameter identifiability. The posterior parameter distributions were estimated by the SCEM-UA method. The highest values of these probability density functions were used to indicate parameter identifiability. In distributed rainfall-runoff modeling, KWMSS1 generally provided higher MPII values than KWMSS2. This finding may be interpreted to mean that adding model parameters increases the degree of freedom of a parameter and eventually decreases parameter identifiability. However, the simplest model, SFM, had the worst parameter identifiability. Therefore, it is also noteworthy that an increased parameter number influences parameter identifiability either solely or in combination with other factors such as the model structure itself and the spatial scale of the model-building unit due to different DEM sizes.

## Acknowledgment

This subject is supported by Korea Ministry of Environment (MOE) as “GAIA Program-2014000540005”.

## References

1. Beven, K. (2001), Revisiting the problem of model choice in Rainfall-runoff modeling: The primer, Chichester, John Wiley and Sons, pp. 297~306
2. Beven, K. (2002), Towards an alternative blueprint for a physically-based digitally simulated hydrologic response modeling system, *Hydrological Processes*, Vol. 16, No. 2, 189~206.
3. Beven, K. (2006), A manifesto for the equifinality thesis, *Journal of Hydrology*, Vol. 320, pp. 18~36.
4. Beven, K. and Binley, A. M. (1992), The future of distributed models: model calibration and uncertainty prediction, *Hydrological Processes*, Vol. 6, No. 3, pp. 279~298.
5. Beven, K. and Freer, J. (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems, *Journal of Hydrology*, Vol. 249, pp. 11~29.
6. Boyle, D. P., Gupta, H. V. and Sorooshian, S. (2000), Toward improved calibration of hydrological models: Combining the strengths of manual and automatic methods, *Water Resources Research*, Vol. 36, No. 12, pp. 3663~3674.
7. Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z. and Smith, M. (2001), Toward improved streamflow forecast: The value of semi-distributed modeling, *Water Resources Research*, Vol. 37, No. 11, pp. 2739~2759.
8. Duan, Q., Gupta, V. K. and Sorooshian, S. (1993), Shuffled complex evolution approach for effective and efficient global minimization, *Journal of Optimization Theory and Application*, Vol. 76, No. 3, pp. 501~521.
9. Duan, Q., Sorooshian, S. and Gupta, V. K. (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, Vol. 28, No. 4, pp. 1015~1031.
10. Duan, Q., Sorooshian, S. and Gupta, V. K. (1994), Optimal use of the SCE-UA global optimization method for calibrating watershed models, *Journal of Hydrology*, Vol. 158, pp. 265~284.
11. Feyen, L., Vrugt, J. A., Nualláin, B. Ó., Knijff, J. and Roo, A. D. (2007), Parameter optimization and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model, *Journal of Hydrology*, Vol. 332, pp. 276~289.
12. Freer, J., Beven, K. and Ambrose, B. (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of GLUE approach, *Water Resources Research*, Vol. 32, No. 7, pp. 2161~2173.
13. Gan, T. Y., Dlamini, E. M. and Biftu, G. F. (1997), Effects of model complexity and structure data quality, and objective functions on hydrologic modeling, *Journal of Hydrology*, Vol. 192, pp. 81~103.
14. Gelman, A. and Rubin, D. B. (1992), Inference from iterative simulation using multiple sequences, *Statistical Science*, Vol. 7, No. 4, pp. 457~472.
15. Gupta, H. V., Sorooshian, S., Hogue, T. S. and Boyle, D. P. (2003), Advances in automatic calibration of watershed models, in *Advances in calibration of watershed models*, edited by Q. Duan, S. Sorooshian, H.V. Gupta, Rosseau, A. and R. Turcotte, Washington, AGU, pp. 9~28.
16. Gupta, H. V., Sorooshian, S. and Yapo, P. O. (1998), Toward improved calibration of hydrologic models: Multiple and non-commensurable measures of information, *Water Resources Research*, Vol. 134, No. 4, pp. 751~763.
17. Ichikawa, Y., Shiiba, M., Tachikawa, Y. and Takara, K. (2000), Object-oriented hydrological modeling system, *Proceeding of 4th International Conference Hydro- informatics 2000 CD-ROM*, Iowa, pp. 8~15.
18. Jakeman, A. J. and Hornberger, G. M. (1993), How much complexity is warranted in a rainfall-runoff model?, *Water Resources Research*, Vol. 29, No. 8, pp. 2637~2649.
19. Kuczera, G. and Parent, E. (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *Journal of Hydrology*, Vol. 211, pp. 69~85.
20. Refsgaard, J. C. and Knudsen J. (1996), Operational validation and intercomparison of different types of hydrologic models,

- Water Resources Research, Vol. 32, No. 7, pp. 2189~2202.
21. Sorooshian, S. and Dracup, J. A. (1980), Stochastic parameter estimation procedures for hydrologic rainfall-runoff models: Correlated and heteroscedastic error cases, Water Resources Research, Vol. 16, No. 2, pp. 430~442.
  22. Spear, R. C. and Hornberger, G. M. (1980), Eutrophication in Peel Inlet - Identification of critical uncertainties via generalized sensitivity analysis, Water Research, Vol. 14, pp. 43~49.
  23. Tachikawa, Y., Nagatani, G. and Takara, K. (2004), Development of stage-discharge relationship equation incorporating saturated-unsaturated flow mechanism, Annual Journal of Hydraulic Engineering, Vol. 48, pp. 7~12 (in Japanese).
  24. Takasao, T. and Shiiba, M. (1988), Incorporation of the effect of concentration of flow into the kinematic wave equations and its applications to runoff system lumping, Journal of Hydrology, Vol. 102, pp. 301~322.
  25. Takasao, T., Shiiba, M. and Ichikawa, Y. (1996), A runoff simulation with structural hydrological modeling system, Journal of Hydroscience and Hydraulic Engineering, Vol. 14, No. 2, pp. 47~55.
  26. Uhlenbrook, S., Seibert, J., Leibundgut, C. and Rohde, A. (1999), Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structures, Hydrological Sciences Bulletin, Vol. 44, No. 5, pp. 779~797.
  27. Vrugt, J. A., Gupta, H. V., Bouten, W. and Sorooshian, S. (2003), A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resources Research, Vol. 39, No. 8, pp. 1~14.
  28. Wagener, T. (2003), Evaluation of catchment models, Hydrological Processes, Vol. 17, pp. 3375~3378.
  29. Wagener, T., Boyle, D. P., Leeds, M. J., Wheater, H. S., Gupta, H. V. and Sorooshian, S. (2001), A framework for development and application of hydrological models, Hydrology and Earth System Sciences, Vol. 5, pp. 13~26.
  30. Wagener, T. and Gupta H. V. (2005), Model identification for hydrological forecasting under uncertainty, Stochastic Environmental Research and Risk Assessment, Vol. 19, pp. 378~387.
  31. Wagener, T., Wheater, H. S. and Gupta, H. V. (2003a), Identification of Watershed Models in Advances in calibration of watershed models, edited by Q. Duan, S. Sorooshian, H.V. Gupta, Rosseau, A. and R. Turcotte, Washington, AGU, pp. 29~47.
  32. Wagener, T., Leeds, M. J., Wheater, H. S. and Gupta, H. V. (2003b), Towards reduced uncertainty in conceptual rainfall-runoff modeling: dynamic identifiability analysis. Hydrological Processes, Vol. 17, No. 2, pp. 455~476.
  33. Wagener, T., Wheater, H. S. and Gupta, H. V. (2004), Rainfall-runoff modeling in gauged and ungauged catchments, London, Imperial College Press, 306 pp.
  34. Yapo, P. O., Gupta, H. V. and Sorooshian, S. (1998), Multi-objective global optimization for hydrologic models, Journal of Hydrology, Vol. 204, pp. 83~97.
  35. Yu, P. S., Yang, T. C. and Chen, S. J. (2001), Comparison of uncertainty analysis methods for a distributed rainfall-runoff model, Journal of Hydrology, Vol. 244, pp. 43~59.