HYDROLOGICAL PROCESSES
Hydrol. Process. **21**, 1179–1195 (2007)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/hyp.6678

Hydrological model performance comparison through uncertainty recognition and quantification

Shen Chiang, Yasuto Tachikawa* and Kaoru Takara
Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan

Abstract:

In this study, a methodology which is capable of identifying the uncertainty of prediction through recognizing and quantifying the different sources of uncertainty in hydrologic models is applied for model comparison. The methodology is developed to recognize and quantify different uncertainty sources through observing hydrologic model behaviour under increasing input uncertainty levels. Based on the methodology, an index, which originates from the Nash–Sutcliffe efficiency named Model Structure Indicating Index (MSII) developed by the authors is applied to evaluate the reliability of model structure. A ranking of the adequacy of the hydrologic models to the watershed can be achieved by applying MSII. The hydrologic models Storage Function Method (SFM), TOPMODEL and KW-GIUH are used for model quantitative comparison in this study. Of these, a parameter-constrained SFM is used as an example of a poor-structured model; and two versions of TOPMODEL with different vertical flux calculation processes are used to demonstrate the behaviour of different model components. The results show that, at small input uncertainties, no distinction can be made between the capabilities of the hydrologic models to adapt themselves to error-contaminated data. As the input uncertainty increased, however, the distinction between the models becomes larger and the accuracy of the model structures could be quantified through MSII. The results prove that the index can be used as a tool for implementing quantitative model comparison/selection. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS model quantitative comparison; prediction uncertainty; predictions in ungauged basins (PUB)

Received 29 July 2005; Accepted 20 June 2006

INTRODUCTION

A hydrologic model is an integration of mathematical descriptions of conceptualized hydrologic processes, which serves a specific purpose in water resources engineering. Consequently, the spatial scale, temporal scale, model structure, architecture, and applicability of a hydrologic model are restricted by the hypothesis of the hydrologic model in most of the cases. As a result, numerous hydrologic models have been developed to suit various requirements, and the development of new hydrologic models or the improvement of existent models continues all over the world. Along with this scenario, the credibility of model outputs is becoming an important issue while applying hydrologic models not only for flood or drought simulations but also for other purposes such as decision making in relation to hydrological designs.

With rapid advances in computing technology, remote sensing, GIS and DBMS, the role of hydrologic models is enhanced as a tool in planning and decision making, and tends to be incorporated with other process models such as economic, social, political, administrative, and judicial. Thus, watershed hydrologic models will become a component of a larger management strategy. Furthermore, these models will become more global, not only in the sense of spatial scale but also in the sense of hydrologic details (Singh and Woolhiser, 2002). Therefore, a methodology for model comparison and evaluation of the adequacy for adopting hydrologic models for a given purpose is required.

It is doubtless that the performance of a hydrologic model is highly dependent on the hypothesis of the model, the data for calibration, the data for simulation input and the model structure. Underestimating or misunderstanding of these factors and the relationships among them may tremendously mislead the interpretation of the results of the hydrologic models.

Hydrologists have been interested in the effects of these factors on the accuracy and reliability of the estimation of catchment hydrological variables such as peak flow and flood volume. Some focus on the spatial and temporal variation in input data (precipitation data or remote sensing data) and its influence on the runoff (e.g.: Storm *et al.*, 1988; Lamb *et al.*, 1998; Andreassian *et al.*, 2001); others focus on the uncertainty of model structure (e.g.: Singh and Woolhiser, 1976), the uncertainty of model parameter (e.g.: Spear *et al.*, 1994; Johnson, 1996; Eckhardt *et al.*, 2003), scale issues (e.g.: Wolock, 1995), uncertainty propagation (e.g.: Crosetto *et al.*, 2001), or model validation (e.g.: Klemes, 1986; Bathurst *et al.*, 2004).

Recent research relating to hydrologic model uncertainty mostly refers to the identification of parameter uncertainty (e.g.: Uhlenbrook *et al.*, 1999), or parameter calibration (e.g.: Ajami *et al.*, 2004; Eckhardt *et al.*, 2005) and their impact to simulation results (e.g.: Freer

* Correspondence to: Yasuto Tachikawa, Disaster Prevention Research Institute, Kyoto University, Gokasho, Uji, Kyoto, 611-0011, Japan.
E-mail: tachikawa@mbox.kudpc.kyoto-u.ac.jp

et al., 1996; Kuczera and Parent, 1998). Many methodologies have also been proposed to identify and analyse the uncertainty of hydrologic simulation, such as BATEA (Kuczera and Franks, 2002), DYNIA (Wagener *et al.*, 2003), GLUE (Freer *et al.*, 1996), NLFIT (Kuczera, 1983a,b; Kuczera and Parent, 1998; Kuczera, 2004) etc. Among those, the Generalized Likelihood Uncertainty Estimation (GLUE) methodology offers a path of identifying parameter uncertainty. Nevertheless, parameter equifinality became the conclusion of GLUE; uncertainty related to input data and other factors are excluded. These uncertainties could also be included in GLUE but this has not normally been done (Beven, 2001). The uncertainty related studies give hydrologists the knowledge about the way in which uncertainty factors impact on the hydrologic models through observing the way models respond to them. A methodology or strategy for hydrologic model comparisons can be viewed as an application and extension of such knowledge.

The World Meteorological Organization (WMO) sponsored three studies intercomparing watershed hydrologic models. The first study (WMO, 1975) dealt with conceptual models used in hydrologic forecasting. The second study (WMO, 1986) dealt with an intercomparison of models that simulate flow rates, including snowmelt. The third study (WMO, 1992) dealt with models for forecasting streamflow in real time (Singh and Woolhiser, 2002). Another joint effort of model intercomparison is the Distributed Model Intercomparison Project (DMIP), which was held by the Hydrology Laboratory (HL) of the US NOAA's (National Oceanic & Atmospheric Administration) National Weather Services (NWS). One of the major goals of DMIP was to understand the capabilities of existing distributed modelling methods and identify promising directions for future research and development (Reed *et al.*, 2004). For lumped hydrologic models, Perrin *et al.* (2001) made an extensive comparative performance assessment of the structures of 19 daily lumped models by modelling 429 catchments located in France, the United States, Australia, the Ivory Coast, and Brazil.

It should be noted that the majority of the studies mentioned above were based on model performance, in which the model response surface was generated by giving forcing input data (precipitation, temperature, etc.) and evaluated by comparing the output with the observed watershed response data with unknown precision. Subsequently, the credibility of the results of model comparison is questionable. Available literature shows that the sensitivity or performance of a particular model is related to the spatial or temporal variability of input data, and not the sensitivity or performance of the actual basin. As a result, model assessment is a tricky exercise and the conclusions from such experiments generally depend on the methodology of the comparisons and the characteristics of the test catchment (Perrin *et al.*, 2001).

Almost all hydrologic models are being used under an unknown magnitude of input uncertainty. There are numerous studies related to the uncertainty of hydrologic model and model comparison, a fact, that is, seldom discussed. It is certain that input errors/uncertainties propagate and persist in hydrologic models, and corrupt the parameter estimation processes. The capability of a hydrologic model to regenerate the watershed response series from input data containing errors of a known magnitude is highly related to the model structure. Thus, the evaluation of the model performance–not only the model simulation outcomes but also outcomes during calibration–to different input data errors leads to a better understanding of the efficiency of the hydrologic model's structure.

Based on this idea, a methodology was developed to recognize and quantify the predicting uncertainty from a given input uncertainty to perform a hydrologic model quantitative comparison (Chiang *et al.*, 2005). A ranking of the adequacy of hydrologic models can be achieved by observing model behaviours under increasing input uncertainty. In practice, firstly, the Monte-Carlo simulation method is applied to add a bias item to the model input data series (rainfall), then rainfall realizations, parameter space, and model outcomes (outflow discharge) under different bias levels are acquired. Secondly, by examining the relationship between model simulation outcomes, calibration outcomes and observed watershed response series (discharge), an uncertainty structure can be recognized. Finally, model structure uncertainties caused by input data uncertainty are recognized and quantified through an index originating from the Nash–Sutcliffe efficiency called Model Structure Indicating Index (MSII), which is used as an instrument for implementing model quantitative comparison/selection.

Three hydrologic model bases are used for model quantitative comparison in this study. These are Storage Function Method (SFM) (Kimura, 1961), TOPMODEL (Beven and Kirkby, 1979) and KW-GIUH (Lee and Yen, 1997). Within these, a parameter-constrained SFM is used as an example of a poorly structured model, and two versions of TOPMODEL with different vertical flux calculation processes are used to demonstrate the behaviour of a different model component. The results show that within a small magnitude of input uncertainty, there are no apparent distinctions between the capabilities of the different hydrologic models to adapt themselves to the error-contaminated data. With increasing input uncertainty, however, the difference becomes larger and can be quantified by the index MSII, with a larger value of MSII indicating a poorly structured model. Finally, the different behaviours of the different hydrologic models under differing levels of input uncertainty are discussed by means of the model structure and catchment characteristics.

UNCERTAINTY QUANTIFICATION IN HYDROLOGIC MODELS

The performance of hydrologic models is profoundly affected by sources of uncertainty. In brief, these sources in hydrologic modelling are the observed data, the data used for model calibration, and the model structure.

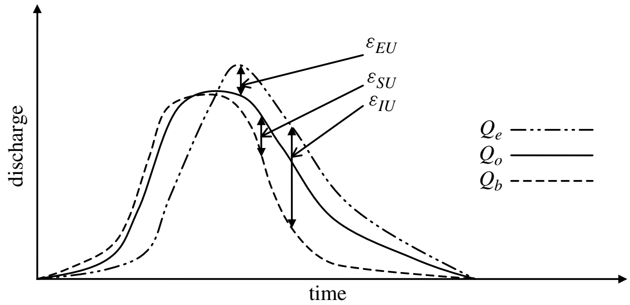

Figure 1. Schematic diagram of the uncertainty structure defined in the study. Q_o: observed discharge; Q_b: a model outcome with the best-fitted parameter set; Q_e: a model outcome with a parameter set within the whole parameter space. ε_{SU}, ε_{EU} and ε_{IU} are used to acquire system uncertainty, entire uncertainty and inherent uncertainty respectively

Data uncertainty is the most influential and contaminates other sources of uncertainty. Underestimating or misunderstanding these sources of uncertainty and the interrelation between them may mislead the interpretation of the results of the hydrologic models. In this section, prediction uncertainty sourced from the three kinds of sources mentioned previously is classified into four categories: system uncertainty, entire uncertainty, inherent uncertainty, and structure uncertainty. Figure 1 depicts a schematic diagram of the uncertainty structures. In Figure 1, ε_{SU}, ε_{EU} and ε_{IU} represent the differences in hydrographs among those observed, estimated with a parameter set in the parameter space, and estimated with the best-fitted parameter set. The system uncertainty, entire uncertainty and inherent uncertainty are evaluated using ε_{SU}, ε_{EU} and ε_{IU} in terms of Nash–Sutcliffe efficiency. The definition and the procedure of recognizing and quantifying the uncertainties are described in the subsequent text.

System uncertainty

A hydrologic model is an approximation of the real phenomena based on the hydrologic cycle, still it is vital to stress that there always exits a discrepancy between model outcome and observed data, no matter how precise the model is and how perfect the model is calibrated. This is due to the model's predicting limitation associated with the hypothesis and architecture of the model, termed the system uncertainty in this study.

The system uncertainty can be recognized by evaluating the discrepancies between the observed watershed response series and the model outcome during the process of model parameter calibration. The discrepancy is supposed to be the minor value by comparison to the model outcome by using different periods for the forcing input data. It has often been observed that the goodness-of-fit between observed data and estimated data during calibration is better than that in validation, not to mention in implementation. Given this fact, the uncertainty occurring here denotes the predicting limitation of the model, since this is the best performance that a model can achieve. In agreement with this definition, it is clear that the uncertainty comes from the process of calibration. The uncertainty source is the calibration data. The index for quantifying the system uncertainty is defined as:

$$SU = 1 - \frac{\sqrt{\sum_{i=1}^{n}(Q_{o_i} - Q_{b_i})^2}}{\sqrt{\sum_{i=1}^{n}(Q_{o_i} - \overline{Q}_o)^2}} \quad (1)$$

where Q_o and Q_b indicate the observed watershed response series and model outcome by using the best-fit parameter set, n is the time step of the time series, and $\overline{Q}_o$ denotes the time average of the whole observed watershed response series. SU elucidates the performance of the best-fitted simulated outcomes.

This is a measure of the model performance during calibration, denoting the predictive capability of a hydrologic model. It has been demonstrated that the model performance against independent data not used for calibration is generally poorer than the performance achieved in the calibration situation (Refsgaard and Henriksen, 2004). As a result, the system uncertainty is less than the entire uncertainty, which is described next.

Entire uncertainty

After calibrating the model parameter, the calibrated parameter space reflects its variance through the model structure and propagates through to the response surface. This uncertainty can be recognized by examining the discrepancy between the observed watershed response data and the model outcome by using input data and parameter sets in the calibrated parameter space. This is the method used by most of the current research dealing with parameter sensitivity. This uncertainty is defined as the entire uncertainty in this study, since, actually, this is the utmost uncertainty a model could have under existing input uncertainty. The index for quantifying the entire uncertainty is:

$$EU = 1 - \frac{\sqrt{\sum_{i=1}^{n}(Q_{o_i} - Q_{e_i})^2}}{\sqrt{\sum_{i=1}^{n}(Q_{o_i} - \overline{Q}_o)^2}} \quad (2)$$

where Q_e is the model outcomes acquired by using a parameter set within the whole parameter space and entire rainfall realizations.

Inherent uncertainty

The inherent uncertainty represents the variability of the parameter space, which is determined according to the input uncertainty. This can be examined by analysing the discrepancy between model outcomes using parameter values within parameter space and the outcome with the best-fit parameter set. Observed data is not used here, which indicates that the consistency of the model outcomes generated from the parameter set can

DOI: 10.1002/hyp

be observed. The index for quantifying the inherent uncertainty is defined as:

$$IU = 1 - \frac{\sqrt{\sum_{i=1}^{n}(Q_{b_i} - Q_{e_i})^2}}{\sqrt{\sum_{i=1}^{n}(Q_{b_i} - \overline{Q}_b)^2}} \quad (3)$$

where Q_e is calculated with a possible parameter set identified by other calibrated model outcomes, thus, the discrepancy between Q_b and Q_e indicates the error inevitable in a model formulation. The observed watershed response data is not used here. The outcomes during calibration are used as criteria to be compared with other outcomes, which are generated by other parameter sets using rainfall realization with the same noise variance as forcing input data.

Structure uncertainty–model structure indication index (MSII)

In order to implement a dynamic view of the hydrologic model's behaviour, the relationship among the system uncertainty, the entire uncertainty, and the inherent uncertainty caused by input data uncertainty is used to formulate a MSII (Chiang *et al.*, 2005) defined as:

$$MSII = \frac{IU - EU}{SU} \quad (4)$$

The difference between the entire and the inherent uncertainty is used as the numerator in the equation, while the system uncertainty is used as a denominator. The numerator is expected to be a smaller value when the model is more accurately reproducing the watershed response series. It is a measure of the possibility of a model adapting itself to the input uncertainty. The larger the magnitude the poorer the possibility of the model adapting itself to the error-contaminated input data. This indicates that the calibrated parameter space lacks the capability to drive the model to accurately reproduce the watershed response because of an insufficient structure of the hydrologic model.

The numerator of the MSII shows a measure of the adaptiveness of a hydrologic model to the error-contaminated input data, while the denominator of MSII indicates the effectiveness of the model calibration results and the predictive capability of the model. This enables the MSII to reflect how well the calibration scheme functions. The index interprets the variance caused by calibration process and model structure in a dimensionless form. During the implementation of model evaluation, a system uncertainty of less than zero is excluded from the calculation of MSII for its insignificance. Hence the smaller value of the MSII represents better model structure, and the range of MSII is: $0 \leq MSII < \infty$.

The logic behind uncertainty recognition and quantification

All hydrologic models are being manipulated under an unknown magnitude of input uncertainty. In this study, a good hydrologic model is assumed and expected to be capable of assimilating the input uncertainty and simulating the watershed behaviour to a certain level of precision. The methodology applied herein is designed to test the stability of the predictive capability possessed by a hydrologic model under a certain magnitude of input uncertainty. The idea can be clarified through considering a simplified modelling procedure, depicted in Figure 2, which has been modified according to Sargent (1999).

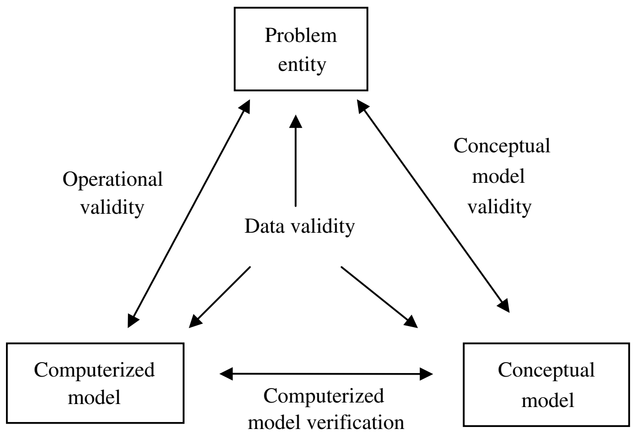

Figure 2. Modelling Process (quoted and modified from Sargent, 1999)

In Figure 2, the problem entity is the system to be modelled; the conceptual model is the mathematical representation of the problem entity developed for a particular study; and the computerized model is the conceptual model implemented on a computer. The conceptual model is developed through an analysis and modelling phase; the computerized model is developed through a computer programming and implementation phase, and inferences about the problem entity are obtained by conducting computer experiments on the computerized model in the experimentation phase. The conceptual model validity is defined as determining that the theories and assumptions underlying the conceptual model are correct and that the model representation of the problem entity is 'reasonable' for the intended purpose of the model. Computerized model verification is defined as ensuring that the computer programming and implementation of the conceptual model is correct. Operational validity is defined as determining that the model's output behaviour has sufficient accuracy for the model's intended purpose over the domain of the model's intended applicability. Data validity is defined as ensuring that the data necessary for model building, model evaluation and testing, and conducting the model experiments to solve the problem are adequate and correct (Sargent, 1999).

Input data uncertainty makes data validity difficult to perform. Since there is no way to assure the accuracy of the data used for model calibration, the subsequent result is that the best calibrated parameter set may or may not equal to the 'effective value' which will make a hydrologic model work properly. A feasible alternative is to know that within a certain level of input uncertainty,

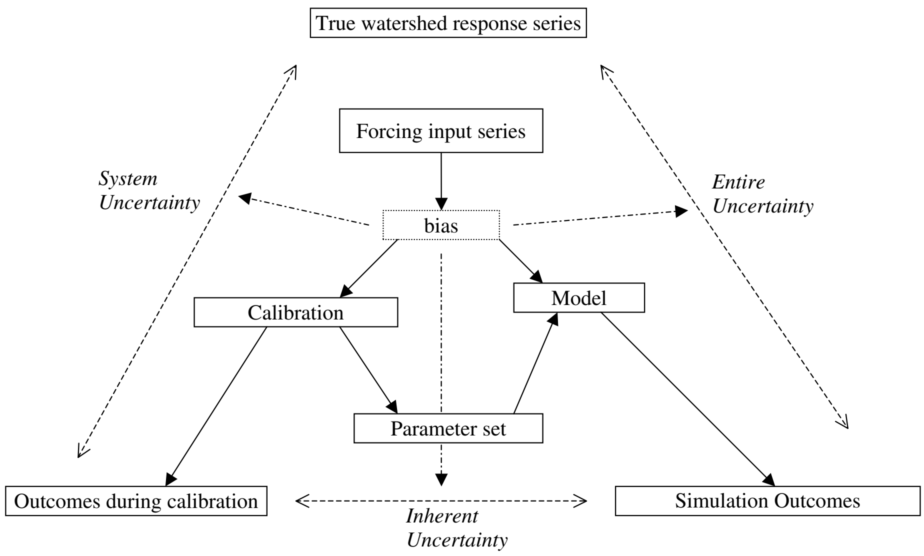

Figure 3. Schematic diagram of the uncertainty acquisition process

there is a possibility that a hydrologic model is still capable of regenerating true watershed characteristics. In this sense, both model calibration processes and the error propagation scheme induced by the model structure must be taken into consideration.

Figure 3 depicts the schematic diagram of the methodology proposed in this study. The bias item located at the centre of the structure dominates the whole uncertainty propagation scheme. If the focus is only on the effect of changing the bias item on the model outcome, referred to as entire uncertainty in this study, it can be seen as a sensitivity analysis of input data error; the system uncertainty indicates the predictive capability of the model under input uncertainty. The distance between system uncertainty and entire uncertainty represents the accuracy of the calibrated parameter sets, which can be referred to as a measure of model divergence, a term proposed by Sage and Melsa (1971).

Inherent uncertainty represents the variability of the parameter sets generated from specified input uncertainty levels. The distance between inherent uncertainty and entire uncertainty indicates the capability of a model to adapt itself to the specified input uncertainty, which is dominated by the model structure. The quantified and categorized uncertainty: system uncertainty, entire uncertainty and inherent uncertainty, are integrated into MSII, which enables the evaluation of how well the model structure behaves under a certain magnitude of input uncertainty.

ALGORITHM AND DATA APPLIED FOR UNCERTAINTY IDENTIFICATION

The methodology to generate the system uncertainty, entire uncertainty, inherent uncertainty and MSII is described in the subsequent text, and is detailed in Figure 4.

Instead of sampling the parameter space directly like GLUE, the study here generates the parameter set space by introducing a noise item into the input data with a probability distribution specified by using the Monte Carlo method.

1. For each different level of input uncertainty, a Monte Carlo simulation is applied to sample many rainfall realizations according to a real recorded event by adding a noise item to it. Here normal distribution with a mean of zero and a standard deviation equal to the input uncertainty (1·0–9·0 mm/h) is used to acquire model parameter spaces and outcomes under different levels of input uncertainty. Here, 100 rainfall realizations are generated for each level of input uncertainty.
2. Using the least square sum of errors, 100 parameter sets are determined for each level of input uncertainty. For each specified input uncertainty, 10 000 model outcomes were derived from the combination of 100 rainfall series realizations and 100 parameter sets through the model calibrations.
3. For each level of input uncertainty, the system uncertainty is acquired as a mean value by examining the discrepancy between the 100 best-fit model outcomes during calibrations and the observed watershed response.
4. For each level of input uncertainty, the entire uncertainty is acquired as a mean value by measuring the discrepancy between the observed watershed response and 10 000 model outcomes derived from the combination of applying 100 rainfall realizations and its corresponding 100 calibrated parameter sets as inputs.
5. For each level of input uncertainty, the inherent uncertainty is acquired as a mean value by examining the discrepancy between the best-fit model outcome during calibration and the rest of model outcomes. In this case, for one rainfall realization time series, the mean of the discrepancy between the best fit model outcome and 99 model outcomes using the parameter sets identified for 99 other rainfall realizations is

DOI: 10.1002/hyp

For input uncertainty σ_x = 1.0, 2.0,…, 5.0,…, 9.0 (unit: mm/hr)

1. Use Monte Carlo simulation to sample 100 rainfall realizations according to a real recorded event by adding noise item $\xi_{j,n}$ on it. Where $\xi \sim N(0, {\sigma_x}^2)$, *j* represents the *j*-th realization; *n* denotes the specific time step of the simulation time series.

2. Use LSE to determine 100 parameter sets $\theta_j, j = 1,...,100$, regarding to its corresponding rainfall realization.

$$S_{LS}(\theta_j) = \min\left(\sum_{n=1}^{N}(Y_n - f(x_n + \xi_{j,n}, \theta_j))^2\right)$$

where Y_n is the observed watershed response at specific time step *n*, *f* is the function of the model that used for calculating the model outcome, x_n is the observed watershed response at specific time step *n*.

3. System uncertainty SU_j is acquired using the Nash-Sutcliffe efficiency by examining the discrepancy $\varepsilon_{SU,j}$ between the best fit model outcome during calibration and the observed watershed response.

$$\varepsilon_{SU,j,n} = Y_n - f(x + \xi_{j,n}, \theta_j), j = 1,...,100$$

Mean value of SU_j for 100 cases are used for uncertainty quantification.

4. Entire uncertainty EU_{ij} is acquired using the Nash-Sutcliffe efficiency by examining the discrepancy $\varepsilon_{EU,i,j}$ between observed watershed response and the model outcome with *i*-th rainfall realization and the *j*-th parameter set.

$$\varepsilon_{EU,i,j,n} = Y_n - f(x + \xi_{i,n}, \theta_j)$$
$$i = 1,...,100; j = 1,...,100$$

Mean value of EU_{ij} for 10000 cases are used for uncertainty quantification.

5. Inherent uncertainty IU_{lk} is acquired using the Nash-Sutcliffe efficiency by examining the discrepancy $\varepsilon_{IU,l,k}$ between the best fit model outcome during calibration and the rest model outcomes with the *l*-th rainfall realization and the *k*-th parameter set.

$$\varepsilon_{IU,l,k,n} = y_{l,n} - f(x_n + \xi_{l,n}, \theta_k), l = 1,...,100; k = 1,...,100; l \neq k$$

where $y_{l,n} = f(x_n + \xi_{l,n}, \theta_l), l = 1,...,100$. Mean value of *IU* for 9900 cases are used for uncertainty quantification.

6. Apply Eq. (4) to acquire MSII of the Model.

End

Figure 4. Algorithm for uncertainty recognition and quantification

acquired; the calculations are made for 100 rainfall realizations; and finally, the inherent uncertainty is obtained as the total mean value.

A Japanese basin, the Yasu river basin, located in the Shiga prefecture, is used in this study. The main stream length of the Yasu river is about 65 km and the total basin area is around 387 km^2. It is a mountainous watershed with around 55% of the watershed having a slope greater than 0·1.

The methodology is applied to a selected rainfall event. The rainfall data was collected from four rainfall gauging stations inside the Yasu river basin as shown in Figure 5. The average precipitation was calculated according to the weight of each rainfall station through the Thiessen polygon method. The rainfall record of the event used here is applied for rainfall realization by adding a noise item, which is generated by using a normal distribution with a mean of zero and a specified standard deviation.

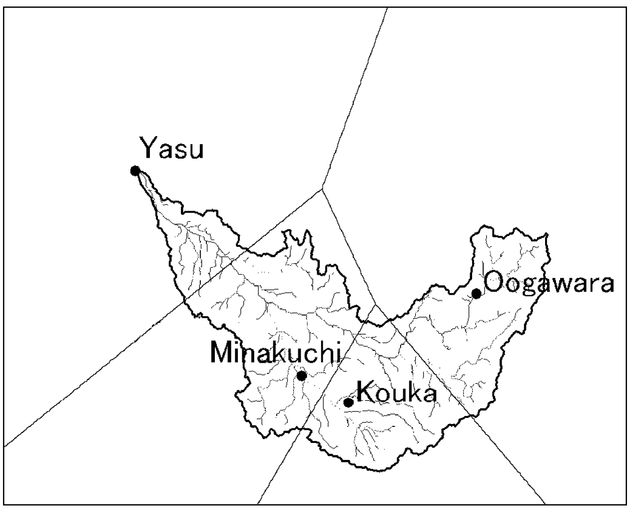

Figure 5. Yasu river basin with the main stream length of 65 km and the catchment area of 387 km^2

DOI: 10.1002/hyp

The negative rainfall was modified to zero in the rainfall realization process. A more realistic input-error model can be applied in this study but the error model with a normal distribution with a mean of zero and a specified standard deviation was chosen for its simplicity and for exaggeration purposes. The water level data acquired at the Yasu gauging station is in an hourly time step. A rating curve is used for transforming water level to discharge. During the observation period, there is no artificial operation in the upstream area.

DESCRIPTION OF THE MODELS USED FOR COMPARISON

Three hydrologic models are applied for model quantitative comparison. They are: SFM (Kimura, 1961), TOPDMODEL (Beven and Kirkby, 1979; Beven *et al*., 1984) and KW-GIUH (Lee and Yen, 1997). Table I summarizes the models used in the study. The brief descriptions are as follows:

Storage function model

The SFM was proposed by Kimura (1961). The form of SFM is as below:

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

$$r_e = \begin{cases} f \times r, & if \quad \sum r \leq R_{SA} \\ r, & if \quad \sum r > R_{SA} \end{cases} \tag{5}$$

where S is water storage height; r is rainfall intensity; q is runoff height; t is time step; and T_1 is the lag time. This model is often used for the flood runoff calculation in a basin with an area of less than five hundred square kilometers in Japan.

Parameter p is a constant, commonly the value is 0·6; f is the ratio of contribution area of the watershed, which generates outflow; and R_{SA} is the accumulated saturated rainfall. Parameter k is solved by the equation proposed by Nagai *et al*. (1982) which is:

$$k = 2{\cdot}5\left(\frac{n}{\sqrt{i}}\right)^{0{\cdot}6} A^{0{\cdot}24} \tag{6}$$

where n denotes effective roughness coefficient, i denotes average slope of the watershed, and A is catchment area. A fully functional SFM with parameter T_1, f, and R_{SA} is used.

Also, for contradistinction, a poorer-structured model with comparison to the original SFM by fixing the value of R_{SA} to 0·0 (which makes T_1 became the only functional parameter of the parameter-constrained SFM), is manipulated in this study.

TOPMODEL

TOPMODEL is almost 30 years old and has been the subject of numerous applications in a wide variety of catchments. The code used herein is a modified version based on the TOPMODEL 95·02 acquired from the official website of TOPMODEL (http://www.es.lancs.ac.uk/hfdg/freeware/hfdg_freeware_top.htm). TOPMODEL is a set of programs for rainfall-runoff modelling in single

Table I. Summary of the hydrologic models applied for model comparison

Hydrologic model	SFM	Parameter-constrained SFM	Five-parameter TOPMODEL	Six-parameter TOPMODEL	KW-GIUH
Lumped or distributed	Lumped model	Lumped model	Semi-distributed model	Semi-distributed model	Lumped model
Soil moisture storage	No	No	Yes	Yes	No
Channel routing	No	No	No	No	Yes
Ground water runoff	No	No	Yes	Yes	No
Surface runoff	Using time lag	Using time lag	Using time-area method	Using time-area method	Kinematic wave approximation
Infiltration	No	No	No	No	No
Linearity	Non-linear	Non-linear	Non-linear	Non-linear	Non-linear
Evapotranspiration	No	No	No	No	No
Surface runoff mechanism			Dunne-Black	Dunne-Black	Hortonian
Number of parameter	3: R_{sa}, f, t_l	1: t_l	5: *szm*, *sr0*, T_0, *RV*, q_o	6: *szm*, *sr0*, T_0, *RV*, q_o, t_d	2: roughness coefficient for overland flow *No* and roughness coefficient for channel flow *Nc*

or multiple subcatchments in a semi-distributed way and using raster elevation data for the catchment area. It is considered a physically based model as its parameters can be measured *in situ* (Beven and Kirkby, 1979; Beven, 1997). Subcatchment discharge is routed to the catchment outlet by using a time-area diagram with a constant velocity over the entire catchment area, a parameter that needs to be calibrated. Topographic index derivation was obtained by using DEM algorithm. The infiltration excess mechanism and the evapotransportation mechanism are not included in this study.

In TOPMODEL, there are several parameters that need to be calibrated before model validation and implementation. Basically they are: m, the decay factor which controls the rate of decline of transmissivity with increasing storage deficit; T_o, the hydraulic transmissivity; t_d: the time delay constant for vertical flux calculation; RV, the overland flow velocity; Q_o, the initial base flow; and S_{r0}, the initial root zone storage deficit, which are specified at the start of the simulation.

Inside TOPMODEL, the vertical drainage q_v from an unsaturated zone storage at any point of topographic index class i is calculated by Equation (7):

$$q_{v_i} = \frac{S_{uz}}{S_i t_d} \tag{7}$$

where S_{uz} is the storage in the unsaturated zone, t_{d} is a time delay constant and S_i is a local storage deficit; or by Equation (8) (Beven *et al.*, 1995):

$$q_{v_i} = \alpha K_0 e^{-S_i/m} \tag{8}$$

where α is the effective vertical hydraulic gradient, K_0 is the saturated conductivity at the surface, and m is a model parameter controlling the rate of decline of transmissivity with increasing storage deficit. If the value of α is set to unity, thus assuming that the vertical flux is equal to the saturated hydraulic conductivity just at the water table, it is eliminated as a parameter (Beven *et al.*, 1995).

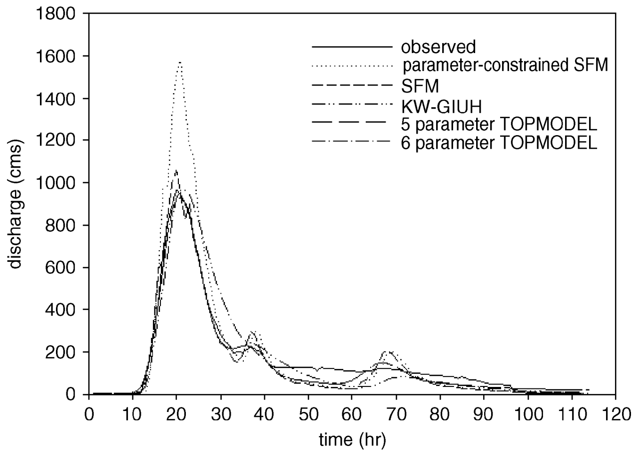

Figure 6. Best-fit hydrograph simulated by each model

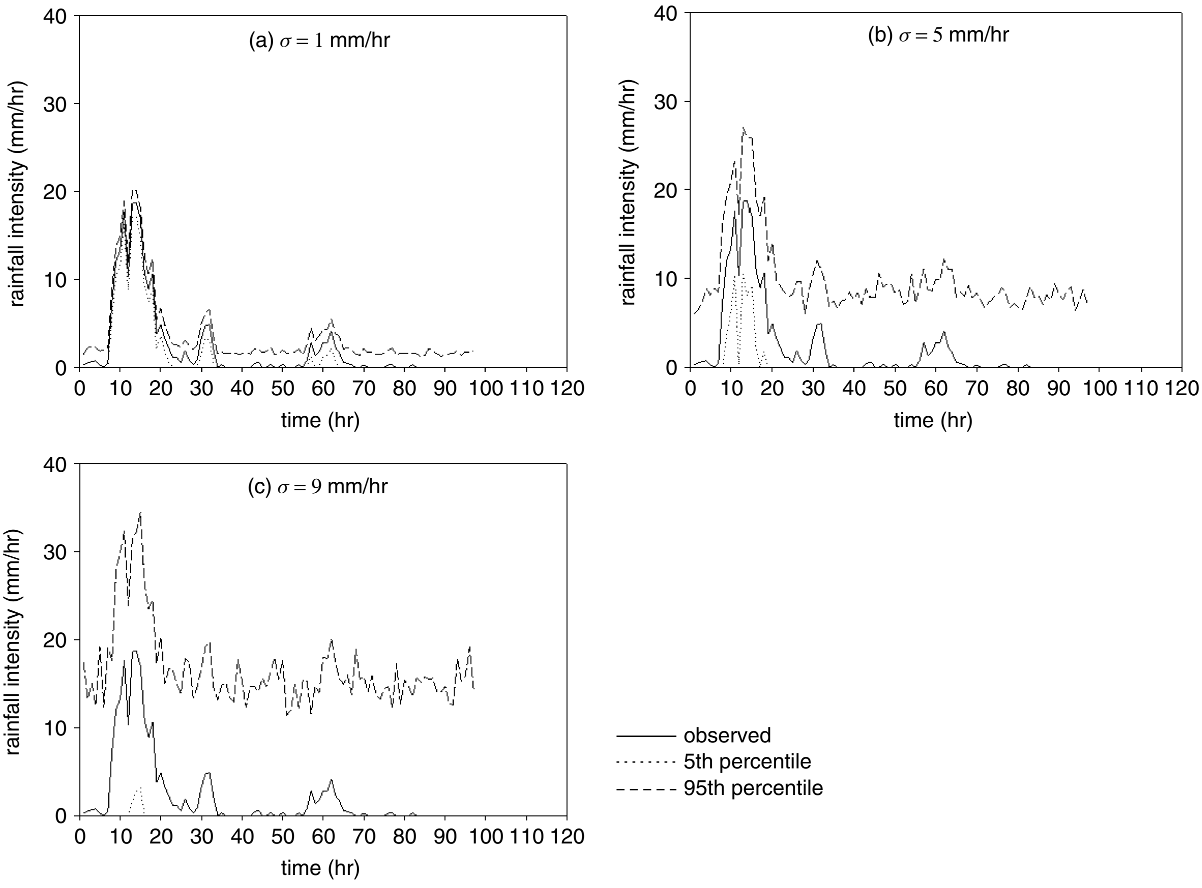

Figure 7. Rainfall realizations according to the recorded event. (a), (b) and (c) represent fifth and ninety-fifth percentile for each time step of the 100 rainfall realizations with the standard deviation equal to 1, 5 and 9 mm/h respectively

DOI: 10.1002/hyp

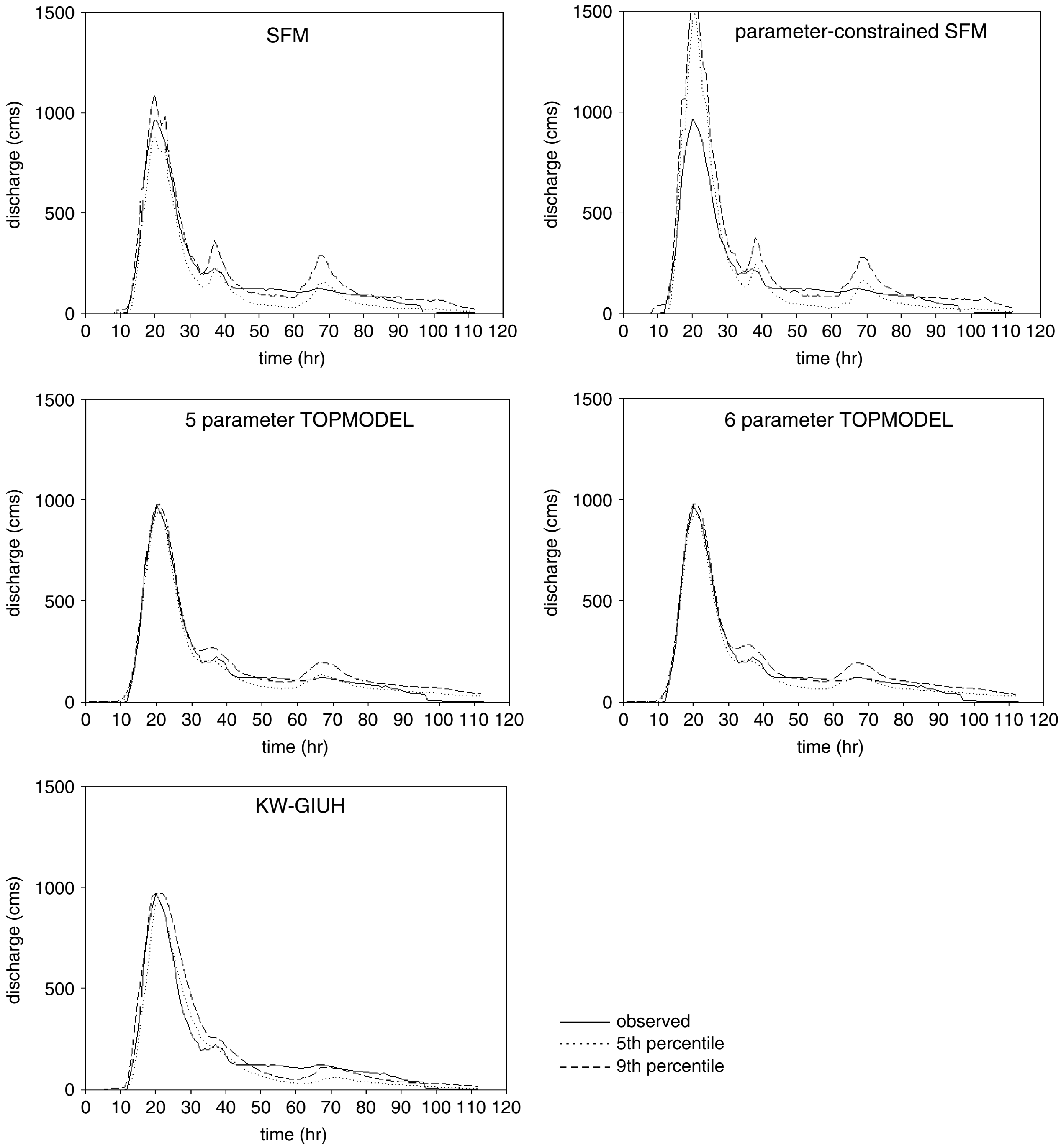

Figure 8. Outcomes during calibration of SFM, parameter-constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH with input uncertainty = 1 mm/h

Equation (7) is the equation of a linear store but with a time constant $S_i t_d$ that increases with increasing depth of the water table (Beven *et al*., 1995). In the model, the K_0 is acquired by T_o, that means the number of parameters are five (m, T_o, Q_o, RV and S_{r0}) or six (m, T_o, t_d, Q_o, RV and S_{r0}). Two TOPMODEL models with different vertical flux computing components are applied for model quantitative comparison.

KW-GIUH

Rodriguez-Iturbe and Valdes (1979) and Gupta *et al*. (1980) developed the IUH by using geomorphic stream order information. If one excess unit of rainfall falls on the watershed instantaneously, by assuming that the raindrops are independent and isolated from each other and ignoring the raindrops falling on the river, then the distribution of the number of raindrops appearing at the outlet to time is the instantaneous unit hydrograph of the watershed. The movement of the surface water, which came from raindrops inside the watershed, is described in terms of a probability distribution.

KW-GIUH (Lee and Yen, 1997) is a refinement of the GIUH method by using kinematic wave approximation to calculate the travel time of overland flow inside the

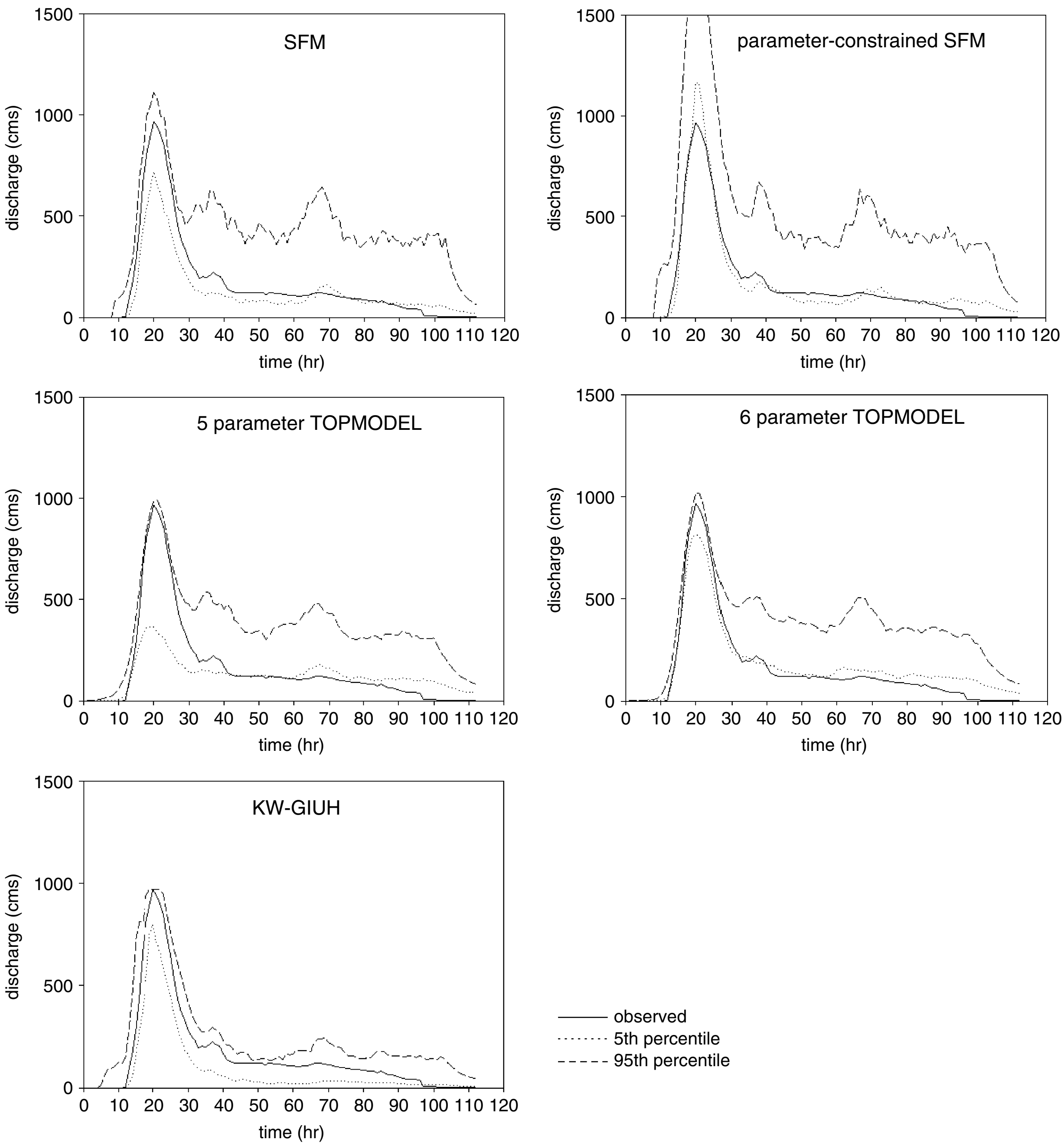

Figure 9. Outcomes during calibration of SFM, parameter-constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH with input uncertainty = 5 mm/h

catchment. In KW-GIUH, an *i*th-order subbasin of the watershed is conceptually simplified as consisting of two identical rectangular overland-flow planes. Each plane contributes a lateral discharge into a channel of constant cross section and slope. For implementation, the channel width on the outlet of the watershed is acquired by a field survey; the rest of the geomorphic factors can be acquired from a topographic map or through raster elevation data and DEMs by some algorithms.

RESULTS OF APPLICATIONS AND DISCUSSION

The methodology described above is applied here for model comparison. Figure 6 is the hydrograph of the calibration results of the five models. The performances of the models are evaluated by using the Nash–Sutcliffe efficiency (Nash and Sutcliffe, 1970). Table II is the summary of the results. The TOPMODEL is the best while the parameter constrained SFM is the worst.

The rainfall record of the event used here is applied for a generation of 100 rainfall realizations for every input uncertainty increment; a normal distribution with a mean of zero and a specified standard deviation as input uncertainty is manipulated for the random process. Figure 7 shows the examples of rainfall realizations by specifying the input uncertainty magnitude equal to 1·0, 5·0 and 9·0 mm/h. For each time step, the fifth and the ninety-fifth percentile of the realized rainfall intensity

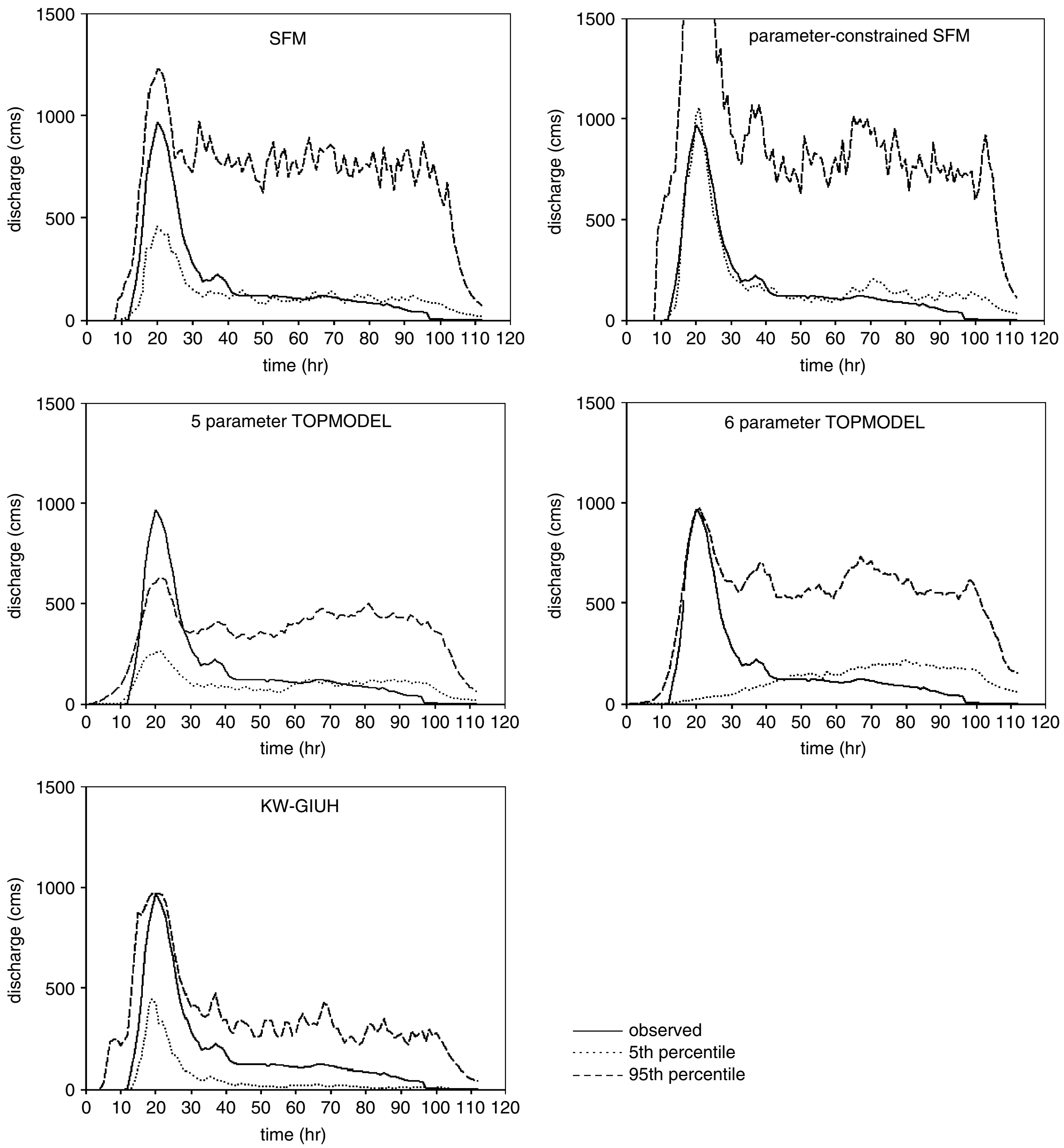

Figure 10. Outcomes during calibration of SFM, parameter-constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH with input uncertainty = 9 mm/h

Table II. Value of the Nash–Sutcliffe coefficient for each model with the best–fit model parameter set

Hydrologic Model	Parameter-constrained SFM	SFM	Five-parameter TOPMODEL	Six-parameter TOPMODEL	KW-GIUH
Nash–Sutcliffe coefficient	0·45	0·80	0·86	0·86	0·77

are drawn. It can be seen that the spectrum of the rainfall becomes broader while the input uncertainty is higher. The outcomes during calibration of each model are plotted in Figures 8 to 10, and the simulation results of each model are plotted in Figures 11 to 13. The fifth and ninety-fifth percentiles of the 100 hydrographs of each time step of the specified input uncertainty (1, 5 and 9 mm/h) are plotted. In Figures 8 and 11, except for the parameter-constrained SFM, the models perform without obvious distinction. While in Figures 9

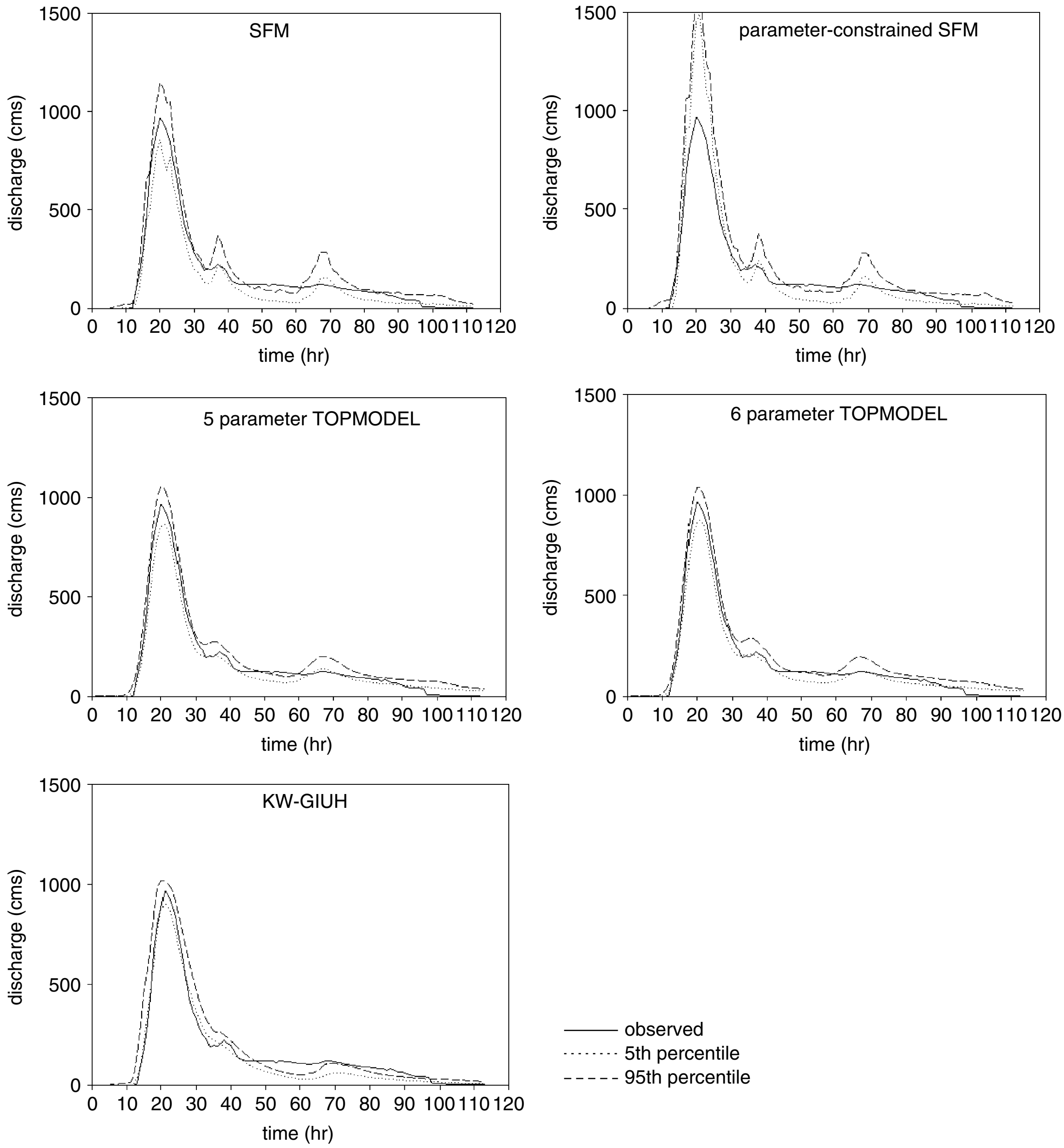

Figure 11. Simulation results of SFM, parameter-constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH with input uncertainty = 1 mm/h

and 12, the characteristics of the models were revealed by the spectrum of the hydrograph. Since the size and the shape of the spectrum of the model outcomes reflect the behaviour of a model structure response to the error-contaminated input data. It can be seen that both in calibration phase and simulation phase, the outcomes spectrum of the KW-GIUH envelopes the observed watershed response data the narrowest. While other models, especially the parameter-constrained SFM, envelop the observed response data poorly, indicating a smaller value of the system uncertainty as is revealed in Figure 14. Figures 10 and 13 can be seen as an extreme condition for a large magnitude of input uncertainty, it can be seen that only KW-GIUH still holds the capability to produce a simulated hydrograph, that is, similar to the observed watershed response.

Figure 14 is the entire, inherent and system uncertainty of each model. It can be seen that the entire uncertainty becomes larger as input uncertainty increases. It is always expected that entire uncertainty increases as input uncertainty increases; the discrepancy between the entire uncertainty and the inherent uncertainty shows the same tendency. The system uncertainty is located between entire and inherent uncertainty. The range between entire and inherent uncertainty indicates the capability of a model that adapts itself to the error-contaminated data. In other words, it is the capability of the model to assimilate the input error through the adjustment of the model parameter. The broader the distance, the poorer the prediction capability of the model, meaning that there is less possibility that the model is capable of producing an accurate watershed response series.

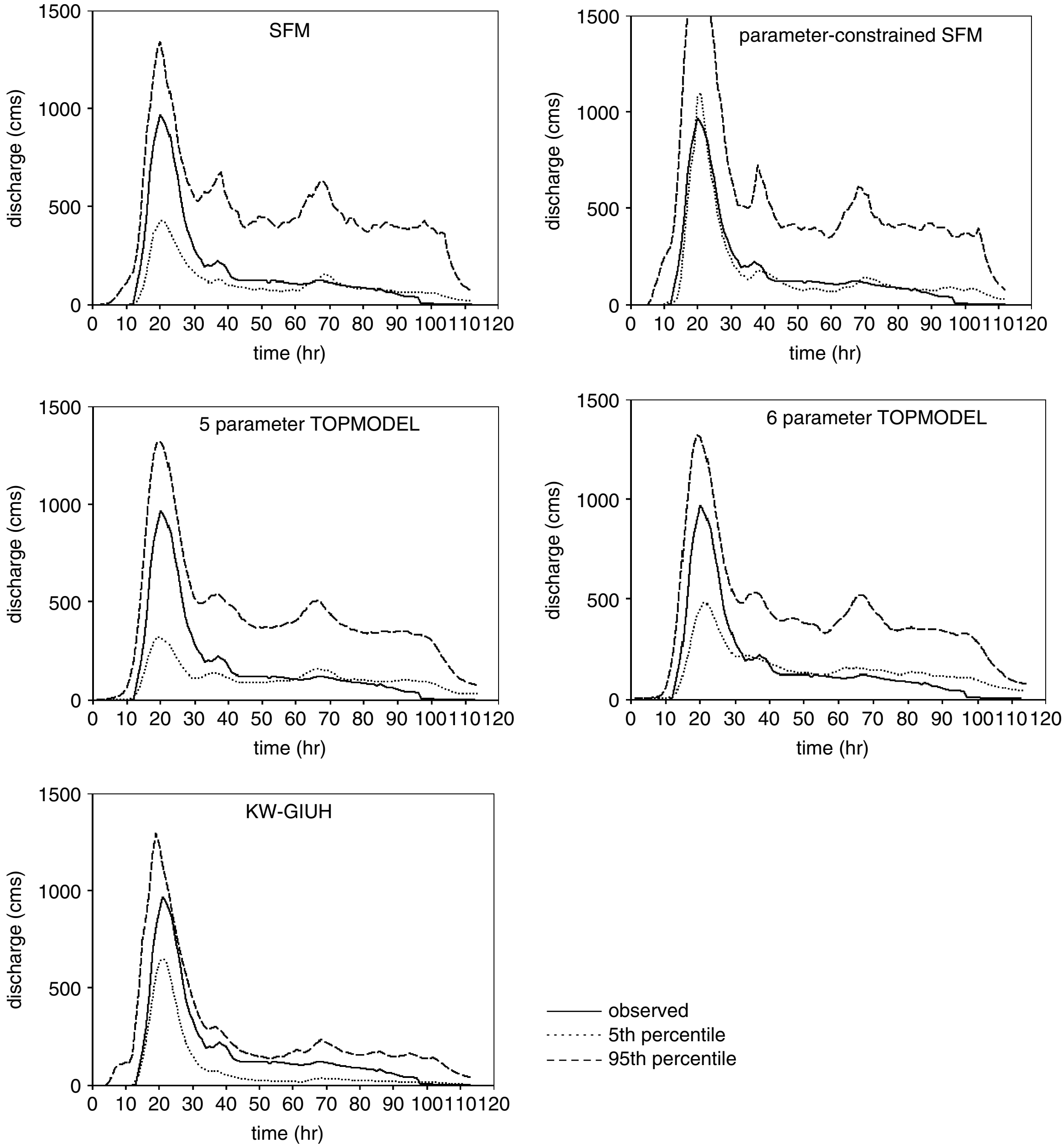

Figure 12. Simulation results of SFM, parameter-constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH with input uncertainty = 5 mm/h

The system uncertainty indicates the accuracy of the model calibration result. It is shown that within an input uncertainty of less than 2·0 mm/h, there is little distinction in the entire uncertainty and system uncertainty between the five- and the six-parameter TOPMODEL. However, the higher magnitude of the inherent uncertainty of the five-parameter TOPMODEL indicates the poorer capability of the model adapting itself to the error-contaminated data. Yet with an input uncertainty larger than 5 mm/h, the entire uncertainty and system uncertainty of the six-parameter TOPDMODEL becomes worse than the five-parameter TOPMODEL. The results from the higher inherent uncertainties indicate the same. This is evidence that the five-parameter TOPMODEL is more capable of reproducing the watershed response series than the six-parameter TOPMODEL when the magnitude of the input uncertainty is high. By composing the possibility (the distance between entire and inherent uncertainty) and the capability (the system uncertainty), the model structure of a model can be evaluated.

Figure 15 shows the MSII of each model. Except the parameter constrained SFM, there is no apparent distinction between the other four models with small input uncertainties. However, KW-GIUH is structurally more stable than SFM and TOPMODEL with increasing input uncertainty. This also proves that even TOPMODEL performs the best during the calibration process, though with increasing input uncertainties, the capability of TOPMODEL to adapt itself to the error-contaminated input data is still weak. Although KW-GIUH does not perform as well as TOPMODEL with small input uncertainties, with increasing input uncertainty, the capability of the

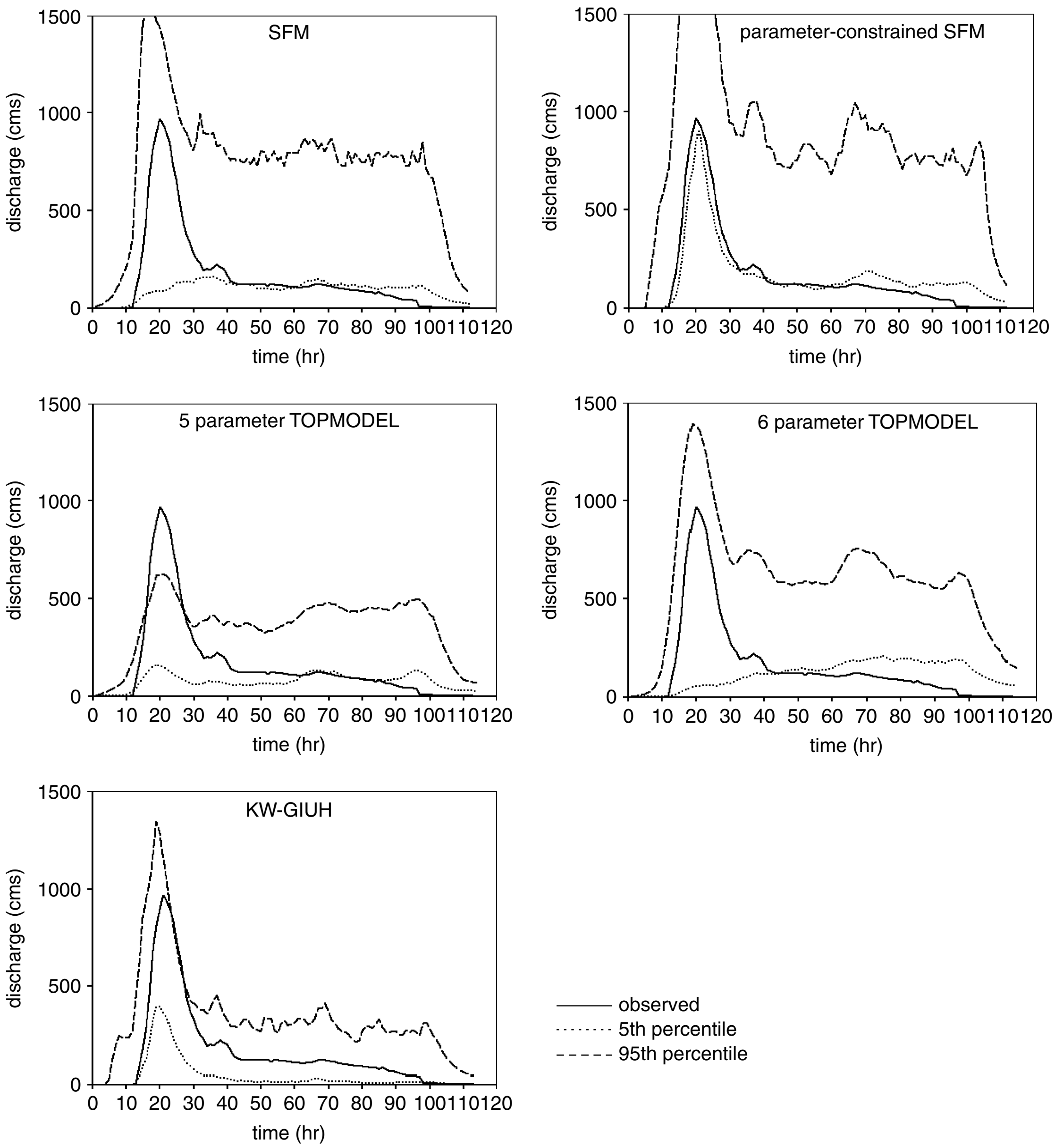

Figure 13. Simulation results of SFM, parameter-constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH with input uncertainty = 9 mm/h

model structure to adapt itself to the error-contaminated data is better than that of TOPMODEL. For SFM, with small magnitudes of input uncertainty, the capability is as good as TOPMODEL, however, the rapid ascending trend shows that it fails the test under high levels of input uncertainty.

The reason for the superior performance of the KW-GIUH under higher magnitudes of input uncertainty can be explained by the structure of MSII and Figures 10, 13 and 14. As mentioned previously, when the distance between the entire uncertainty and the inherent uncertainty is close, a small value of MSII is shown, and the model is considered to have a better chance of a better fitting simulation result. It is a measure of the capability of a model adapting itself to the input uncertainty. The higher the performance of the system uncertainty, the better the model structure is. It is shown in Figures 10 and 13 that the spectrum of the model outcomes during the calibration phase and simulation phase of KW-GIUH are narrower than the other models. This indicates a higher inherent uncertainty. Also, the average of the spectrum, which can be expressed by the system uncertainty and the entire uncertainty, which are shown in Figure 14, is better than the other models under high magnitude of the input uncertainty. The behaviour of the system uncertainty under increasing input uncertainty indicates the capability of the model to assimilate the input uncertainty by using the model parameter. This is evidence that a higher performance of the system uncertainty denotes a better model structure.

As a result, the higher performance of the system uncertainty makes KW-GIUH superior to the five-parameter TOPMODEL under a high magnitude of input uncertainty. In addition to the reasons mentioned above,

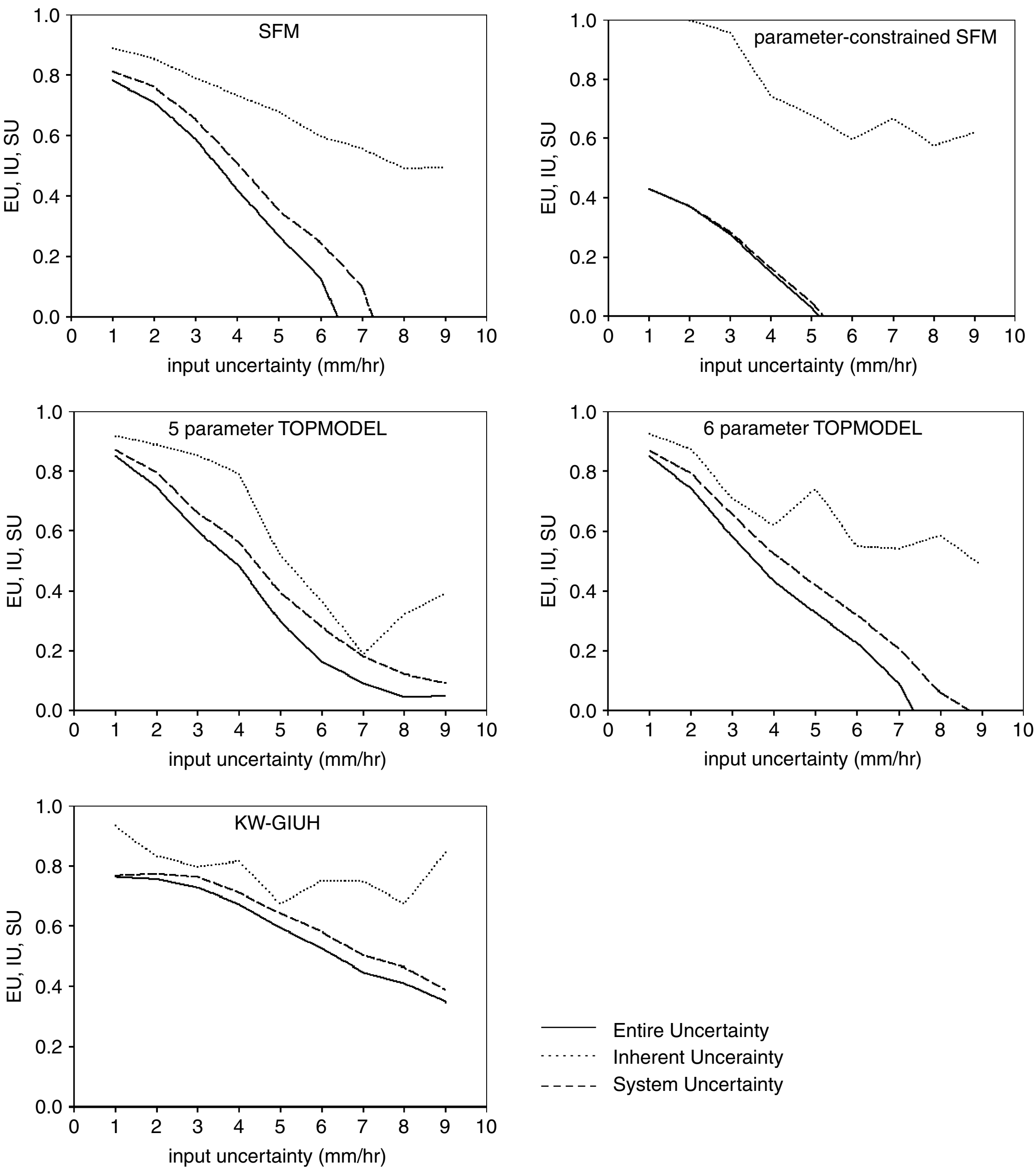

Figure 14. Entire, inherent and system uncertainty of SFM, constrained SFM, five-parameter TOPMODEL, six-parameter TOPMODEL and KW-GIUH

the watershed mountainous terrain characteristics and the selection of heavy rainfall event areas, are also considered as some of the reasons for the better performance of the KW-GIUH. As a consequent result, the capability of KW-GIUH to adapt itself to high input uncertainty is better than the other hydrologic models by the selected input-error model.

CONCLUSIONS

In this study, through uncertainty recognition and quantification, a methodology for hydrological model comparison is proposed. A uniform distribution with a specific standard deviation is applied to randomly generate the bias item, which is added to a true rainfall event for every time step to generate rainfall realizations. The parameter space and model outcomes (outflow discharge) under different input uncertainty levels are then acquired. Instead of sampling the parameter space directly as GLUE Methodology, the methodology generates the parameter set space by introducing noise to the input data with a specified probability distribution. This reflects the truth that parameter uncertainty comes from the uncertainty in the data and the way the model structure responds to it.

Finally, by examining the interrelationship between model simulation outcomes, model outcomes during calibration and observed watershed response series (discharge), different categorized uncertainties can be recognized and quantified by a predefined index with its corresponding input uncertainty level. An index which originates from the Nash–Sutcliffe efficiency named MSII is applied to quantify model structure uncertainty. The

DOI: 10.1002/hyp

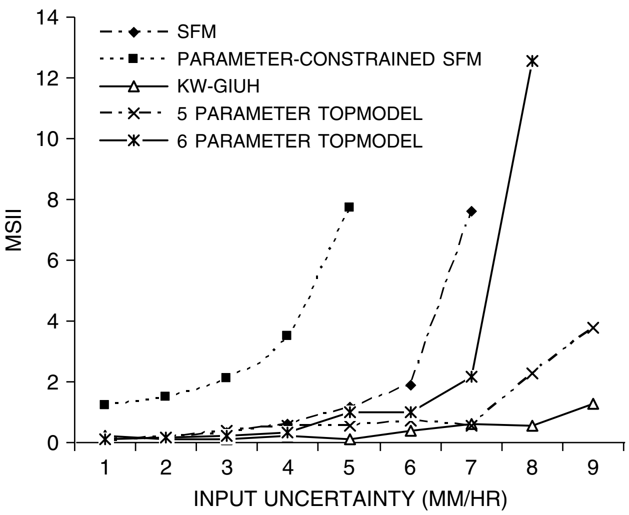

Figure 15. MSII of five-parameter TOPMODEL, six-parameter TOPMODEL, SFM, parameter constrained SFM and KW-GIUH

index can reveal the capability of a model to adapt itself to the error-contaminated input data under a specific level of uncertainty in a relative display frame. The study has shown that the index can be used as an indication for implementing model quantitative comparison/selection.

Five hydrologic models are used for model comparison in this study, which are: SFM, parameter constrained SFM, TOPMODEL with five parameters, TOPMODEL with six parameters and KW-GIUH. It is shown that with a small magnitude of input uncertainty, the model structure of parameter constrained SFM is the worst; TOPMODEL (both the five- and six-parameter versions), SFM, KW-GIUH perform similarly. With increasing input uncertainty KW-GIUH becomes the best, then TOPMODEL (five-parameter version), TOPMODEL (six-parameter version), SFM and then the parameter–constrained SFM.

The input-error model manipulated in this study was chosen for its simplicity and exaggeration purposes. The performance of KW-GIUH is superior under a high magnitude of the input uncertainty. The results show that the methodology is a good reference for a hydrological modeller or decision maker, it should be noted that the results are for certain input error assumptions and certain watershed topographic features.

The aim of this paper is to apply the methodology of predicting the uncertainty quantification caused by input data error and model calibration processes. The input-error model is not the main concern here. Theoretically, the methodology is capable of performing hydrologic model comparison with any input-error model. That is, the hydrologic model can be compared under different input data error assumptions. Compared to the parameter space sampling based approach, the methodology applied in this study treats the input uncertainty as an intrinsic component of the uncertainty identification process, and evaluates the model structure according to the ability of the model to adapt itself to the error input. The methodology would be enhanced if the uncertainty of the observed response data is involved in the procedure for uncertainty identification. Also, applying various input error models, especially a more realistic one to test the robustness of the methodology, will be considered in a further study. The calculation cost should also be taken into consideration since the computational burden will be several times larger than the present approach.

Only one rainfall event is applied in this study, further investigations are suggested to incorporate a result updating scheme into the methodology. A mathematical description of the methodology is needed so that an adequate updating procedure can be developed. The Bayesian approach seems to be one of the options to solve the problem.

REFERENCES

Ajami NK, Gupta H, Wagener T, Sorooshian S. 2004. Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system. *Journal of Hydrology* **298**: 112–135.

Andreassian V, Perrin C, Michel C, Usart-Sanchez I, Lavabre J. 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology* **250**: 206–223.

Bathurst JC, Ewen J, Parkin G, O'Connell PE, Cooper JD. 2004. Validation of catchment models for predicting land-use and climate change impacts. 3. Blind validation for internal and outlet responses. *Journal of Hydrology* **287**: 74–94.

Beven K. 1997. Topmodel: a Critique. *Hydrological Processes* **11**: 1069–1085.

Beven K. 2001. How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences* **5**: 1–12.

Beven K, Kirkby MJ. 1979. A Physically based variable contributing area model of catchement hydrology. *Hydrological Sciences Bulletin* **24**: 43–69.

Beven KJ, Kirkby MJ, Schofield N, Tagg AF. 1984. Testing a physically-based flood forecasting-model (Topmodel) for 3 Uk catchments. *Journal of Hydrology* **69**: 119–143.

Beven K, Lamb R, Quinn P, Romanowicz R, Freer J. 1995. TOPMODEL. In *Computer Models of Watershed Hydrology*, Singh VP (ed.). Water Resources Publications: Highlands Ranch, Colorado, USA, 627–668.

Chiang S, Tachikawa Y, Takara K. 2005. An approach to uncertainty identification in hydrological modeling. *Annual of Hydraulic Engineering, JSCE* **49**: 211–216.

Crosetto M, Ruiz JAM, Crippa B. 2001. Uncertainty propagation in models driven by remotely sensed data. *Remote Sensing of Environment* **76**: 373–385.

Eckhardt K, Breuer L, Frede HG. 2003. Parameter uncertainty and the significance of simulated land use change effects. *Journal of Hydrology* **273**: 164–176.

Eckhardt K, Fohrer N, Frede HG. 2005. Automatic model calibration. *Hydrological Processes* **19**: 651–658.

Freer J, Beven K, Ambroise B. 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* **32**: 2161–2173.

Gupta VK, Waymire E, Wang CT. 1980. A representation of an instantaneous unit-hydrograph from geomorphology. *Water Resources Research* **16**: 855–862.

Johnson PA. 1996. Uncertainty of hydraulic parameters. *Journal of Hydraulic Engineering-ASCE* **122**: 112–114.

Kimura T. 1961. *The Flood Runoff Analysis by Using the Storage Function Method (in Japanese)*. The Public Works Research Institute, Ministry of Construction: Japan.

Klemes V. 1986. Operational testing of hydrological simulation-models. *Hydrological Sciences Journal-Journal des Sciences Hydrologiques* **31**: 13–24.

Kuczera G. 1983a. Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty. *Water Resources Research* **19**: 1151–1162.

Kuczera G. 1983b. Improved parameter inference in catchment models. 2. Combining different kinds of hydrologic data and testing their compatibility. *Water Resources Research* **19**: 1163–1172.

DOI: 10.1002/hyp

Kuczera G. 2004. *NLFIT A Bayesian Nonlinear Regression Suite, Users Manual*. Department of civil Engineering and Surveying, University of Newcastle.

Kuczera G, Parent E. 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* **211**: 69–85.

Kuczera G, Franks SW. 2002. On the testing of hydrological models: falsification or fortification? In *Mathematical Models of Watershed Hydrology*, Singh VP, Frevert DK, Meyer SP (eds). Water Resources Publications: Littleton, CO.

Lamb R, Beven K, Myrabo S. 1998. Use of spatially distributed water table observations to constrain uncertainty in a rainfall-runoff model. *Advances in Water Resources* **22**: 305–317.

Lee KT, Yen BC. 1997. Geomorphology and kinematic-wave-based hydrograph derivation. *Journal of Hydraulic Engineering-Asce* **123**: 73–80.

Nagai A, Kadoya M, Sugiyama H, Suzuki K. 1982. *Synthesizing Storage Function Model for Flood Runoff Analysis (in Japanese)*, Vol. 25 B-2. Disaster Prevention Research Institute, Kyoto University, Annuals; 207–220.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, Part 1, a discussion of principles. *Journal of Hydrology* **10**: 282–290.

Perrin C, Michel C, Andreassian V. 2001. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology* **242**: 275–301.

Reed S, Koren V, Smith M, Zhang Z, Moreda F, Seo DJ. 2004. Overall distributed model intercomparison project results. *Journal of Hydrology* **298**: 27–60.

Refsgaard JC, Henriksen HJ. 2004. Modelling guidelines—terminology and guiding principles. *Advances in Water Resources* **27**: 71–82.

Rodriguez-Iturbe I, Valdes JB. 1979. The geomorphologic structure of hydrologic response. *Water Resources Research* **15**: 1409–1420.

Sage AD, Melsa JL. 1971. *Estimation Theory with Applications to Communication and Control*. McGraw-Hill: New York.

Sargent RG. 1999. Validation and verification of simulation models. In *Proceedings of the 1999 Winter Simulation Conference*, Squaw Peak, Phoenix, AZ, The 1999 Winter Simulation Conference, 5–8 December 1999; 39–59.

Singh VP, Woolhiser DA. 1976. Sensitivity of linear and nonlinear surface runoff models to input errors. *Journal of Hydrology* **29**: 243–249.

Singh VP, Woolhiser DA. 2002. Mathematical modeling of watershed hydrology. *Journal of Hydrologic Engineering* **7**: 270–292.

Spear RC, Grieb TM, Shang N. 1994. Parameter uncertainty and interaction in complex environmental-models. *Water Resources Research* **30**: 3159–3169.

Storm B, Jensen KH, Refsgaard JC. 1988. Estimation of catchment rainfall uncertainty and its influence on runoff prediction. *Nordic Hydrology* **19**: 77–88.

Uhlenbrook S, Seibert J, Leibundgut C, Rodhe A. 1999. Prediction uncertainty of conceptual rainfall-runoff models caused by problems in identifying model parameters and structure. *Hydrological Sciences Journal-Journal des Sciences Hydrologiques* **44**: 779–797.

Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. *Hydrological Processes* **17**: 455–476.

Wolock DM. 1995. Effects of subbasin size on topographic characteristics and simulated flow paths in sleepers river watershed, Vermont. *Water Resources Research* **31**: 1989–1997.

World Meteorological Organization (WMO). 1975. *Intercomparison of Conceptual Models Used in Operational Hydrological Forecasting*, *WMO series, no. 429*, *Operational hydrology report, no. 7*. WMO, Geneva.

World Meteorological Organization (WMO). 1986. *Intercomparison of Models of Snowmelt Runoff*, *WMO series*, no. 646, *Operational hydrology report, no. 23*. WMO, Geneva.

World Meteorological Organization (WMO). 1992. *Simulated Real-time Intercomparison of Hydrological Models*, *WMO series*, no. 779, *Operational hydrology report no. 38*. WMO, Geneva.