# Hydrological model evaluation and comparison through uncertainty recognition and quantification

**SHEN CHIANG[1], YASUTO TACHIKAWA[2] & KAORU TAKARA[2]**

1 *Disaster Prevention Research Institute, Kyoto University & CREST, JST, Uji, Kyoto, 611-0011, Japan*
johnson@rdp.dpri.kyoto-u.ac.jp

2 *Disaster Prevention Research Institute, Kyoto University, Uji, Kyoto, 611-0011, Japan*

**Abstract** A method to identify hydrological prediction uncertainties is proposed through recognizing and quantifying different uncertainty sources in hydrological modelling. Based on this methodology, an index which originates from the Nash-Sutcliffe efficiency measure, named the Model Structure Indicating Index (MSII) is developed to quantify model structure uncertainty. The results of hydrological model comparisons show that the MSII can reflect the goodness of a hydrological model structure well. The index can be used as a tool for implementing quantitative comparison and selection of models.

**Key words** hydrological prediction; model comparison; model evaluation; prediction uncertainty

## INTRODUCTION

A hydrological model is an integration of mathematical descriptions of conceptualized hydrological processes. It is difficult to completely represent hydrological phenomena in mathematical descriptions. Thus the performance of a hydrological model is highly dependent on the hypothesis of the model, the hydrological data used for calibration and simulation, the calibration method and its model structure. Underestimating or misunderstanding of these factors and the interrelation among them may result in the tremendous misinterpretation of the results of hydrological models. For many years, hydrologists have tried to understand the effects of these factors on the accuracy and reliability of estimated hydrological variables such as peak flow and flood volume.

However, it is not well understood how most of those hydrological models are used under an unknown magnitude of input uncertainty. A calibration process makes the model functionally capable of generating a good simulation of the watershed response data series. To ensure the calibration process is acceptable and workable, a validation process is essential. Again, the precision or the accuracy of the data, which is used for its validation, is unknown. The methodology proposed in this study tries to cut the Gordian knot by acknowledging that input uncertainty is inevitable in hydrological modelling.

The methodology is developed to recognize and quantify the different uncertainty sources by observing model behaviour under different levels of input uncertainty. The adequacy of hydrological models applied to the same watershed is then ranked. Instead of sampling the parameter space directly as the Generalized Likelihood Uncertainty Estimation (GLUE; Beven & Binley, 1992) procedure does, the methodology proposed here generates the parameter space by introducing noise into the input data with a specified probability distribution function. This approach reflects the problem that partial parameter uncertainty comes from the uncertainty of data at hand and the way the model structure responds to it.

The Nash-Sutcliffe efficiency (Nash & Sutcliffe, 1970) is used as a measure to quantify predictive uncertainty; then an index which originates with the Nash-Sutcliffe efficiency named the Model Structure Indicating Index (MSII) is used to quantify model structure uncertainty. The index is used as a tool for implementing quantitative comparison and selection of models.

Several hydrological models are used for model comparison in this study. They are: Storage Function Method (SFM) (Kimura, 1961), TOPMODEL (Beven & Kirkby, 1979), and KW-GIUH (Lee & Yen, 1997). The comparison results show that within a small magnitude of input uncertainty, there appears to be no apparent distinction between the capability of the hydrological models to adapt themselves to the error contaminated data, while, with increasing input uncertainty, the differences become larger and are quantified by the index MSII (a larger value of MSII indicates a worse model structure). As a final result, among the candidate models in this study, KW-GIUH is the most structurally stable model, followed by TOPMODEL, then SFM and the worst is the parameter-constrained SFM. For watershed management or planning, the methodology proposed here enables the selection of a hydrological model as a good reference.

## UNCERTAINTY QUANTIFICATION IN HYDROLOGICAL MODELS

The performance of hydrological models is profoundly affected by the sources of uncertainty; briefly they are observed data, model calibration, and model structure. Among those, input data uncertainty occupies the greatest source of uncertainty and readily contaminates the other sources. In this study, prediction uncertainty, which comes from the three kinds of sources mentioned above, is classified into four categories: system uncertainty, entire uncertainty, inherent uncertainty, and structure uncertainty. The definitions and the procedures of recognizing and quantifying them are described below:

### System uncertainty

Even if it is well known that a hydrological model is an approximation of the real phenomena based on the hydrological cycle, it still needs to be stressed that existing discrepancies are always between the model outcome and the observed data, no matter how precise and how perfectly calibrated the model is. This is the model predicting limitation beneath the hypothesis and the architecture of a hydrological model, and is defined as the system uncertainty in this study.

The system uncertainty is given by evaluating the discrepancy between the observed watershed response series and the model outcome during the process of model parameter calibration. The discrepancy is supposed to be the minimum value of the discrepancy to the possible occurrence of the event. It has often been observed that the goodness-of-fit between the observed data and the estimated data during the calibration process is better than during the validation process. Due to this observation, the uncertainty occurring here denotes the predicting limitation of the model, since this is the best performance that the model can achieve. In agreement with this definition, it is clear that the uncertainty comes from the process of calibration and thus, the real uncertainty source is the data used for calibration. The index for quantifying the system uncertainty is defined as:

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

where $Q_o$ and $Q_b$ indicate the observed watershed response series and model outcome by using the best performing parameter set. $n$ is the time step of the time series. $SU$ elucidates the performance of the most suitable simulated outcomes.

This is a measure of the model performance during calibration, which denotes the predictive capability of the 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 & Henriksen, 2004). As a result of this, the system uncertainty will be less than the entire uncertainty, which is described in the following.

## Entire uncertainty

After calibrating the model parameters, the calibrated parameter space reflects its variance through the model structure and propagates to the response surface. The entire uncertainty is recognized by examining the discrepancy between observed watershed response data and model outcome using input data and the parameter sets in the calibrated parameter space. The entire uncertainty is the utmost uncertainty that a model could have under existing input uncertainty. The index for quantifying the entire uncertainty is defined as:

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

where $Q_e$ is the model outcome acquired by using a parameter set within the whole parameter space.

## Inherent uncertainty

The inherent uncertainty represents the variability of the parameter space, which is determined according to the input uncertainty. This uncertainty is examined by calculating the discrepancy between model outcomes using parameter values within the parameter space and the outcome with the best-fit parameter set. The index for quantifying the inherent uncertainty is defined as:

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

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

**Structure uncertainty—Model Structure Indication Index (MSII)**

In order to implement a dynamic view of the relationship among the system uncertainty, the inherent uncertainty and the entire uncertainty caused by input data uncertainty, the Nash-Sutcliffe efficiency is used to formulate a Model Structure Indicating Index (MSII, Chiang *et al.*, 2005) defined as:

$$MSII = \frac{IU - EU}{SU} \tag{4}$$

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, while the model is considered to have a better chance of reproducing a more real watershed response series. It is a measure of the possibility of a model that adapts itself to the input uncertainty. The larger the magnitude the poorer the possibility that the model adapts itself to the error contaminated input data; this indicates that the calibrated parameter space lacks the capability to drive the model to simulate the watershed response well because of shortcomings in the structure of the hydrological model.

The numerator shows a measure of the adaptiveness of a hydrological model to uncertainty contained input data, while only the numerator is not enough to reveal the model structure; a component showing the model predictive capability is needed. The denominator of MSII indicates the effectiveness/goodness of the model calibration results and the predictive capability of the model, which enables the explicit reflection of the calibration scheme. The index interprets the variance caused by the calibration process and the model structure in a dimensionless form through the Nash-Sutcliffe efficiency. During the implementation of model evaluation, the system uncertainty, which is less than zero, is excluded from the calculation of MSII due to its insignificance. Hence the smaller value of the MSII represents better model structure, and the range of MSII is: $0 \le MSII < \infty$.

## ALGORITHM AND DATA APPLIED FOR UNCERTAINTY IDENTIFICATION

Instead of sampling the parameter space directly, like GLUE does, this study generates the parameter space by introducing noise items into the input data with specified probability distributions. Here, a normal distribution with a mean equal to zero and standard deviation from 1.0 to 9.0 (mm $h^{-1}$) is used to acquire the model parameter space and its outcomes under different input uncertainty. For each specified standard deviation of input uncertainty, 10 000 model outcomes were derived from the combinations of 100 rainfall series realizations and 100 fitted parameter sets through the model calibration.

The Yasu River basin, which is located in the Shiga prefecture of Japan, was used in this study. The main stream length of the Yasu River is about 65 km and the total basin area is approximately 387 $km^2$. The rainfall data was collected from four rainfall-gauging stations inside the Yasu River basin. The catchment average precipitation was calculated by using the Thiessen polygon method. The rainfall record of the event used here was applied for rainfall realizations with noise, which were generated using a normal distribution with mean equal to zero and the specified standard deviation as input uncertainty. The water level data acquired at the Yasu gauging station has an hourly time step. A rating curve was used to transform the water levels to discharge.

## RESULTS

The methodology described above is applied here for model performance comparisons of four hydrological models. They are TOPMODEL, KW-GIUH, SFM and a parameter-constrained SFM. The equations of SFM are:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = r_e(t - T_l) - q, \quad S = kq^p$$
$$r_e = \begin{cases} f \times r, & \text{if} \quad \sum r \le R_{SA} \\ r, & \text{if} \quad \sum r > R_{SA} \end{cases} \tag{5}$$

where $S$ = water storage height; $r_e$ = effective rainfall intensity; $q$ = runoff height; $t$ = time step; and $T_l$ = the lag-time. 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 accumulating saturated rainfall. A fully functional SFM is used with parameters $T_l$, $f$, and $R_{SA}$. For an example of a poorer-structured model with comparison to the original SFM, a parameter-constrained SFM is manipulated by fixing the value of $R_{SA}$ to 0.0.

Figure 1 shows the calibration results of the models. The performance is evaluated by the Nash-Sutcliffe efficiency, which is summarized in Table 1. It can be seen that TOPMODEL is the best; and the parameter-constrained SFM is the worst. Figures 2–5 show 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 the entire uncertainty will increase as input uncertainty increases; and the discrepancy between the entire and the inherent uncertainty has the same tendency.

The system uncertainty is supposed to be located in the middle, between the entire and the inherent uncertainty. The range between the entire and inherent uncertainty indicates the goodness of a model structure; that is the ability of the model to adapt itself to input data error. The broader range indicates that the model is less structurally sound.

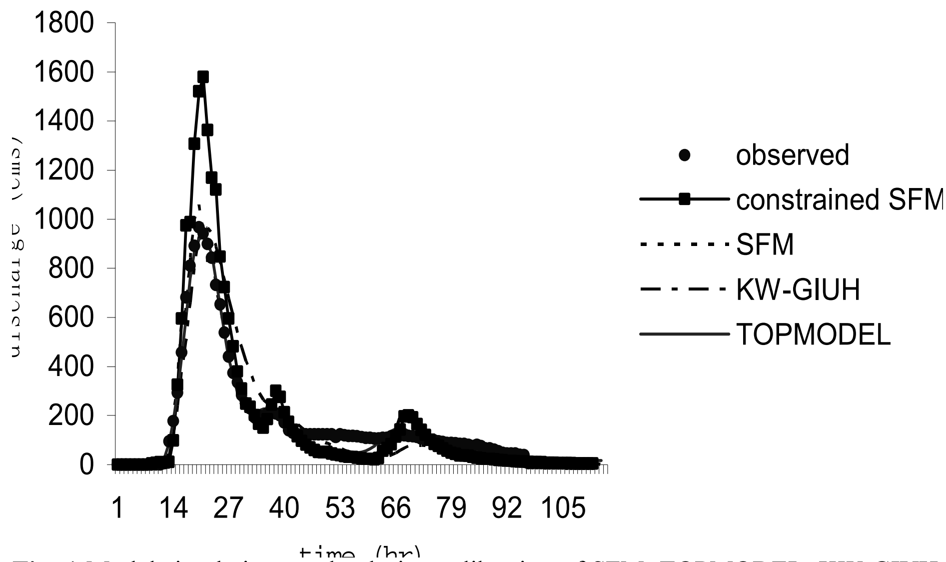


**Fig. 1** Model simulation results during calibration of SFM, TOPMODEL, KW-GIUH and parameter-constrained SFM.

**Table 1** Summary of the model performance during calibration.

| Model | SFM | TOPMODEL | KW-GIUH | Parameter-constrained SFM |
|---|---|---|---|---|
| Nash-Sutcliffe efficiency | 0.795 | 0.864 | 0.773 | 0.454 |

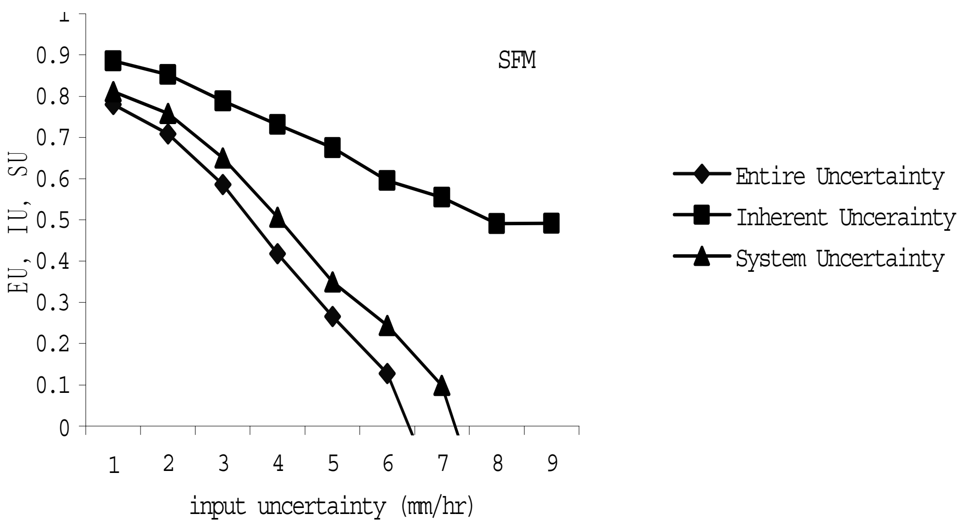


**Fig. 2** Entire, inherent and system uncertainty of SFM.

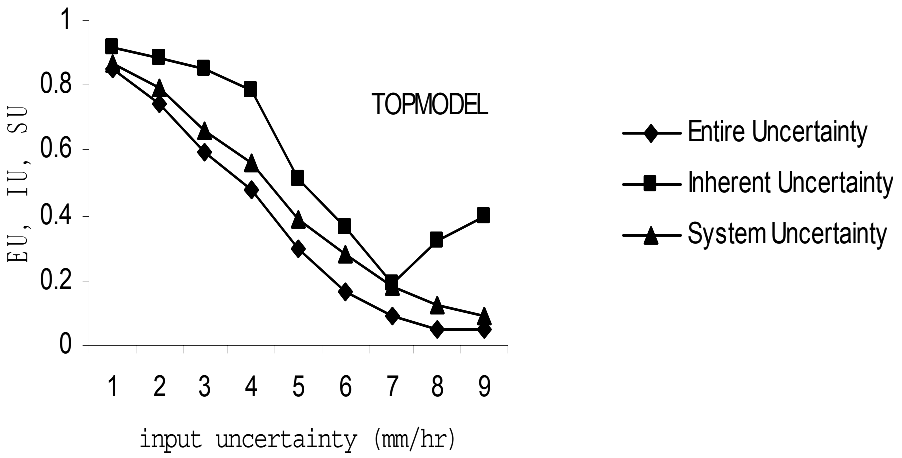


**Fig. 3** Entire, inherent and system uncertainty of TOPMODEL.

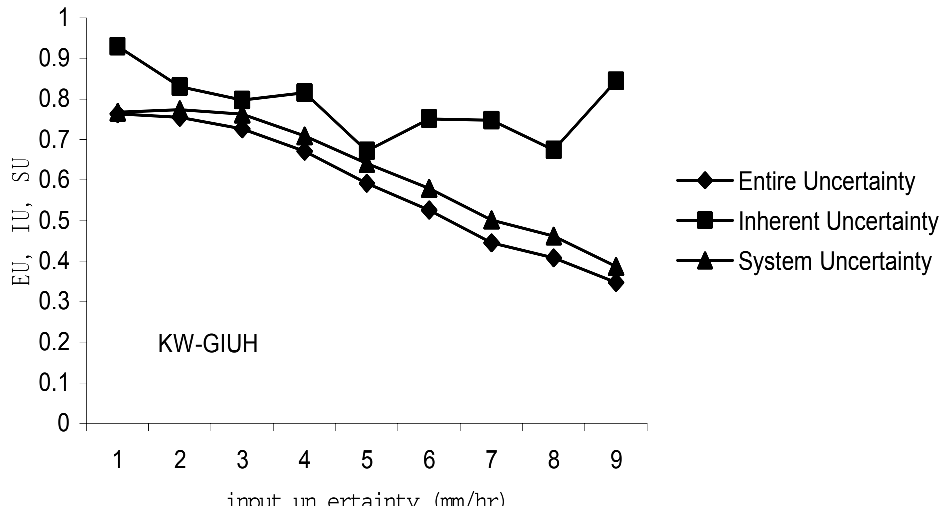


**Fig. 4** Entire, inherent and system uncertainty of KW-GIUH.

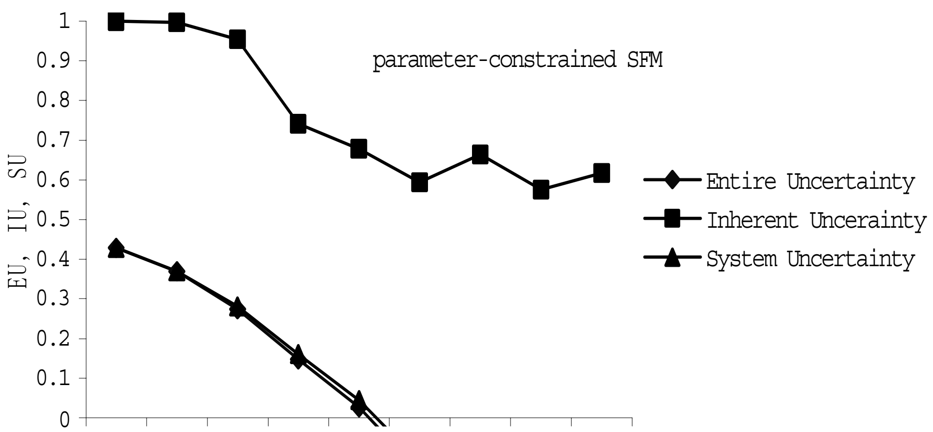


**Fig. 5** Entire, inherent and system uncertainty of parameter-constrained SFM.

Figure 6 shows the MSII of each model. Except for the parameter-constrained SFM, there is no apparent distinction between the other three candidate models within a small magnitude of the input uncertainty. However, with increasing input uncertainty, KW-GIUH becomes structurally more stable than SFM and TOPMODEL in this study.

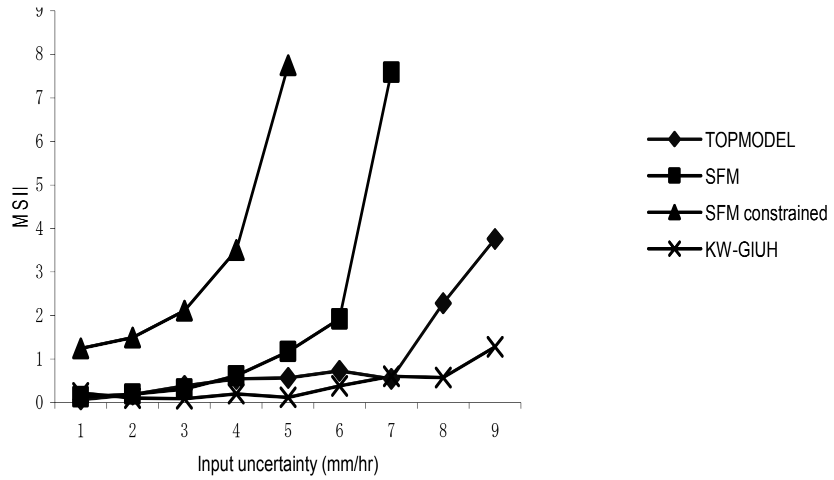


**Fig. 6** MSII of TOPMODEL, SFM, parameter-constrained SFM and KW-GIUH.

## CONCLUSIONS

Although TOPMODEL performs best during the calibration process, when the input uncertainty increases the capability of TOPMODEL to adapt itself to the error contaminated data fails. This result reveals a very interesting point; that is, even if the performance of a model during calibration is better than that of other models, it may not be the most stable model while input uncertainty exists. KW-GIUH does not perform as well as TOPMODEL does with small input uncertainty, but with increasing input uncertainty, the capability of KW-GIUH to adapt itself to the error contaminated data is better than TOPMODEL. For SFM, with small magnitude input uncertainty, the model performance is not so different from TOPMODEL. However, the steep slope of MSII in Fig. 6 indicates that it lacks this ability at high input uncertainty levels.

## REFERENCES

Beven, K. & Kirkby, M. J. (1979) A Physically based variable contributing area model of catchment hydrology. *Hydrol. Sci. Bull*. **24**, 43–69.

Beven, K. J. & Binley, A. M. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Processes* **6**, 279–298.

Chiang, S., Tachikawa, Y. & Takara, K. (2005) An approach to uncertainty identification in hydrological modelling. *Ann. Hydraulic Engng, JSCE* **49**, 211–216.

Kimura, T. (1961) The flood runoff analysis by using the storage function method (in Japanese). The Public Works Research Institute, Ministry of Construction, Japan.

Lee, K. T. & Yen, B. C. (1997) Geomorphology and kinematic-wave-based hydrograph derivation. *J. Hydraulic Engng-ASCE* **123**, 73–80.

Nash, J. E. & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I – A discussion of principles. *J. Hydrol*. **10**, 282–290.

Refsgaard, J. C. & Henriksen, H. J. (2004) Modelling guidelines—terminology and guiding principles. *Adv. Water Resour*. **27**, 71–82.