



Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models

Lu Li^{a,b,c,*}, Jun Xia^a, Chong-Yu Xu^c, V.P. Singh^{d,e}

^a Key Laboratory of Water Cycle & Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China

^b Graduate University of Chinese Academy of Sciences, Beijing 100049, China

^c Department of Geosciences, University of Oslo, Oslo, Norway

^d Department of Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843-2117, USA

^e Department of Civil and Environmental Engineering, Texas A&M University, College Station, TX 77843-2117, USA

ARTICLE INFO

Article history:

Received 25 March 2010

Received in revised form 28 June 2010

Accepted 30 June 2010

This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Vazken Andréassian, Associate Editor

Keywords:

Uncertainty assessment

Hydrological models

GLUE

Bayesian methods

WASMOD

DTVGM

SUMMARY

Quantification of uncertainty of hydrological models has attracted much attention in the recent hydrological literature. Different results and conclusions have been reported which result from the use of different methods with different assumptions. In particular, the disagreement between the Generalized Likelihood Uncertainty Estimation (GLUE) and the Bayesian methods for assessing the uncertainty in conceptual watershed modelling has been widely discussed. What has been mostly criticized is the subjective choice as regards the influence of threshold value, number of sample simulations, and likelihood function in the GLUE method. In this study the impact of threshold values and number of sample simulations on the uncertainty assessment of GLUE is systematically evaluated, and a comprehensive evaluation about the posterior distribution, parameter and total uncertainty estimated by GLUE and a formal Bayesian approach using the Metropolis Hasting (MH) algorithm are performed for two well-tested conceptual hydrological models (WASMOD and DTVGM) in an arid basin from North China. The results show that in the GLUE method, the posterior distribution of parameters and the 95% confidence interval of the simulated discharge are sensitive to the choice of the threshold value as measured by the acceptable samples rate (ASR). However, when the threshold value in the GLUE method is high enough (i.e., when the ASR value is smaller than 0.1%), the posterior distribution of parameters, the 95% confidence interval of simulated discharge and the percent of observations bracketed by the 95% confidence interval (P-95CI) for the GLUE method approach those values estimated by the Bayesian method for both hydrological models. Second, in the GLUE method, the insufficiency of number of sample simulations will influence the maximum Nash–Sutcliffe (MNS) efficiency value when ASR is fixed. However, as soon as the number of sample simulations increases to 2×10^4 for WASMOD and to 8×10^4 for the DTVGM model the influence of number of sample simulations on the model simulation results becomes of minor importance. Third, the uncertainty in simulated discharges resulting from parameter uncertainty is much smaller than that resulting from the model structure uncertainty for both hydrological models. Fourth, the goodness of model fit as measured by the maximum Nash–Sutcliffe efficiency value is nearly the same for the GLUE and the Bayesian methods for both hydrological models. Thus this study provides useful information on the uncertainty assessment of hydrological models.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The assessment of uncertainty of hydrological models is of major importance in hydrologic modelling. Different methods have been used for uncertainty assessment and different results have been obtained. Uncertainty assessment methods can be broadly

* Corresponding author at: Key Laboratory of Water Cycle & Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China. Tel.: +86 64889010.

E-mail address: marylilu420@gmail.com (L. Li).

classified into two groups, i.e., the Bayesian methods (Krzysztofowicz, 1999; Thieman et al., 2001; Engeland et al., 2005) and the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992; Beven and Freer, 2001). In the literature, the former method is named as the formal Bayesian approach and the latter is called the informal Bayesian approach. Based on these two methods or a combination thereof, many methodologies have been developed during the past two decades to better treat uncertainty and they have been applied to various catchments models (Vrugt et al., 2003; Blasone et al., 2008a,b; Vogel et al., 2008; Xiong and O'Connor, 2008). Both methods have been discussed with

respect to their philosophies and the mathematical rigor they rely on (Gupta et al., 2003; Beven, 2006; Mantovan and Todini, 2006; Todini, 2007; Beven et al., 2008; Yang et al., 2008; Vrugt et al., 2009; Jin et al., 2010).

When applying the formal Bayesian method, the difficulty is to find a proper statistical model which fits the data. Statistical transforms have been employed to get normally distributed residuals. Two transform methods that have been widely used in uncertainty analyses of hydrological models include the Box-Cox method (Engeland et al., 2005; Yang et al., 2007; Wang et al., 2009; Jin et al., 2010) and the Normal Quantile Transform method (NQT) (Van der Waerden 1952, 1953a,b; Krzysztofowicz, 1997, 1999; Todini, 2008). In the hydrological uncertainty processor (HUP), which aims to estimate the predictive uncertainty of hydrological models, Krzysztofowicz (1999) introduced the idea of converting both observation and simulation data into a normal space by means of the Normal Quantile Transform (NQT). With this method, the residuals in a normal space can be used conveniently by the Bayesian inference to get the posterior probability density function. Recently, this method has been widely used in hydrological uncertainty analyses for deriving the joint distribution and predictive conditional distribution to form a multivariate distribution (Krzysztofowicz and Kelly, 2000; Maranzano and Krzysztofowicz, 2004; Montanari and Brath, 2004; Reggiani and Weerts, 2008; Todini, 2008). Engeland et al. (2010) compared three statistical models which are auto-regressive model for Box-Cox transformation, auto-regressive model for NQT-transformation and separate models for the positive and negative errors. The results showed that the first two statistical models have almost identical results. Besides, even input and model structure errors will bring correlated residuals because of the memory inherent in hydrological processes. Both discrete-time and continuous-time auto-regressive error models have been used recently (Yang et al., 2007).

On the contrary, GLUE is popular for its conceptual simplicity, ease of implementation and flexibility of less modification to existing source codes of hydrological models. However, GLUE has some drawbacks. It is dependent on subjective decisions and not based on the statistically consistent error models (Blasone et al., 2008a). It has an informal measure which has weak procedures to summarize parameter and predictive distributions (Vrugt et al., 2009), which will be “incoherent and inconsistent” with less formal likelihood functions (Mantovan and Todini, 2006). In response, Beven et al. (2007, 2008) argue that the problem is the likelihood function that overestimates the data information and GLUE will yield results identical to those of the formal Bayesian method if the likelihood function used was correct. Therefore, the comparison of these two methods has been done in recent years. Some studies show that the Bayesian method and GLUE can generate very similar estimates of stream flow uncertainty under certain conditions (Vrugt et al., 2009; Jin et al., 2010). On the other hand, other studies point out that prediction limits estimated by GLUE can be very different from prediction limits estimated by the Bayesian methods (Montanari, 2005; Mantovan and Todini, 2006; Stedinger et al., 2008). Todini (2007) states that the formal Bayesian inference can largely reduce the prior predictive uncertainty and should be used in preference to GLUE. Yang et al. (2008) demonstrated that the GLUE method has a widest marginal parameter uncertainty intervals compared with Bayesian methods.

Previously published results have shown that the determination of the likelihood function is the key point in the Bayesian method which determines its success. The GLUE method is influenced by some subjective decisions. Stedinger et al. (2008) point out that GLUE methodology can be a useful tool for model calibration and uncertainty analyses only if a proper likelihood function for a normal independent distribution error can be found and the imposition of an arbitrary behavioural threshold cannot be helpful.

However, other studies show that threshold values and ranges of parameters are the important factors influencing the results of the GLUE method (Yang et al., 2008; Jin et al., 2010).

Survey of literature shows that although many studies have been reported on the application and comparison of the formal Bayesian and GLUE methods, the following issues have not been addressed which motivated the present study: (1) there is a need for a systematic evaluation and quantification of the effect of threshold values in GLUE on the parameter and model uncertainties; (2) there is a need for a systematic evaluation and quantification of the effect of number of sample simulations in GLUE on the parameter and model uncertainties; and (3) there is a need for efficient and objective evaluation criteria for uncertainty assessment with different methods and models.

This study makes a comprehensive evaluation about the parameter and total uncertainty estimated by GLUE and a formal Bayesian approach using Metropolis Hasting (MH) algorithm, for two well-tested conceptual hydrological models (WASMOD and DTVGM) in an arid basin from North China. Special attention is paid (1) to quantify the impact of threshold values or the acceptable sample rate (ASR) on the GLUE method; (2) to quantify the impact of number of sample simulations on the results of GLUE; and (3) to compare the performance of the Bayesian method and the GLUE method with different choices of threshold values and number of sample simulations. It also shows how to correctly set the subjective decisions of the GLUE method with models to assure the reasonableness of the results of uncertainty analysis which is more consistent with the Bayesian method. For deriving the likelihood function for the Metropolis Hasting algorithm, it is convenient to have the simulation errors which are normally distributed with zero mean and constant variance, and time independent. In this study the Normal Quantile Transform (NQT) is used to transform observed and simulated discharges into a Gaussian distribution and the AR (1) Gaussian error model is used to remove the time dependence of residuals. As for the GLUE method, the dependence of the results on ASR and the number of samples are investigated.

2. Method

For the sake of comparison and understanding of the results, the methods and the procedures used in this study are briefly described in what follows.

2.1. Bayesian method

2.1.1. Bayesian inference

The Bayesian theorem is expressed as follows:

$$\pi(\varphi|\eta) = \frac{f(\eta|\varphi) \cdot f(\varphi)}{\int f(\eta|\varphi) \cdot f(\varphi) d\varphi} \quad (1)$$

where $\varphi = \{\theta, \omega\}$, in which θ represents the hydrological model parameters and ω represents statistical parameters. The posterior density $\pi(\varphi|\eta)$ can be derived from the prior density $f(\varphi)$ and the likelihood function $f(\eta|\varphi)$. Variable η is a transformed variable from the original space.

2.1.2. Transformation

The Box-Cox transformation has been used by Engeland and Gottschalk (2002) and Yang et al. (2007) and it can be expressed as:

$$h(y, \lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y) & \lambda = 0 \end{cases} \quad (2)$$

The square-root transformation is a special case of the Box-Cox transformation. λ -parameter is used in the residual test for choosing a statistical model in this study (Engeland et al., 2005; Wang et al., 2009; Jin et al., 2010).

The NQT method has been described in detail by Krzysztofowicz and Kelly (2000) and Todini (2008). Let y_t and y_t^M represent the observed and simulated streamflow, respectively. Then, η_t and η_t^M are Gaussian variants which are obtained by matching an empirical probability distribution $\Gamma(\cdot)$ with a Gaussian distribution $Q(\cdot)$ and then performing an inversion $Q^{-1}(\cdot)$:

$$\eta_t = Q^{-1}(\Gamma(y_t)) \quad (3)$$

$$\eta_t^M = Q^{-1}(\Gamma(y_t^M)) \quad (4)$$

In which Γ is the empirical distribution. The inverse NQT is represented as $\Gamma^{-1}(Q(\cdot))$ (Reggiani and Weerts, 2008). Prior densities of all (model parameters, statistical parameters, and runoff) are considered to be uniform.

2.1.3. Likelihood function

The likelihood function is the key issue in the Bayesian method, which is defined by the distribution of residuals. ξ_t represents the residual of $\eta_t(y_t)$ and $\eta_t^M(y_t^M)$ as follows:

$$\xi_t = \eta_t(y_t) - \eta_t^M(y_t^M) \quad (5)$$

It is needed to check whether ξ_t is independent. If not, an AR (1) model is used to make the residuals independent (Eq. 6).

$$\xi_t = a \cdot \xi_{t-1} + \varepsilon \quad (6)$$

In this study NQT was used to transform the observed and simulated discharges and the resulting residuals are normally distributed. Then the Jarque–Bera test (Carlos and Bera, 1980) was used to check the independence of residuals, and the results showed that an AR (1) model was needed to make the NQT-residuals independent. The resulting likelihood function is:

$$f(\xi|\varphi) = \frac{1}{\sigma} q\left(\frac{\xi_0}{\sigma}\right) \cdot \prod_{t=1}^T \frac{1}{\sigma} q\left(\frac{\xi_t - a\xi_{t-1} - b}{\sigma}\right) \quad (7)$$

In which φ represents all parameters; T represents the time. Where q represents the normal density operator. Because there is no systematic error, b is set to zero, which results in the following likelihood function:

$$f(\xi|\varphi) = \frac{1}{\sigma} q\left(\frac{\xi_0}{\sigma}\right) \cdot \prod_{t=1}^T \frac{1}{\sigma} q\left(\frac{\xi_t - a\xi_{t-1}}{\sigma}\right) \quad (8)$$

2.1.4. Posterior density

There is in general no information about the distribution of parameters. The prior probability density of the parameters is generally taken as non-informative multi-uniform distribution in hydrological applications (Engeland et al., 2005; Liu et al., 2005; Yang et al., 2008; Jin et al., 2010). The effect of a prior density tends to vanish compared with that of the likelihood when dealing with large samples of observations. In this study, uniform priori distribution is also used for most parameters. Parameter σ is unknown standard deviation with Jeffreys' uninformative prior which is proportional to σ^{-1} (Bernardo and Smith, 1994; Yang et al., 2007). The prior densities of all parameters in the two hydrological models are shown in Table 3. With the priori density $f(\varphi)$ considered to be uniform, the posterior density $\pi(\varphi|\xi)$ is given as follows:

$$\pi(\varphi|\xi) = \frac{f(\xi|\varphi) \cdot f(\varphi)}{\int f(\xi|\varphi) \cdot f(\varphi) d\varphi} \quad (9)$$

Substituting Eq. (8) into Eq. (9) results in:

$$\pi(\varphi|\xi) = \frac{\frac{1}{\sigma} q\left(\frac{\xi_0}{\sigma}\right) \cdot \prod_{t=1}^T q\left(\frac{\xi_t - a\xi_{t-1}}{\sigma}\right) \cdot f(\varphi)}{\int \frac{1}{\sigma} q\left(\frac{\xi_0}{\sigma}\right) \cdot \prod_{t=1}^T q\left(\frac{\xi_t - a\xi_{t-1}}{\sigma}\right) \cdot f(\varphi) d\varphi} \quad (10)$$

In which $\sigma > 0$, $a \in [0, 1)$.

2.2. Metropolis Hasting algorithm

The Metropolis Hasting (MH) algorithm is an algorithm for constructing Markov chains to draw samples from the Bayesian posterior distribution. The Markov chain starts from a random initial value and the new sample is derived from the proposal distribution, then it is judged by the accept probability to know whether or not to accept the new sample. After that, the algorithm is run for many iterations until this initial sample is “forgotten”. These discarded samples constitute what is known as the “burn in” period. The remaining set of accepted values represents a sample which will converge to the posterior distribution after ‘burn in’. Details of the Metropolis Hasting method have been reported by Chib and Greenberg (1995), Kuczera and Parent (1998), and Engeland and Gottschalk (2002), among others. The main steps are as follows:

1. $x^{(L)}$ is the current state and $L = 1, 2, \dots, m$.
2. Draw I from $1, 2, \dots, n$, randomly and only once for each iteration. n is the number of parameters.
3. x^* is derived from the proposal distribution d :

$$x^* \sim d_I(x^{(L)}, x^*) \quad \text{where} \quad x_j^* = x_j^{(L)} \quad \forall j \neq I \quad (11)$$

4. Then the new state is chosen according to the accept probability:

$$a(x^{(L)}, x^*) = \min \left\{ 1, \frac{\pi(x^*) \cdot d_I[x^*, x^{(L)}]}{\pi(x^{(L)}) \cdot d_I[x^{(L)}, x^*]} \right\} \quad (12)$$

$$x_I^{(L+1)} = \begin{cases} x_I^* & \text{with probability } a(x^{(L)}, x^*) \\ x_I^{(L)} & \text{with probability } 1 - a(x^{(L)}, x^*) \end{cases} \quad (13)$$

By definition, the random variable in normal distribution has a range of $(-\infty, \infty)$. Normal distribution is not appropriate for $\sigma > 0$. However, in this study, the result shows that the mean of σ is much larger than the standard deviation of σ , and the mean of σ minus three sigma of σ is still larger than zero. It means that the area covered by the normal distribution density function lies outside the bounds is nearly zero. So in this particular case, the use of normal distribution even though σ is bounded by zero is justified. This is perhaps the reason that normal distribution is widely used for hydrological data which, in most cases, are bounded by zero. Other statistical parameters and model parameters in this paper satisfy the same assumption. Then random walk Metropolis Hasting algorithm is used and the proposal distributions are normal distributions as follows:

$$x^* \sim d(x^{(L)}, x^*) \quad (14)$$

$$d(\sigma) \sim N[\sigma^L, S_\sigma] \quad (15)$$

$$d(a) \sim N[a^L, S_a] \quad (16)$$

$$d(\theta) \sim N[\theta^L, S_\theta] \quad (17)$$

The algorithm works best if the proposal density matches the target distribution which is unknown normally. A proposal density is used and the variance parameter has to be tuned during the “burn in” period. Chib and Greenberg (1995) pointed out that the tuning of the variance is done by calculating the acceptance rate,

which depends on the target distribution. It has been shown that the ideal acceptance rate is approximately 40–50% for one parameter updating, decreasing to approximately 20–30% for multi-parameter updating in one block (Engeland and Gottschalk, 2002). The standard deviations S_σ , S_α , S_θ should be tuned by the acceptance rate. If the standard deviation is too small the chain will mix slowly, while if the standard deviation is too large the acceptance rate will be very low. In this study, there are five parallel chains and each of them contains 5000 iterations; first 1000 of which will be removed as burn in period. The resulting acceptable rate is about 23%.

Furthermore, convergence of the chain is another key problem in the Metropolis Hasting method. In this study, the Scale Reduction Score \sqrt{R} (Gelman and Rubin, 1992) was used to check whether the MC chain converges. The value of \sqrt{R} is calculated by the following function:

$$\sqrt{R} = \sqrt{\frac{n-1}{n} + \frac{m+1}{m} \cdot \frac{B}{n} \cdot \frac{1}{W}} \quad (18)$$

$$\text{in which } \frac{B}{n} = \frac{\sum_{i=1}^m (u_i - \bar{u})^2 / (m-1)}{\quad} \quad (19)$$

$$W = \frac{\sum_{i=1}^m s_i^2 / m}{\quad} \quad (20)$$

where m is the number of sequences, n represents the length of the sequence, u_i is the mean of the n samples within-sequence, \bar{u} is the mean of u_i , s_i^2 is the variances of the n samples within-sequence, each based on $n-1$ degrees of freedom, W is the average of s_i^2 , and B/n is the variance between the m sequence means u_i . The chain converges when \sqrt{R} is close to 1.0.

2.3. GLUE

In the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992) a large number of model runs are made with many different randomly chosen parameter values selected from a priori probability distribution. The acceptability of each run is evaluated against observed values and, if the acceptability is below a certain subjective threshold, the run is considered to be “non-behavioral” and that parameter combination is removed from further analysis. In this method, the likelihood values serve as relative weights of each parameter set or simulated value. It is noted that the likelihood function and the threshold are subjectively determined and this was discussed by Freer et al. (1996). In this study, the Nash–Sutcliffe (NS) value was chosen as the likelihood function:

$$L(\theta_i|Y) = 1 - \frac{\sum_{t=1}^T (R_{obs,t} - R_{sim,t})^2}{\sum_{t=1}^T (R_{obs,t} - \bar{R}_{obs})^2} = 1 - \frac{\sigma_i^2}{\sigma_{obs}^2} \sigma_i^2 < \sigma_{obs}^2 \quad (21)$$

where $L(\theta_i|Y)$ is the likelihood measure; $R_{obs,t}$ is the observed discharge; $R_{sim,t}$ is the simulated discharge, which is depending on the model parameter θ_i ; \bar{R}_{obs} is the average value of $R_{obs,t}$; σ_i^2 is the variance of errors for the given parameter set θ_i and the observed discharge data set Y ; and σ_{obs}^2 is the variance of the observed data set. In this study, different threshold values and sampling sizes were chosen to evaluate their impacts on the simulation results of the GLUE method.

2.4. Ninety-five percent confidence intervals of discharge

Two thousand discharge values for each month are obtained by running the hydrological models with 20,000 parameter sets, which are from the Metropolis Hasting samples. The 95% confidence intervals for discharge due to parameter uncertainty are estimated by these discharge samples. The 5 percentile and 95 per-

centile of discharge are derived by sorting ascending of 20,000 discharge values at each month. The intervals in GLUE method consider weighted values, while in the Bayesian method they do not. It means that at each month, after sorting ascending of all discharge, the quantiles of discharge in Bayesian and GLUE methods are calculated as follows:

$$P_B(y < y_i) = i/n \quad (22)$$

$$P_G(y < y_i) = NS_i / \sum_{i=1}^n NS_i \quad (23)$$

In which, i is the position order of y_i after sorting ascending; n is the number of sample simulations; $P_B(y < y_i)$ is the quantile of discharge y_i in Bayesian method; $P_G(y < y_i)$ is the quantile of discharge y_i in GLUE method; NS_i is the weighted Nash–Sutcliffe values of y_i , which is equal to $L(\theta_i|Y)$.

The 95% confidence intervals for Bayesian method due to parameter uncertainty and model uncertainty for discharge were calculated by adding the model residuals in the form of a normal random uncertainty with zero mean and variance σ^2 to each of the 20,000 sample discharge values that are available for each time step. For the Bayesian method the inverse of NQT-transformation should be used

$$\eta_{i,t} = \eta_{i,t-1}^M + a \cdot (\eta_{i,t-1} - \eta_{i,t-1}^M) + rnorm(0, \sigma)_i \quad (24)$$

$$y_{i,t} = Q^{-1}(\Gamma(\eta_{i,t})) \quad (25)$$

where i is the sample index; t is the time index; $rnorm$ is the random value from a normal distribution with zero mean and variance σ^2 ; and Q^{-1} represents the inverse of NQT. Then the 5 percentile and 95 percentile of discharge due to parameter uncertainty and model uncertainty are derived by sorting ascending of the new 20,000 discharge values $y_{i,t}$ which are derived from Eq. (25) at each month.

2.5. Criteria for comparison

In this paper, three indices were used to compare the derived 95% confidence interval (95CI) of monthly discharge, which are the Average Relative Interval Length (ARIL) defined by Jin et al. (2010), the percent of observations bracketed by the 95CI (P-95CI) used by Li et al. (2009), and the maximum Nash–Sutcliffe value (MNS) (Nash and Sutcliffe, 1970). These indices are expressed as follows:

$$ARIL = \frac{1}{n} \sum \frac{Limit_{Upper,t} - Limit_{Lower,t}}{R_{obs,t}} \quad (26)$$

$Limit_{Upper,t}$ and $Limit_{Lower,t}$ are the upper and lower boundary values of 95CI; n is the number of time steps; and $R_{obs,t}$ is the observed discharge

$$P - 95CI = \frac{NQ_{in}}{n} \times 100\% \quad (27)$$

where NQ_{in} is the number of observations which are contained in 95CI.

The goodness of the simulation was judged on the basis of the closeness of ARIL to 0 and the P-95CI to 100%.

$$MNS = \max_{i=1}^N \{NS_i\} \quad (28)$$

$$NS = 1 - \frac{\sum_{t=1}^T (R_{obs,t} - R_{sim,t})^2}{\sum_{t=1}^T (R_{obs,t} - \bar{R}_{obs})^2} \quad (29)$$

in which i is the acceptable sample index; t is the time index; $R_{obs,t}$ is the observed discharge; $R_{sim,t}$ is the simulated discharge; and \bar{R}_{obs} is the average value of $R_{obs,t}$. MNS should be as close to 1 as possible.

3. Study area and hydrological models

The study area is Chao River basin upstream of the Miyun reservoir with a drainage area of 5300 km², accounting for 40% of the Miyun Reservoir catchment area, which is one of the most important surface water resources for the Beijing water supply (Fig. 1). The hydrological data from 1973 to 1982 are used in the study, which have undergone serious and strict quality control measures in previous studies (Wang, 2005). The models were calibrated against observed discharges at the watershed outlet (Xiaohui station). The mean annual precipitation is 494 mm, of which 80% occurs in the rainy season from June to September. The mean runoff coefficient of the catchment is 0.19. The mean annual temperature is about 11–12 °C in the plain area and 8–11 °C in the hill area of the upstream region. There is rarely little precipitation in the winter season in the basin and hence snow is rare.

3.1. WASMOD

To quantify the effects of threshold values and the number of sample simulations on the GLUE simulation results, and compare them with the Bayesian method, a well-tested simple conceptual water balance model, WASMOD was used. The input data to the model include precipitation, potential evapotranspiration, while the output data include fast flow, slow flow, actual evapotranspiration, and soil moisture. The model works well for different spatial and time scales (Xu, 2002; Widen-Nilsson et al., 2007; Gong et al., 2009; Widen-Nilsson et al., 2009; Jin et al., 2010), and the monthly version of the model on a catchment scale was used in this study. There are three parameters in version of WASMOD used in this study. And the main equations are shown in Table 1.

3.2. DTVGM

For comparison purpose, the monthly distributed time-variant gain model (DTVGM), which was developed based on the time-variant gain model (TVGM) (Xia et al., 1997), was also used in the study. The model has been successfully applied in the Chaobai River basin and Heihe River basin (Xia et al., 2005; Wang, 2005). The input data of the model are monthly precipitation, potential evapotranspiration, air temperature, DEM, and land use. The model outputs are monthly river flow and other water cycle components,

such as actual evapotranspiration, and soil-moisture storage. There are four parameters in the version of DTVGM used in this study. And the primary equations of the model are presented in Table 2.

4. Results

There are 5 chains in the Bayesian method to simulate in parallel and each of them has 5000 iterations. The accept rate tuning was 23%. In order to get rid of the influence of “burn in” phase, the beginning 1000 iterations were removed. Then the total samples of each parameter were 20,000. The Scale Reduction Score \sqrt{R} of all the parameters was nearly 1.0, which means the chains had converged.

4.1. Sensitivity of GLUE simulations to the choice of threshold values

Previous studies have shown that the choice of threshold values for the likelihood measures is particularly important for the GLUE method. In order to quantify the effect of threshold values on model simulation results, a series of threshold values have to be investigated. A threshold is either defined in terms of a certain allowable deviation of the highest likelihood value in the sample, or sometimes as a fixed percentage of the total number of simulations (Blasone et al., 2008a). The fixed percentage method was used in the study which is named as the acceptable sample rate (ASR) of 10%, 8%, 7%, 6%, 5%, 4%, 2.50%, 2%, 1.50%, 1%, 0.50%, 0.40%, 0.25%, 0.20%, and 0.10%. The relationship between threshold values and acceptable sample rate (ASR) are shown in Fig. 2. It is seen from the figure that ASR has a good linear relation with the threshold value for the two hydrological models which have different slopes of the regression lines.

The effects of the acceptable sample rate (ASR) on the two evaluation criteria, i.e., ARIL and P-95CI are shown in Figs. 3 and 4, respectively. It is seen from Fig. 3 that the Average Relative Interval Length (ARIL) increases with the decrease of the threshold value, i.e., the increase of ASR. But ARIL was nearly the same when ASR was <2% for the two hydrological models. It can be seen that the larger the ASR, i.e., the lower the threshold value, the larger the difference in ARIL between two hydrological models. Fig. 3 also shows that there is an inflexion point in both lines, which means ARIL would increase slowly when ASR is increasing over an inflexion value, which is 1% for ASR for both hydrological models. When ASR is

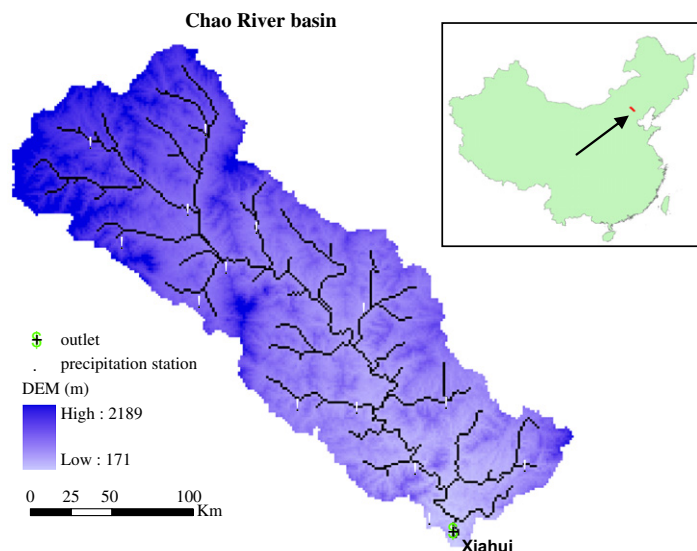


Fig. 1. The location and river system of Chao River Basin in North China with meteorological stations and discharge station.

Table 1

The main equations of WASMOD.

Actual evapotranspiration	$e_t = \min\{(sm_{t-1} + p_t)(1 - \exp(-a1 \cdot ep_t)), ep_t\}$	$0 \leq a1 \leq 1$
Slow flow	$s_t = a2 \cdot (sm_{t-1})^2$	$0 \leq a2 \leq 1$
Fast flow	$f_t = a3 \cdot (sm_{t-1})(p_t - ep_t(1 - \exp(-p_t / \max(ep_t, 1))))$	$0 \leq a3 \leq 1$
Total flow	$d_t = s_t + f_t$	
Water balance	$sm_t = sm_{t-1} + p_t - e_t - d_t$	

In which, t is t th month; p_t is precipitation, ep_t is potential evaporation, sm_t is soil moisture at the end of t th monthly, and $a1$, $a2$ and $a3$ are model parameters.

Table 2

Principal equations for the DTVGM model.

Snowmelt	$S_M = MF \cdot (T_{av} - T_{mlt})$	$0 \leq MF$
Actual evapotranspiration	$ET_a = [(1 - KAW) \cdot f(P/ET_p) + KAW \cdot AW/WM] \cdot ET_p$	$0 \leq KAW \leq 1 < 1$
Surface runoff	$\begin{cases} Rs = g1 \cdot (AW/WM)^{g2} \cdot P & AW \geq WM_i \\ Rs = g1 \cdot (WM_i/WM)^{g2} \cdot P & AW < WM_i \end{cases}$	$0 \leq g1 \leq 1$ $0 \leq g2$
Subsurface runoff	$Rg = Kr \cdot AW \cdot ThinkU$	$0 \leq Kr \leq 1$
Total runoff	$R = Rg + Rs$	

In which, MF is snowmelt rate ($\text{mm}^\circ\text{C}^{-1} \text{ month}^{-1}$); T_{mlt} is the Snowmelt temperature ($^\circ\text{C}$); T_{av} is mean monthly air temperature ($^\circ\text{C}$); P is precipitation (mm); $g1$ and $g2$ are coefficients of time-variant gain factor, related to surface runoff generation; AW is soil moisture content (%); WM is saturated soil moisture content (%); WM_i is minimum soil moisture content (%); KAW is coefficient for calculating actual evapotranspiration; ET_p is potential evapotranspiration (mm month^{-1}), EP_0 is actual evapotranspiration (mm month^{-1}); $ThinkU$ is soil depth (mm); Kr is the storage-outflow coefficient related to subsurface runoff generation (month^{-1}).

Table 3

The prior distributions of all parameters in two hydrological models.

Hydrological models	WASMOD				
Parameters	$a1$	$a2$	$a3$	a	σ^b
Prior distributions ^a	$U[0, 1]$	$U[0, 1]$	$U[0, 1]$	$U[0, 1]$	$\propto \sigma^{-1}$
Hydrological models	DTVGM				
Parameters	KAW	$g1$	$g2$	Kr	a
Prior distributions ^a	$U[0, 1]$	$U[0, 1]$	$U[0, 10]$	$U[0, 1]$	$\propto \sigma^{-1}$

^a $U[a, b]$ means the prior distribution of the parameter is uniform over the interval $[a, b]$.

^b $\propto \sigma^{-1}$ means Jeffreys' uninformative prior of parameter σ is proportional to $1/\sigma$.

bigger than the inflexion value, ARIL is less sensitive to the change in ASR. It is also seen that as ASR increases continuously, the rates of ARIL increase are different for the two hydrological models. There is even nearly no increase of ARIL for WASMOD when ASR is larger than 1%. It means that the 95% confidence interval of discharge in WASMOD is narrower than that in DTVGM. Fig. 4 shows that the relationship of P-95CI and ASR also has an inflexion value. When ASR increases to more than 1%, the increasing speed of P-

95CI obviously reduces for both hydrological models. This is because P-95CI is also impacted by other factors, such as errors in input data, the error of output data and so on, such that the 95% confidence interval of discharge cannot increase to close 95% or 100% by increasing ASR alone. From Fig. 4 it can be seen that P-95CI is different for the two hydrological models, as in the case for ARIL. It is illustrated (Fig. 5) that ARIL and P-95CI have a propor-

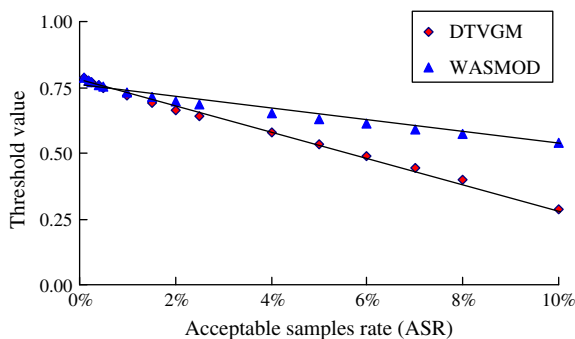


Fig. 2. The relation of threshold value and acceptable sample rate (ASR) for WASMOD and DTVGM.

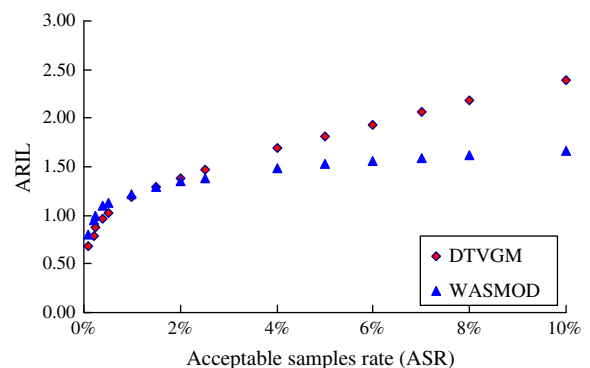


Fig. 3. The relation of ARIL and ASR for WASMOD and DTVGM.

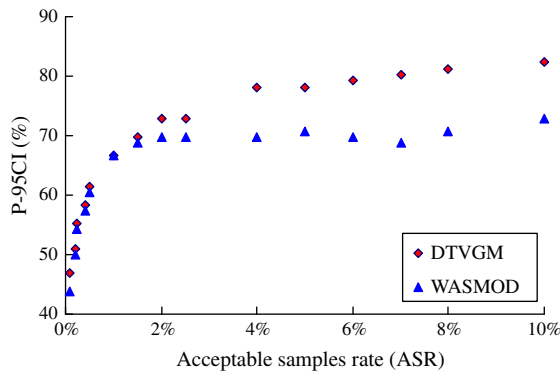


Fig. 4. The relation of P-95CI and ASR for WASMOD and DTVGM.

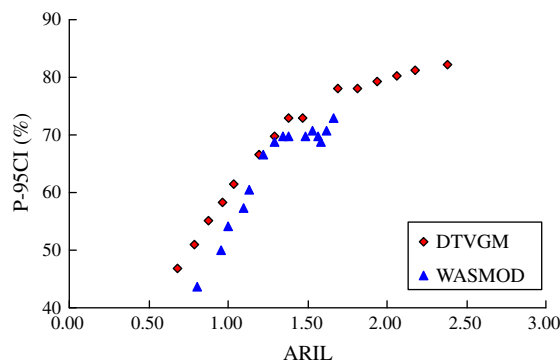


Fig. 5. The relation of ARIL and P-95CI for WASMOD and DTVGM.

tional relationship for the two hydrological models. This is because when ARIL is smaller the opportunity of the observed discharge points falling within the 95% confidence interval is also smaller. Fig. 5 also shows that there is an inflexion point in the regression lines at about $ARIL = 1.4$ for both models. Before the inflexion point, P-95CI increases faster as ARIL increases than that after the inflexion point. At the end of the line, P-95CI nearly has no change with the increase in ARIL for both hydrological models. The explanation is that the number of acceptable samples is the key factor that influences P-95CI and ARIL when ASR is quite small. As ASR increases, i.e., the threshold value decreases, the acceptable samples increases, which causes a faster increase for ARIL than that for P-95CI. Finally, P-95CI will have no change with the increase in ARIL and ASR. In this case, other factors that are not considered in the analysis might have different impacts on P-95CI and on ARIL. From Fig. 5, one can also see that there is a remarkable difference between the two hydrological models. With the increase of ASR, P-95CI and ARIL of WASMOD are smaller than those of DTVGM.

4.2. Sensitivity of GLUE simulations to number of sample simulations

Considering ASR to be 1% and changing the samples size to: 800,000, 700,000, 600,000, 500,000, 400,000, 200,000, 100,000, 80,000, 50,000, 20,000, and 10,000, relationships between number of sample simulations and ARIL, P-95CI and MNS were analysed. Fig. 6 shows that there is nearly no change in the threshold value when the number of sample simulations increased from 1×10^4 to 80×10^4 . It means that the threshold value will be nearly stable with the increase in number of sample simulations when ASR is fixed at 1%. Fig. 7 shows the relationship between maximum Nash–Sutcliffe (MNS) value and samples size when ASR is 1%. We can see in the beginning MNS increases as number of sample sim-

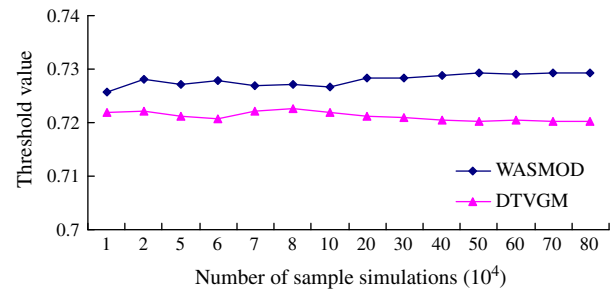


Fig. 6. The relation of threshold value and number of sample simulations for WASMOD and DTVGM when the ASR is 1%.

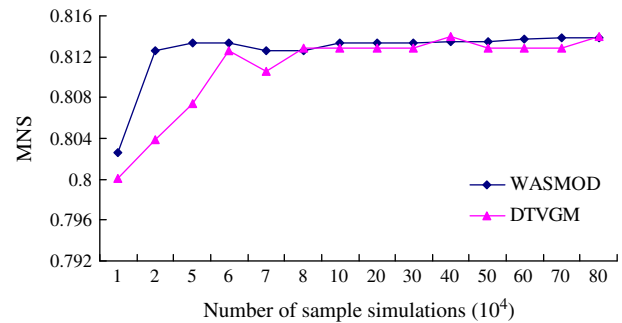


Fig. 7. The relation of maximum Nash–Sutcliffe (MNS) value and number of sample simulations for WASMOD and DTVGM when the ASR is 1%.

ulations increases. For WASMOD the MNS value keeps unchanged after number of sample simulations equals or is larger than 2×10^4 , which can be considered as a minimum and optimum number of sample simulations. In the same case the minimum and optimum number of sample simulations for DTVGM was found to be 8×10^4 . It demonstrates that when number of sample simulations is not big enough it will impact the performance of the simulation models by GLUE method. The minimum number of sample simulations is different for different hydrological models. Furthermore, it can be seen that MNS for two hydrological models are nearly the same when number of sample simulations is big enough. Fig. 7 shows that the best performances of the two models are nearly the same, however the uncertainty ranges as measured by ARIL and the percent of observations bracketed by 95CI (P-95CI) are different for different models (Figs. 3–5).

4.3. Parameter uncertainty

For illustrative purposes, the marginal posterior parameter distributions derived by the Bayesian method and the GLUE method for the WASMOD are shown in Fig. 8. It can be seen that (1) the marginal posterior parameter distributions derived by the Bayesian method and the GLUE method are different. The smaller the ASR (i.e., the larger the threshold value) is, the similar the results of GLUE and Bayesian are, (2) the posterior distributions derived by the Bayesian method are narrower and sharper than those obtained by the GLUE method, which means less uncertainty in parameters, and (3) the posterior parameter distributions are different for different ASR values in the GLUE method. The larger the ASR (i.e., the lower the threshold value) is, the flatter the shape of parameter posterior distribution is. This can also be confirmed by the variances of parameter samples given in Table 4, which reflects the statistical characteristics of the posterior distribution of each parameter derived from the Bayesian and GLUE method with different threshold values. The number of sample simulations in

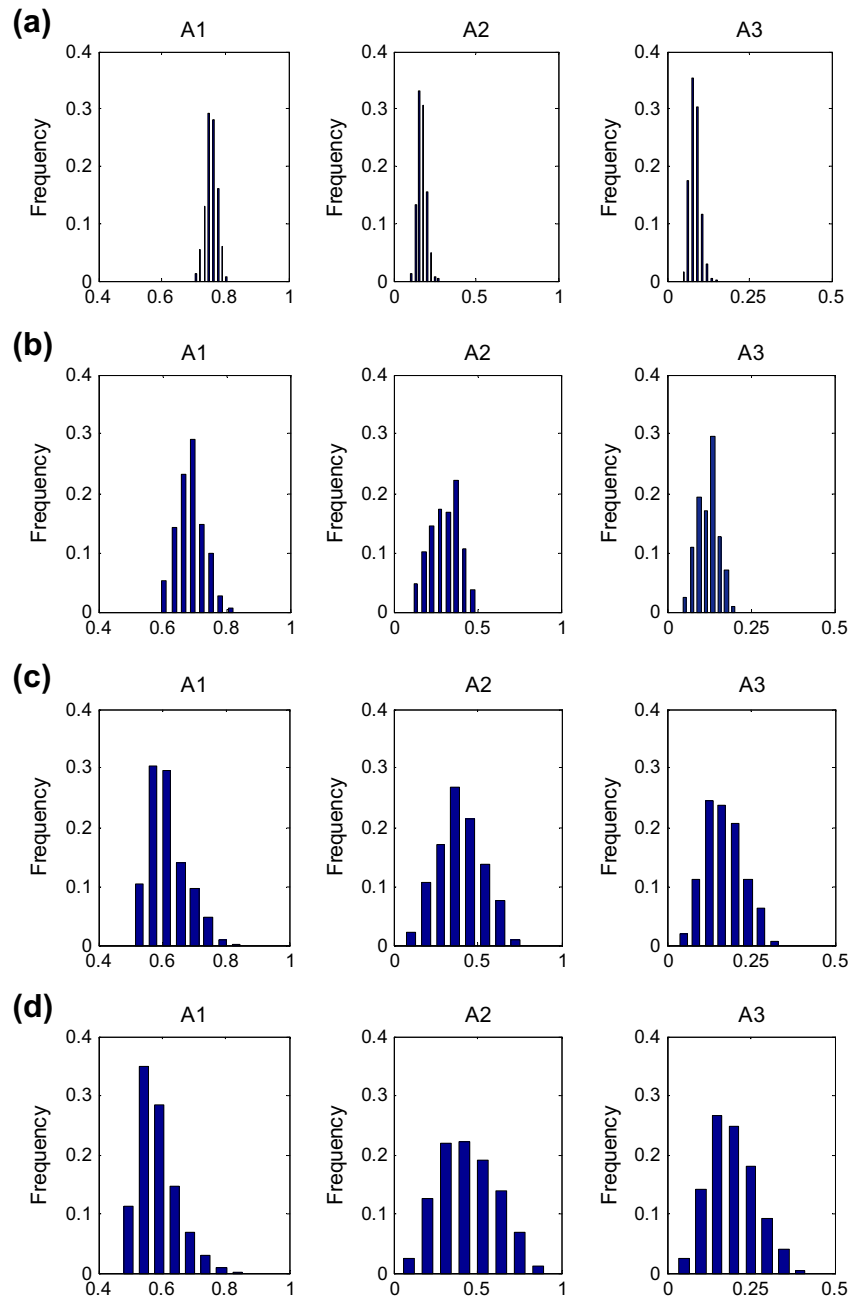


Fig. 8. Histograms of parameters of WASMOD derived by (a) Metropolis Hasting method using the AR (1) Gaussian error method; (b) GLUE method whose number of sample simulations is 200,000 and ASR is 0.2%; (c) GLUE method whose number of sample simulations is 200,000 and ASR is 1%; and (d) GLUE method whose number of sample simulations is 200,000 and ASR is 2%.

GLUE method is 200,000. From Table 4, one can see that (1) the variances and effective parameter space obtained by the Bayesian method are the smallest compared with those obtained by the GLUE method with different ASR values and (2) the variances and effective parameter spaces are increasing with the increase of ASR (i.e., decrease of threshold values). When ASR is chosen to be the smallest value of 0.1% (i.e., the highest threshold value), the statistical characteristics of parameters' posterior distribution obtained by the GLUE method are closer to those obtained by the Bayesian method. However, the shapes of the parameter posterior distributions obtained by the GLUE method are not as good as those obtained by the Bayesian method, which is sharper, more symmetrical and easier to identify the parameters.

Fig. 9 shows the 95% confidence intervals of monthly discharge (1976/1–1983/12) due to parameter uncertainty and observed discharge for the two models. From this figure it can be seen that the 95% confidence interval of discharge due to parameter uncertainty of the Bayesian method is somehow narrower than that of the GLUE method for both models. Fig. 9b and d also show that there is only about 25% and 18.75% of observed discharge values are inside the 95% confidence interval in WASMOD and DTVGM, respectively, which means that the other sources of uncertainty are perhaps more important than parameter uncertainty alone.

Fig. 10 shows the 95% confidence interval due to parameter and model structure uncertainty obtained by the Bayesian method for the two hydrological models. Comparing Fig. 10 with Fig. 9, it is

Table 4

The statistical characteristics of parameters for WASMOD estimated by Bayesian and GLUE methods. The smaller the ASR (i.e., the higher the threshold value) is, the closer the GLUE to Bayesian estimation is.

Method	Parameter	5th-Percentile	95th-Percentile	Median	Variance
Bayesian	a1	0.724	0.786	0.756	0.0003
	a2	0.132	0.218	0.169	0.0007
	a3	0.063	0.110	0.084	0.0002
GLUE with 0.1% ASR	a1	0.603	0.784	0.705	0.0020
	a2	0.141	0.480	0.278	0.0057
	a3	0.063	0.198	0.112	0.0009
GLUE with 0.2% ASR	a1	0.603	0.784	0.691	0.0030
	a2	0.141	0.480	0.305	0.0110
	a3	0.063	0.198	0.123	0.0020
GLUE with 1% ASR	a1	0.530	0.772	0.621	0.0060
	a2	0.140	0.666	0.390	0.0260
	a3	0.067	0.288	0.168	0.0050
GLUE with 2% ASR	a1	0.494	0.762	0.591	0.0070
	a2	0.134	0.792	0.441	0.0410
	a3	0.069	0.352	0.195	0.0070

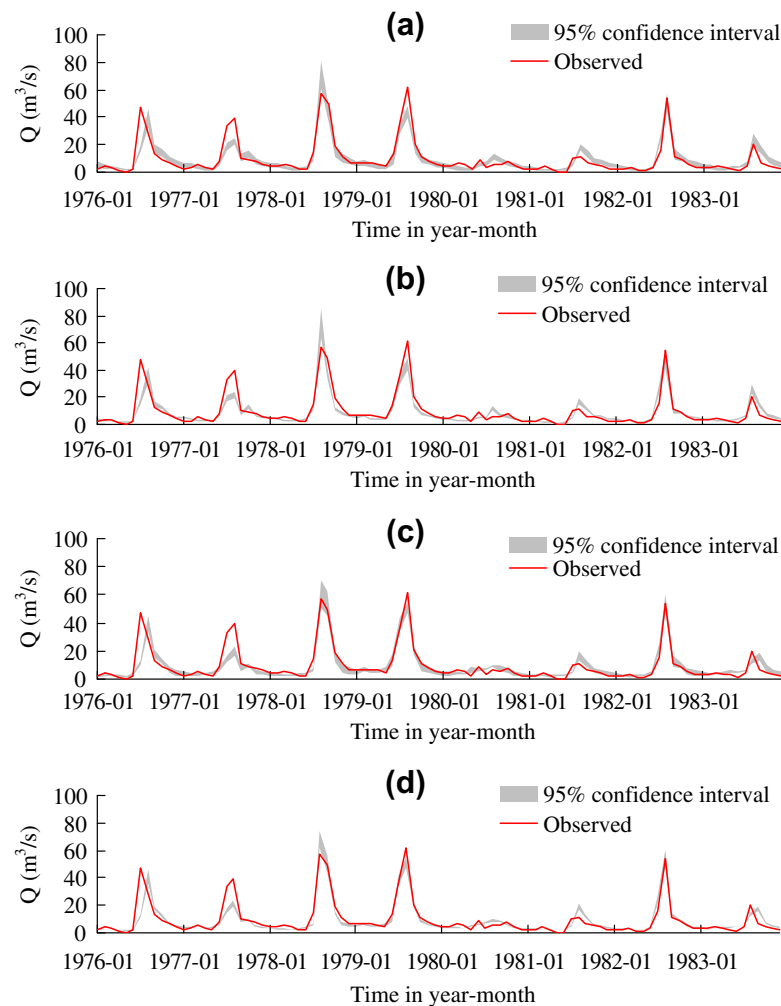


Fig. 9. The 95% confidence intervals of monthly discharge (1976/1–1983/12) due to parameter uncertainty (grey bands) and observed discharge (red solid line). (a) GLUE of WASMOD when number of sample simulations is 200,000 and ASR is 0.1%; (b) Bayesian of WASMOD; (c) GLUE of DTVGM when number of sample simulations is 200,000 and ASR is 0.1%; and (d) Bayesian of DTVGM. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

seen that the uncertainty in discharge simulation due to model structure is more important than that due to parameters.

Detailed comparison of uncertainty measures between the GLUE and the Bayesian methods for the two hydrological models

is shown in Table 5. The table clearly shows that: (1) With ASR = 10%, in WASMOD, 79.167% of observed data are captured by the 95% confidence interval of monthly simulated discharge due to parameter and model uncertainty by the Bayesian method,

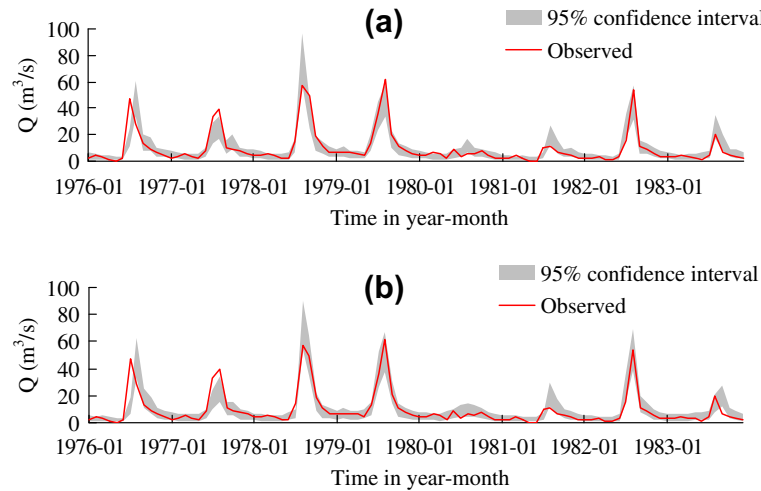


Fig. 10. The 95% confidence intervals of monthly discharge (1976/1–1983/12) due to parameter uncertainty and model uncertainty (grey bands) and observed discharge (red solid line). (a) Bayesian of WASMOD (P-95CI is 79%); (b) Bayesian of DTVGM (P-95CI is 82%). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5

The comparison of uncertainty measures between GLUE and Bayesian methods for two hydrological models.

Methods	DTVGM				WASMOD			
	Threshold value	ARIL	P-95CI (%)	MNS	Threshold value	ARIL	P-95CI (%)	MNS
<i>Bayesian_P</i>	–	0.290	18.75	0.814	–	0.343	25.0	0.814
<i>Bayesian_PM</i>	–	1.685	82.292	0.814	–	1.427	79.167	0.814
<i>GLUE₁₀</i>	0.287	2.385	82.292	0.814	0.541	1.663	72.917	0.814
<i>GLUE₈</i>	0.398	2.179	81.250	0.814	0.574	1.620	70.833	0.814
<i>GLUE₇</i>	0.445	2.058	80.208	0.814	0.592	1.588	68.750	0.814
<i>GLUE₆</i>	0.491	1.936	79.167	0.814	0.612	1.564	69.792	0.814
<i>GLUE₅</i>	0.534	1.815	78.125	0.814	0.631	1.531	70.833	0.814
<i>GLUE₄</i>	0.577	1.690	78.125	0.814	0.651	1.491	69.792	0.814
<i>GLUE_{2.5}</i>	0.643	1.469	72.917	0.814	0.685	1.379	69.792	0.814
<i>GLUE₂</i>	0.666	1.382	72.917	0.814	0.699	1.348	69.792	0.814
<i>GLUE_{1.5}</i>	0.691	1.289	69.792	0.814	0.714	1.294	68.750	0.814
<i>GLUE₁</i>	0.719	1.192	66.667	0.814	0.731	1.223	66.667	0.814
<i>GLUE_{0.5}</i>	0.748	1.032	61.458	0.814	0.754	1.133	60.417	0.814
<i>GLUE_{0.4}</i>	0.756	0.961	58.333	0.814	0.761	1.101	57.292	0.814
<i>GLUE_{0.25}</i>	0.769	0.877	55.208	0.814	0.772	0.996	54.167	0.814
<i>GLUE_{0.2}</i>	0.774	0.789	51.042	0.814	0.776	0.953	50.000	0.814
<i>GLUE_{0.1}</i>	0.786	0.681	46.875	0.814	0.788	0.804	43.750	0.814

GLUE_α ($\alpha = 0.1, 0.2, \dots, 10$) are for the GLUE method with ASR is equal to 0.1, 0.2, ..., 10%;

Bayesian_P represents the 95% confidence interval is only due to parameter uncertainty;

Bayesian_PM represents the 95% confidence interval is due to parameter and model uncertainty.

while the same value by the GLUE method is 72.917%. In DTVGM, the maximum value of P-95CI derived from the GLUE method is 82.292%, which is the same as that derived from the Bayesian method. However, at the same time, ARIL derived from the GLUE method is 2.385 which is much wider than 1.685 derived from the Bayesian method. (2) ARIL and P-95CI decrease as ASR decreases (i.e., the threshold value increases) in the GLUE method for both models. This means it is difficult to compare the results of the Bayesian method with the GLUE method which depends on the choice of the threshold value. (3) It also shows that both ARIL and P-95CI become much wider and larger when both parameter and model uncertainties are included as compared with that when only parameter uncertainty is concerned. (4) The best MNS values are the same (MNS = 0.814) in both the Bayesian method and the GLUE method even with different ASR values.

4.4. Fitting of hydrological models

Fig. 11 compares monthly observed and simulated discharges based on two sets of parameters according to the maximum pos-

terior density by the Bayesian method and the maximum likelihood value by GLUE for two hydrological models. The figure reveals that both the GLUE and the Bayesian methods work equally well to simulate the best fit of the discharge for both models. The maximum Nash–Sutcliffe (MNS) efficiency values obtained by the GLUE and the Bayesian methods are the same, and the simulated discharges are nearly identical.

5. Discussion and conclusion

In recent decades, a lively debate has arisen in the hydrologic literature over the GLUE and Bayesian methods concerning their theory and results. Different studies have resulted in different conclusions. The subjective choice of the threshold values in GLUE and the influence of the sampling size on the results of the GLUE method have been criticized in the literature. In this study, a systematic evaluation and quantification of the effect of threshold values and sampling size in GLUE on the parameter and model uncertainties was performed, and the results were compared with the formal Bayesian approach using the Metropolis Hasting (MH) algorithm

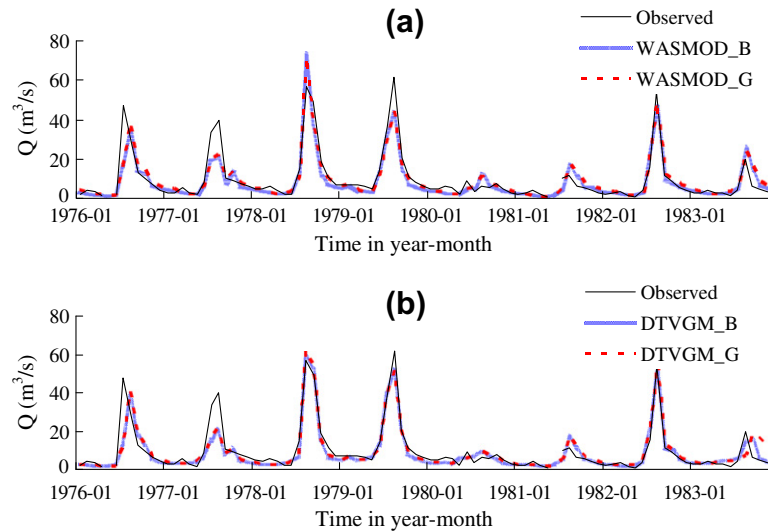


Fig. 11. The comparison of monthly observed and simulated discharges (1976/1–1983/12) based on two sets of parameters according to maximum posterior density from GLUE (_G) and Bayesian method(_B): (a) by WASMOD and (b) by DTVGM.

for two well-tested conceptual hydrological models (WASMOD and DTVGM). The following conclusions are drawn from this study:

1. In the GLUE method, the posterior distribution of parameters and the 95% confidence interval of the simulated discharge are sensitive to the choice of the threshold value as measured by the acceptable samples rate (ASR). The lower the ASR (i.e., the higher the threshold value) the narrower the 95% confidence interval. However, the change in 95% confidence interval and the percent of observations bracketed by the 95% confidence interval (P-95CI) become much slower when ASR is bigger than 1%. There is an inflexion value of ASR at about 1–2%, after which the effect of ASR is minor. It seems that the ASR value has no or very minor influence on the maximum Nash–Sutcliffe (MNS) efficiency value.
2. In the GLUE method, the insufficiency of number of sample simulations influences the maximum Nash–Sutcliffe (MNS) value when ASR is fixed. However, as soon as the number of sample simulations increases to 2×10^4 for WASMOD and to 8×10^4 for DTVGM model the influence of number of sample simulations on the model simulation results becomes of minor importance.
3. The posterior distributions of parameters derived by the Bayesian method are narrower and sharper than those obtained by the GLUE method. However, when the threshold value in the GLUE method is high enough (i.e., when the ASR value is smaller than 0.1%), the posterior distribution of parameters, the 95% confidence interval of simulated discharge and the percent of observations bracketed by the 95% confidence interval (P-95CI) for the GLUE method are approaching those values estimated by the Bayesian method for both hydrological models.
4. For both the Bayesian method and GLUE, the uncertainty in the simulated discharges resulting from parameter uncertainty is much smaller than that resulting from model structure uncertainty. When only parameter uncertainty is a concern, 18.75% and 25% of the observed discharge are found inside the 95% confidence intervals of the simulated discharges from DTVGM and WASMOD, respectively. The same measures become about 82% and 79% when both parameter and model uncertainties are concerned for the DTVGM model and WASMOD, respectively.

5. The goodness of model fit as measured by the maximum Nash–Sutcliffe efficiency value is nearly the same for the GLUE and Bayesian methods for both hydrological models.

It should be noted that although the study yielded interesting findings, the generality of such findings is still need to be verified with more hydrological models and study regions in future studies.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (40730632), the Knowledge Innovation Key Project of the Chinese Academy of Sciences (Kzcx2-yw-126), and the småforsk project of the Norwegian Research Council. The authors would like to thank Geir Olve Storvik, Paolo Reggiani and Stian Solvang Johansen for their advice and help to realize the Bayesian method in this study.

References

- Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. Wiley, Chichester.
- Beven, K.J., Binley, A., 1992. Future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* 6 (3), 279–298.
- Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *Journal of Hydrology* 249 (1–4), 11–29.
- Beven, K.J., 2006. A manifesto for the equifinality thesis. *Journal of Hydrology* 320 (1–2), 18–36.
- Beven, K.J., Smith, P., Freer, J., 2007. Comment on “Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology” by Pietro Mantovan and Ezio Todini. *Journal of Hydrology* 338 (3–4), 315–318.
- Beven, K.J., Smith, P.J., Freer, J.E., 2008. So just why would a modeller choose to be incoherent? *Journal of Hydrology* 354 (1–2), 18–32.
- Blasone, R.S., Madsen, H., Rosbjerg, D., 2008a. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *Journal of Hydrology* 353 (1–2), 18–32.
- Blasone, R.S., Vrugt, J.A., Madsen, H., Rosbjerg, D., Robinson, B.A., Zyvoloski, G.A., 2008b. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Advances in Water Resources* 31 (4), 630–648.
- Carlos, M.J., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6 (3), 255–259.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. *American Statistician* 49 (4), 327–335.
- Engeland, K., Gottschalk, L., 2002. Bayesian estimation of parameters in a regional hydrological model. *Hydrology and Earth System Sciences* 6 (5), 883–898.

- Engeland, K., Renard, B., Steinsland, I., Kolberg, S., 2010. Evaluation of statistical models for forecast errors from the HBV model. *Journal of Hydrology* 384 (1–2), 142–155.
- Engeland, K., Xu, C.Y., Gottschalk, L., 2005. Assessing uncertainties in a conceptual water balance model using Bayesian methodology. *Hydrological Sciences Journal* 50 (1), 45–63.
- Freer, J., Beven, K., Ambrose, B., 1996. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resources Research* 32 (7), 2161–2174.
- Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472.
- Gong, L., Widen-Nilsson, E., Halldin, S., Xu, C.Y., 2009. Large-scale runoff routing with an aggregated network-response function. *Journal of Hydrology* 368 (1–4), 237–250.
- Gupta, H., Thieman, M., Trosset, M., Sorooshian, S., 2003. Reply to comment by K. Beven and P. Young on “Bayesian recursive parameter estimation for hydrologic models”. *Water Resources Research* 39, 1117. doi:10.1029/2002WR001405.
- Jin, X., Xu, C.-Y., Zhang, Q., Singh, V.P., 2010. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *Journal of Hydrology* 383 (3–4), 147–155.
- Krzysztofowicz, R., 1997. Transformation and normalization of variates with specified distributions. *Journal of Hydrology* 197 (1–4), 286–292.
- Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resources Research* 35, 9. doi:10.1029/1999WR900099.
- Krzysztofowicz, R., Kelly, K.S., 2000. Hydrologic uncertainty processor for probabilistic river stage forecasting. *Water Resources Research* 36, 11. doi:10.1029/2000WR900108.
- Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. *Journal of Hydrology* 211 (1–4), 69–85.
- Li, L., Xia, J., Xu, C.-Y., Chu, J., Wang, R., 2009. Analyse the sources of equifinality in hydrological model using GLUE methodology. In: *Hydroinformatics in Hydrology, Hydrogeology and Water Resources*, Proceedings of Symposium JS.4 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009. IAHS Publ. no. 331, pp. 130–138.
- Liu, Z., Martina, M.L.V., Todini, E., 2005. Flood forecasting using a fully distributed model: application of the TOPKAPI model to the Upper Xixian Catchment. *Hydrology and Earth System Sciences* 9 (4), 347–364.
- Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *Journal of Hydrology* 330 (1–2), 368–381.
- Maranzano, C.J., Krzysztofowicz, R., 2004. Identification of likelihood and prior dependence structures for hydrologic uncertainty processor. *Journal of Hydrology* 290 (1–2), 1–21.
- Montanari, A., 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall–runoff simulations. *Water Resources Research* 41, W08406. doi:10.1029/2004WR003826.
- Montanari, A., Brath, A., 2004. A stochastic approach for assessing the uncertainty of rainfall–runoff simulations. *Water Resources Research* 40, W01106. doi:10.1029/2003WR002540.
- Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models – Part 1-a. Discussion of principles. *Journal of Hydrology* 10 (3), 282–290.
- Reggiani, P., Weerts, A.H., 2008. A Bayesian approach to decision-making under uncertainty: an application to real-time forecasting in the river Rhine. *Journal of Hydrology* 356 (1–2), 56–69.
- Stedinger, J.R., Vogel, R.M., Lee, S.U., Batchelder, R., 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research* 44, W00806. doi:10.1029/2008WR006822.
- Thieman, M., Trosset, M., Gupta, H., Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models. *Water Resources Research* 37, 10. doi:10.1029/2000WR900405.
- Todini, E., 2007. Hydrological catchment modelling: past, present and future. *Hydrology and Earth System Sciences* 11 (1), 468–482.
- Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. *International Journal of River Basin Management* 6 (2), 123–138.
- Van der Waerden, B.L., 1952. Order tests for the two-sample problem and their power. *Indagationes Mathematicae* 14 (253), 453–458.
- Van der Waerden, B.L., 1953a. Order tests for the two-sample problem and their power 2. *Indagationes Mathematicae* 15, 303–310.
- Van der Waerden, B.L., 1953b. Order tests for the two-sample problem and their power 3. *Indagationes Mathematicae* 15, 311–316.
- Vogel, R.M., Stedinger, J.R., Batchelder, R., Lee, S.U., 2008. Appraisal of the Generalized Likelihood Uncertainty Estimation (GLUE) Method. American Society of Civil Engineers, 1801 Alexander Bell Drive, Reston, VA, 20191-4400, USA.
- Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* 39, 8. doi:10.1029/2002WR001642.
- Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2009. Equifinality of formal (DREAM) and informal (GLUE) bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment* 23 (7), 1011–1026.
- Wang, G., 2005. Theory and Method of Distributed Time-Variant Gain Model. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing. Ph.D. (in Chinese).
- Wang, G., Xia, J., Chen, J., 2009. Quantification of effects of climate variations and human activities on runoff by a monthly water balance model: a case study of the Chaobai River basin in northern China. *Water Resources Research* 45, W00A11. doi:10.1029/2007WR006768.
- Widen-Nilsson, E., Gong, L., Halldin, S., Xu, C.-Y., 2009. Model performance and parameter behavior for varying time aggregations and evaluation criteria in the WASMOD-M global water balance model. *Water Resources Research* 45, W05418. doi:10.1029/2007WR006695.
- Widen-Nilsson, E., Halldin, S., Xu, C.-Y., 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation. *Journal of Hydrology* 340 (1–2), 105–118.
- Xia, J., O'Connor, K.M., Kachroo, R.K., Liang, G.C., 1997. A non-linear perturbation model considering catchment wetness and its application in river flow forecasting. *Journal of Hydrology* 200 (1–4), 164–178.
- Xia, J., Wang, G., Tan, G., Ye, A., Huang, G.H., 2005. Development of distributed time-variant gain model for nonlinear hydrological systems. *Science in China Series D: Earth Sciences* 48 (6), 713–723.
- Xiong, L., O'Connor, K.M., 2008. An empirical method to improve the prediction limits of the GLUE methodology in rainfall–runoff modeling. *Journal of Hydrology* 349 (1–2), 115–124.
- Xu, C.-Y., 2002. WASMOD – the water and snow balance modelling system. In: Singh, V.P., Frevert, D.K. (Eds.), *Mathematical Models of Small Watershed Hydrology and Applications*. Water Resources Publications, LLC, Chelsea, Michigan, USA, pp. 555–590.
- Yang, J., Reichert, P., Abbaspour, K.C., Xia, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe Basin in China. *Journal of Hydrology* 358 (1–2), 1–23.
- Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007. Hydrological modelling of the Chaohe Basin in China: statistical model formulation and Bayesian inference. *Journal of Hydrology* 340 (3–4), 167–182.