Journal of Hydrology 383 (2010) 147–155

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model

Xiaoli Jin [a,*], Chong-Yu Xu [b], Qi Zhang [c], V.P. Singh [d,e]

[a] Institute of Mountain Hazards and Environment, Chinese Academy of Sciences, Chengdu, Sichuan 610041, China
[b] Department of Geosciences, University of Oslo, Norway
[c] State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, Jiangsu 210008, China
[d] Department of Biological and Agricultural Engineering, Texas A & M University, Scoates Hall, 2117 TAMU, USA
[e] Department of Civil & Environmental Engineering, Texas A & M University, Scoates Hall, 2117 TAMU, USA

ARTICLE INFO

Article history:
Received 1 April 2009
Received in revised form 19 November 2009
Accepted 17 December 2009

This manuscript was handled by L. Charlet, Editor-in-Chief, with the assistance of Jose D. Salas, Associate Editor

Keywords:
Conceptual hydrological model
Uncertainty
Bayesian method
GLUE

SUMMARY

Quantification of uncertainty of hydrological models has attracted much attention in hydrologic research in recent years. Many methods for quantification of uncertainty have been reported in the literature, of which GLUE and formal Bayesian method are the two most popular methods. There have been many discussions in the literature concerning differences between these two methods in theory (mathematics) and results, and this paper focuses on the computational efficiency and differences in their results, but not on philosophies and mathematical rigor that both methods rely on. By assessing parameter and modeling uncertainty of a simple conceptual water balance model (WASMOD) with the use of GLUE and formal Bayesian method, the paper evaluates differences in the results of the two methods and discusses the reasons for these differences. The main findings of the study are that: (1) the parameter posterior distributions generated by the Bayesian method are slightly less scattered than those by the GLUE method; (2) using a higher threshold value (>0.8) GLUE results in very similar estimates of parameter and model uncertainty as does the Bayesian method; and (3) GLUE is sensitive to the threshold value used to select behavioral parameter sets and lower threshold values resulting in a wider uncertainty interval of the posterior distribution of parameters, and a wider confidence interval of model uncertainty. More study is needed to generalize the findings of the present study.

Introduction

Conceptual hydrological models are popular tools for simulating the land phase of the hydrological cycle. They are frequently used for water balance analysis, extending and infilling streamflow records, flow forecasting, reservoir operation, water supply, and watershed management. One distinguishing characteristic of a conceptual model is that one or more of its parameters requires calibration using physically observable catchment responses (Kuczera and Parent, 1998). When parameter calibration is employed, it is easy to show that multiple calibration periods yield multiple optimum parameter sets, and even in a single period, different sets of optimum parameter values may yield similar model performances; this is termed as "equifinality" in the literature. On the other hand, for the same input and output data, different models, with similar calibration results, may produce largely different predictions, as observed by Jiang et al. (2007) in a comparison of hydrological impacts of climate change simulated by six hydrological models. Since a conceptual model can be viewed as an empirical combination of mathematical operators describing the main features of an idealized hydrologic cycle, one cannot rely on a uniquely determined model parameter set or model prediction (Kuczera and Parent, 1998). Consequently, attention should be paid to the uncertainties in hydrological modeling.

Generally speaking, there are three principal sources contributing to modeling uncertainty: errors associated with input data and data for calibration, imperfection in model structure, and uncertainty in model parameters (e.g., Refsgaard and Storm, 1996). Xu et al. (2006) demonstrated that the quality of precipitation data influences both simulation errors and calibrated model parameters. Engeland et al. (2005) showed that the effect of the model structural uncertainty on the total simulation uncertainty of a conceptual water balance model was larger than parameter uncertainty. Marshall et al. (2007) stated that the uncertainty in model structure requires developing alternatives, where outputs from multiple models are pooled together in order to generate an ensemble of hydrographs that are able to represent uncertainty. Kavetski et al. (2002), and Chowdhury and Sharma (2007) investigated input data uncertainty by artificially adding noise to input

* Corresponding author. Tel.: +47 22 855825; fax: +47 22 854215.
E-mail address: yezi1612@hotmail.com (X. Jin).

0022-1694/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2009.12.028

data and then formulating an empirical relationship between this noise and parameter error. Many other examples of the methods dealing with model and data uncertainty are available in the hydrological literature (e.g., Georgakakos et al., 2004; Carpenter and Georgakakos, 2004; Kavetski et al., 2006a,b).

A variety of methods have been developed to deal with parameter uncertainty and modeling uncertainty, i.e., to provide posterior distributions for parameters and runoff, which are of greater interest. Among these methods, the generalized likelihood uncertainty estimation (GLUE) method, developed by Beven and Binley (1992), and the formal Bayesian method using Metropolis–Hastings (MH) algorithm, a Markov Chain Monte Carlo (MCMC) methodology, are extensively used (Freer et al., 1996; Beven and Freer, 2001; Kuczera and Parent, 1998; Bates and Campbell, 2001). According to Beven et al. (2000), GLUE "represents an extension of Bayesian or fuzzy averaging procedures to less formal likelihood or fuzzy measures." The concept of "the less formal likelihood", which represents the main element of differentiation with the Bayesian inference, is a remarkable aspect of the GLUE methodology. The flexible definition of the likelihood function allows for no strong assumptions on the error model. Doubtless, when the assumptions are fulfilled, the formal Bayesian method is the most convenient to use, since "the formal Bayesian approach has its roots within classical statistical theory and applies formal mathematics and MCMC simulation to infer parameter and prediction distributions" (Vrugt et al., 2008). At the same time, variants of these methods have also been developed. These include utilizing a new procedure to partially correct the prediction limit of a hydrological model in GLUE (Xiong and O'Connor, 2008), introducing adaptation into the MCMC method (Kuczera and Parent, 1998; Gilks et al., 1998; Yang et al., 2007, 2008), merging one of MCMC sampler with the SCE-UA global optimization algorithm (Vrugt et al., 2003), and revising GLUE based on the MCMC sampling (Blasone et al., 2008a,b), amongst others.

Compared to other methods, GLUE is easy to implement, requiring no modifications to existing source codes of simulation models. Therefore, many users are attracted by GLUE. On the other hand, controversy over GLUE has recently started to increase "for not being formally Bayesian, requiring subjective decisions on the likelihood function and cutoff threshold separating behavioral from non-behavioral models, and for not implementing a statistically consistent error model" (Blasone et al., 2008b). In the same work, Blasone et al. (2008b) pointed out that "... the GLUE derived parameter distribution and uncertainty bounds are entirely subjective and have no clear statistical meaning." Stedinger et al. (2008) argued that although an absolutely correct likelihood function may be difficult to construct, it should not be an excuse to use any function and calls for a likelihood measure that will yield probabilities with any statistical validity. Liu et al. (2005) and Todini (2008) showed that a formal likelihood measure can be derived by first converting a real data set into a normal space.

Another drawback of GLUE is that it is computationally inefficient and can even lead to misleading results, unless a large sample is drawn (Blasone et al., 2008a). Mantovan and Todini (2006) showed the incoherence of GLUE with Bayesian inference using an experiment, where synthetic input and output and a correct formal likelihood function were applied, and they referred it to as pseudo-Bayes. They proved that the "less formal likelihoods" failed to guarantee the requirements of Bayesian inference process that adding more data did not necessarily add information to the conditioning process, and the equivalence of experimental value irrespective of the order. Subsequently, Beven et al. (2007) responded that GLUE was not developed to deal with cases for which every observation added information to the conditioning process. It was developed to deal with real calibration problems in which both inputs and model structural errors played an important role. They further showed that if a correct formal likelihood was used as a prior in GLUE, the results obtained would be identical with the formal Bayes. Stedinger et al. (2008) showed that using a correct likelihood function GLUE can lead to meaningful uncertainty and prediction intervals. Blasone et al. (2008b) also stated that the computational efficiency of GLUE is improved by sampling the prior parameter space using an adaptive MCMC scheme.

It can be noted here that in many cases it is not easy to compare the original GLUE method and the formal Bayesian method directly. First, the formal Bayesian approaches attempt to disentangle the effect of input, output, parameter and model structural error which, on one hand, are important to improve our hydrologic theory of how water flows through watersheds; on the other hand these attempts make statistical inference difficult (Vrugt et al., 2008). GLUE does not attempt to separate these effects on the total uncertainty which, on one hand, makes it easy to use and understand, but on the other hand it is impossible to pinpoint what elements of the model constitute most uncertainty. Second, in practice the formal Bayesian method is often used to calculate the uncertainty interval of one-step ahead forecasting with a formal or exact likelihood function that is assumed or transformed from a more general and unknown form, while the GLUE method is often used to calculate the uncertainty interval of streamflow simulation with a statistically informal likelihood function. In this study, both methods are applied on the same grounds, i.e., our focus is on the computational efficiency and differences in the results of parameter and model uncertainty calculated by the two methods, and not on the different philosophies and mathematical rigor that both methods rely on. Additional analyses are performed to test how sensitive the GLUE results are to the threshold values of retained solutions.

The specific objectives of this paper are: (1) to assess parameter uncertainty of a conceptual hydrological water balance model WASMOD (Xu, 2002) using GLUE and Bayesian method, and (2) to examine the differences in the results of these two methods and discuss the reasons for the differences.

The organization of this paper is as follows. After this brief introduction, two methods are presented, followed by a discussion of the WASMOD model, study area and data used. Then, parameter and modeling uncertainty of the two methods are compared and discussed, and the results and conclusions are presented.

Methods

Formal Bayesian approach

Consider a random variable Y whose value is denoted by y and a parameter vector θ. $P(\theta)$ is a prior distribution of θ based on historical data or expert knowledge, and $L(Y/\theta)$ is the likelihood function based on data collected by observations. The Bayesian theorem provides a formal mechanism for deriving the posterior distribution, $P(\theta/Y)$, of θ based on the prior distribution and likelihood distribution as:

$$P(\theta/Y) = \frac{L(Y/\theta)P(\theta)}{\int L(Y/\theta)P(\theta)d\theta} \propto L(Y/\theta)P(\theta) \quad (1)$$

Based on the investigation by Xu (2001) on the statistical properties of the simulation error for the WASMOD model, a square-root transformation is applied:

$$\sqrt{Q_{obs,t}} = \sqrt{Q_{sim,t}} + \varepsilon_t \quad (2)$$

where $Q_{obs,t}$ is the observed streamflow at time t, $Q_{sim,t}$ is the simulated streamflow at time t, and ε_t is the residual term. It is assumed that this residual term is mutually independent and approximately follows a normal distribution with mean zero and variance σ^2. Then the formal likelihood function $L(Y/\theta)$ can be formulated as:

$$L(Y/\theta) = (2\pi\sigma^2)^{-n/2} \times \exp\left[-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(\sqrt{Q_{obs,t}} - \sqrt{Q_{sim,t}})^2\right] \quad (3)$$

where n is the number of time steps. In most cases, the posterior distribution is not of the standard form, especially in hydrological modeling. For summarizing information, however, samples can be drawn from the posterior distribution using MCMC. Previous studies (Xu, 2001; Engeland et al., 2005) about the model show that the error model proposed in Eq. (3) seems to be reasonable for this model on a monthly time step. For other models and other time resolutions the assumption made here does not necessarily hold and an alternative error model has to be constructed. In such cases, a procedure to construct the likelihood function and posterior density function of Bayesian method is proposed by Todini (2008).

In the Bayesian method the Metropolis Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970), a Markov Chain Monte Carlo (MCMC) methodology was used to estimate parameters. MCMC generates a sample of parameter values from a carefully constructed Markov chain that converges to the posterior distribution. It starts from an initial parameter value, then generates a new value using a proposal distribution, and computes the acceptance probability and determines whether to accept or reject this new value. After removing the initial "burn-in", samples might converge to the posterior distribution from which quantiles are estimated. Details of this algorithm can be found in Kuczera and Parent (1998), Bates and Campbell (2001), Engeland and Gottschalk (2002) and Engeland et al. (2005), among others.

One problem in the application of the Bayesian method is the choice of a proposal probability density function. The variance of the proposal candidate density influences how far from the current state the candidate state is likely to be and its probability of being accepted. If this variance is too small, the iterative process will induce a Markov chain that does not mix rapidly enough over the parameter space. Hence the tail regions of the posterior may not be sampled sufficiently. If the variance is too large, proposal distribution will be frequently rejected (Bates and Campbell, 2001). Chib and Greenberg (1995) stated that the acceptance rate for a new point in the chain being 45% would obtain a reasonable and efficient chain.

Another problem of the method is the diagnosis of convergence of the chain to the posterior distribution. It can be roughly judged by plotting samples to observe, if the mean and variance are stable. Gelman and Rubin (1992) proposed a quantitative measurement, R statistic, which involves both the between-chain variance and the within-chain variance, to monitor the convergence.

Starting from arbitrary values in the parameter space, this study used a normal distribution as proposal distribution to generate five parallel chains. Each chain contained 20,000 iterations and the first 10,000 were considered as burn-in period and removed, then totally 50,000 samples for each parameter were retained. The acceptance rate was between 40% and 50%. Both visual observations and R statistic calculations were performed to check the chains' convergence before estimating uncertainties.

GLUE method

GLUE (Beven and Binley, 1992) has been developed in the context of multiple sources of uncertainty in real problems and an expectation that the structure of the errors will be non-stationary and complex, hence less formal likelihood function is used (Beven, 1993; Beven et al., 2008). The GLUE method includes several steps: (1) Monte Carlo sampling from a feasible parameter space with uniform distribution; (2) specification of the likelihood function and the threshold value for behavioral parameter sets; (3) calculation of likelihood values of behavioral parameter sets; (4) rescaling of these values to formulate a cumulative distribution; and (5) derivation of quantiles of uncertainty from the distribution. In this method, 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 efficiency (ME) was chosen as the likelihood function as in many other studies:

$$ME = 1 - \frac{\sum(Q_{obs} - Q_{sim})}{\sum(Q_{obs} - \bar{Q}_{obs})} = 1 - \frac{\sigma_i^2}{\sigma_{obs}^2} \quad (4)$$

where $\bar{Q}_{obs}$ represents mean values of observed streamflows. σ_i^2 is the error variance for the ith model (i.e., the combination of the model and the ith parameter set) and σ_{obs}^2 is the variance of observations. In total 100,000 samples were drawn and 18,191 parameter sets with ME bigger than 0.8 were retained as behavioral sets accounting for about 18% of the total samples.

Study area and WASMOD model

Shuntian catchment, a sub-basin of the Dongjiang River basin, in southern China, with a drainage area of 1362 km^2 was selected as the study area (see Fig. 1). It has a sub-tropical climate with mean annual precipitation of 1690 mm, mean annual potential evaporation of 1152 mm, and an approximate runoff coefficient of 0.62. The landscape here is characterized by forest and farmlands, accounting for 45.63% and 33.96% of the basin area, respectively.

Fig. 1 also shows stations where data came from. Rainfall data of seven stations were employed to get areal precipitation. Potential evaporation of one station at the outlet was observed and used to represent the areal mean evaporation. Discharges were recorded at the basin outlet. Monthly precipitation, potential evaporation and discharge from January 1978 to December 1988, a total of 132 months, were used in this study. Data quality control was made and the data had been used in earlier studies (Jiang et al., 2007; Jin et al., 2009).

WASMOD, developed by Xu et al. (1996) and Xu (2002), is a monthly water balance model. The input data include monthly areal precipitation, potential evaporation and/or temperature, and outputs are monthly streamflow, actual evapotranspiration, soil moisture and other water balance components. Main model equations are given in Table 1. The earlier version of this model also includes a snow routine. However, because of rare snowfall in this area, the corresponding formulas are excluded from this table.

Results

Parameter uncertainty

For each method, samples of three posterior parameter distributions were produced. To get a visual impression, histograms of parameter samples are plotted (Fig. 2). All parameters possess well-defined and unimodal posterior distributions. From such distributions, parameter estimates can be unambiguously inferred as modal values, while the shape of the distributions indicates the degree of uncertainty of the estimates. Sharp and peaked distributions are associated with well identifiable parameters, while flat distributions indicate more parameter uncertainty. The modal values for GLUE and the Bayesian method are found to be similar from Fig. 2. The posterior distributions obtained by the Bayesian method are somehow sharper and narrower than those obtained by the GLUE method, indicating a slightly better identified parameters

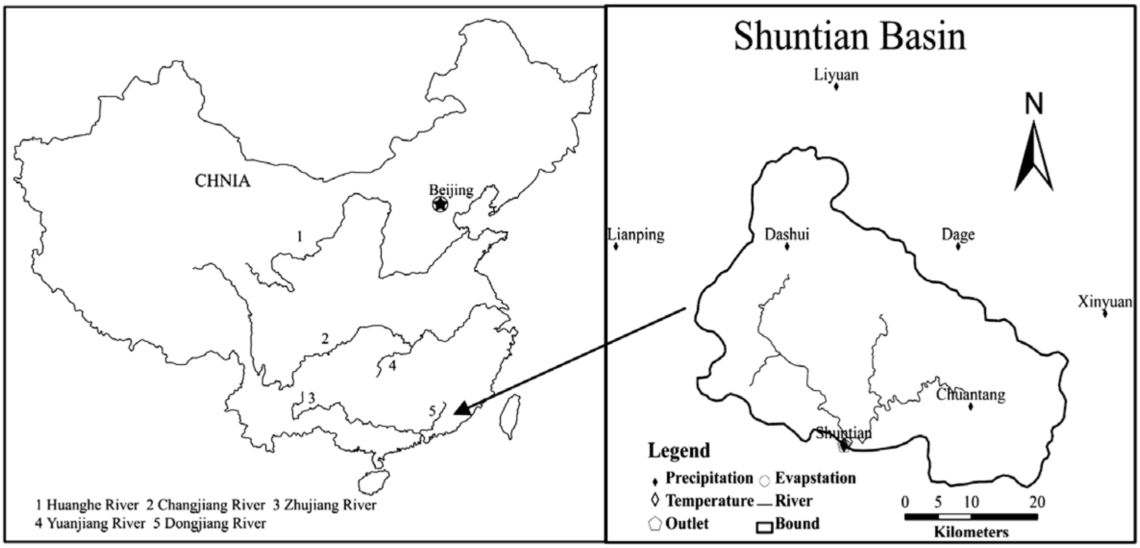

Fig. 1. Location, main rivers and meteorological stations and discharge station of Shuntian basin.

Table 1
Main equations of WASMOD model.

Actual evapotranspiration	$e_t = \min\{(sm_{t-1} + p_t)(1 - \exp(-a_1 ep_t), ep_t)\}$	$0 \leqslant a_1 \leqslant 1$
Slow flow	$s_t = a_2(sm_{t-1})^2$	$a_2 \geqslant 0$
Fast flow	$f_t = a_3(sm_{t-1})(p_t - ep_t(1 - \exp(-p_t/\max(ep_t, 1))))$	$a_3 \geqslant 0$
Total flow	$d_t = s_t + f_t$	
Water balance	$sm_t = sm_{t-1} + p_t - e_t - d_t$	

where a_i are model parameters, t is tth month, p_t is monthly precipitation, ep_t is monthly potential evapotranspiration, sm_{t-1} and sm_t are soil moistures of $t-1$ th and tth months, respectively.

and less uncertainty in parameters. This is confirmed by the variances of parameter samples given in Table 2, which reports posterior summary statistics for each parameter. Several features are of interest. First, both the GLUE and Bayesian methods have similar effective parameter spaces. Second, there is no significant difference between the two sets of mean values. This also holds for the two sets of median values. Third, although very small, differences can still be observed in the two sets of variances, which demonstrate that samples generated by the Bayesian method are less scattered. Also, variances obviously vary across parameters. Fourth, except for parameter a_2 by the GLUE method, the marginal posterior distributions for three parameters are nearly symmetric as indicated by the skewness coefficients given in Table 2 and as visually seen from Fig. 2. Finally, the interval length for parameter a_1 is nearly the same between the two methods, but for the other two parameters, the samples by the GLUE method have slightly larger intervals than those by the Bayesian method.

Correlations between the parameters estimated from the GLUE and Bayesian samples are shown in Table 3. The results show that correlation coefficients range between −0.048 and 0.303 for GLUE, and vary from 0.239 to 0.586 for Bayesian. In the Bayesian method, one would expect an alternative multivariate, instead of univariate, normal distribution as a proposal distribution where several parameters can be updated simultaneously (for instance, Kuczera and Parent, 1998; Engeland and Gottschalk, 2002). However, GLUE cannot deal with these correlations. It draws samples only from a single effective parameter space at one time.

Model fit

Once the posterior density has been derived, parameters can be estimated either in terms of the maximum likelihood value or the expected value. The goodness of model fit is made by comparison of observed against fitted hydrographs based on the posterior medians obtained from the two methods, and a series of plots of transformed residual analysis from the Bayesian method (Fig. 3).

Fig. 3a indicates that a reasonable fit of observed discharges is achieved by the WASMOD model based on posterior parameter median values for both the GLUE and the Bayesian methods. More clearly, the fit can be measured by ME, and the resulting ME for GLUE and the Bayesian method are 0.88 and 0.90, respectively. As the residual properties are also similar for the two methods, in Fig. 3b–d only the results from the Bayesian method are shown for illustrative purposes. The residuals against observed discharge are plotted in Fig. 3b. The figure suggests that the transformed residuals have a near constant variance. A plot of quantiles for residuals versus the standard normal distribution shows that residuals are normally distributed (Fig. 3c). In Fig. 3d merely two outliers from the 95% confidence interval indicates the mutual independence between residuals. These results show that both the chosen model and the likelihood function used seem to be quite realistic in this particular case study.

Uncertainty bounds of model simulation due to parameter uncertainty

To calculate confidence intervals of model uncertainty, 18,191 and 50,000 discharge values for each month were obtained by running the model with the 18,191 (behavioral) and 50,000 parameter sets from the GLUE (when threshold value is 0.8) and the Bayesian samples, respectively. The 95% confidence intervals due to parameter uncertainty were estimated by sorting those discharge samples. Compared to the Bayesian method, discharges in the GLUE method were weighted by the corresponding ME values, while those in the Bayesian method were not.

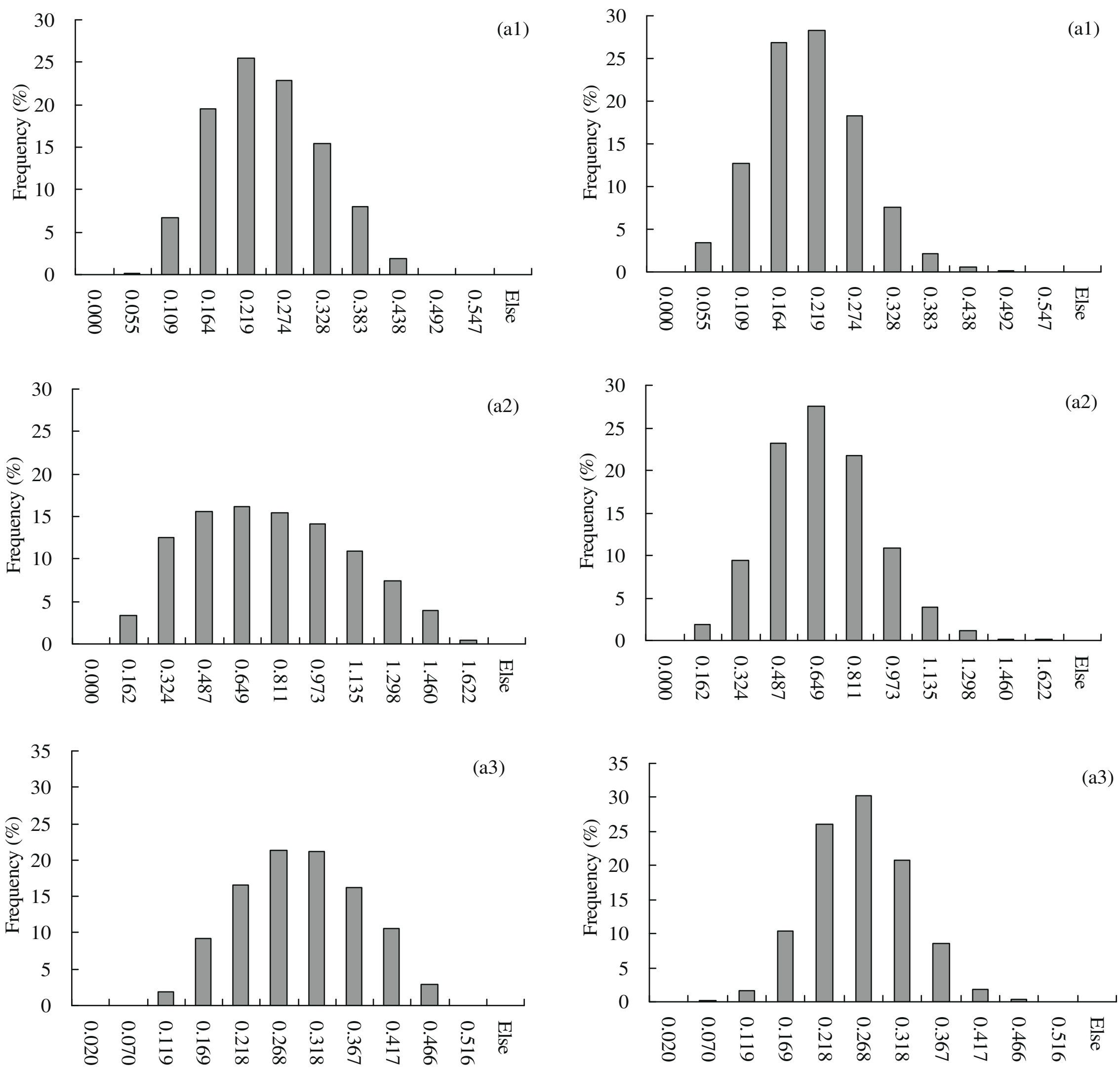

Fig. 2. Frequency histograms of parameters (left: by GLUE method; right: by Bayesian method).

Table 2
Summary of posterior distribution for each parameter.

Statistics	GLUE			Bayesian		
	a_1	a_2	a_3	a_1	a_2	a_3
Minimum	0.046	0.100	0.078	0.0	0.0	0.02
Maximum	0.433	1.531	0.460	0.547	1.622	0.516
Mean	0.220	0.694	0.272	0.181	0.588	0.239
Median	0.215	0.671	0.270	0.177	0.577	0.236
Variance	0.006	0.112	0.006	0.005	0.052	0.004
Skewness	0.242	0.245	0.030	0.336	0.363	0.183
$P_{0.025}$	0.087	0.149	0.124	0.045	0.180	0.123
$P_{0.05}$	0.102	0.189	0.142	0.065	0.241	0.142
$P_{0.95}$	0.355	1.281	0.403	0.303	0.977	0.344
$P_{0.975}$	0.376	1.354	0.421	0.334	1.078	0.366

*Percentiles P_α (α = 0.025, 0.05, 0.95, 0.975) are for the tail areas.

Table 3
Correlation between the parameters estimated by GLUE and Bayesian methods.

	GLUE			Bayesian		
	a_1	a_2	a_3	a_1	a_2	a_3
a_1	1.000	−0.048	0.303	1.000	0.248	0.239
a_2	−0.048	1.000	0.241	0.248	1.000	0.586
a_3	0.303	0.241	1.000	0.239	0.586	1.000

The model uncertainties due to parameter errors are plotted in Fig. 4. No obvious difference between the results of the two methods can be visually ascertained, but there indeed are some small differences in specific quantiles. In order to provide a quantitative evaluation of differences, different criteria might be adopted. For example, in the probability forecasting exercise, the following criteria are used, e.g., reliability, resolution and Brier scores (Lindsay and Cox, 2005; Gneiting et al., 2007). In probability and ensemble forecast the discrete ranked probability score and continuous ranked probability score are used by Toth et al. (2003) and Hersbach (2000). In this paper, a new quality measure, called "Average Relative Interval Length (ARIL)," was defined for comparison covering the whole time period as

$$ARIL = \frac{1}{n}\sum \frac{Limit_{Upper,t} - Limit_{Lower,t}}{Q_{obs,t}} \quad (5)$$

where $Limit_{Upper,t}$ and $Limit_{Lower,t}$ are the calculated upper and lower confidence limits for tth month, respectively; and n is the number of time steps. The term that the upper limit minus the lower limit divided by the observation is called relative interval length. Moreover, the percentages of observations that are contained in the calculated

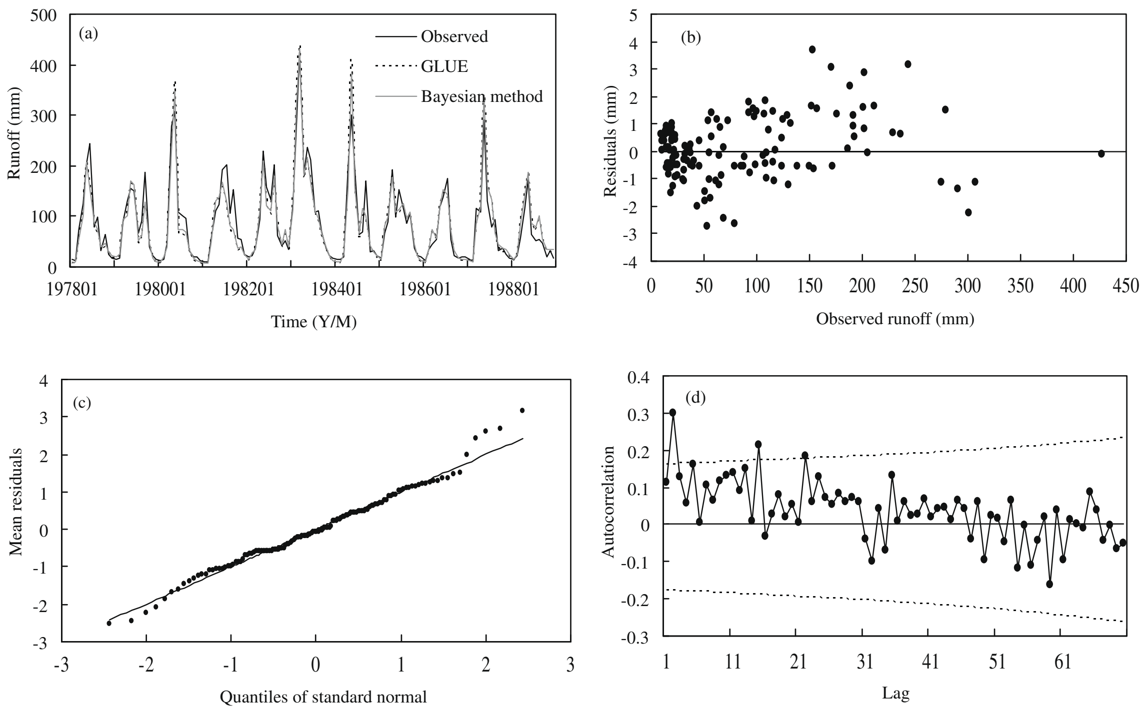

Fig. 3. Model-fit diagnostics: (a) comparison of monthly observed and simulated discharges based on two sets of posterior parameter median values from two methods, (b) residuals of Bayesian method versus observed discharges, (c) quantiles of mean residuals of Bayesian method versus standard normal quantiles, (d) partial autocorrelation function plot of residuals of Bayesian method with dashed lines representing 95% confidence interval about zero.

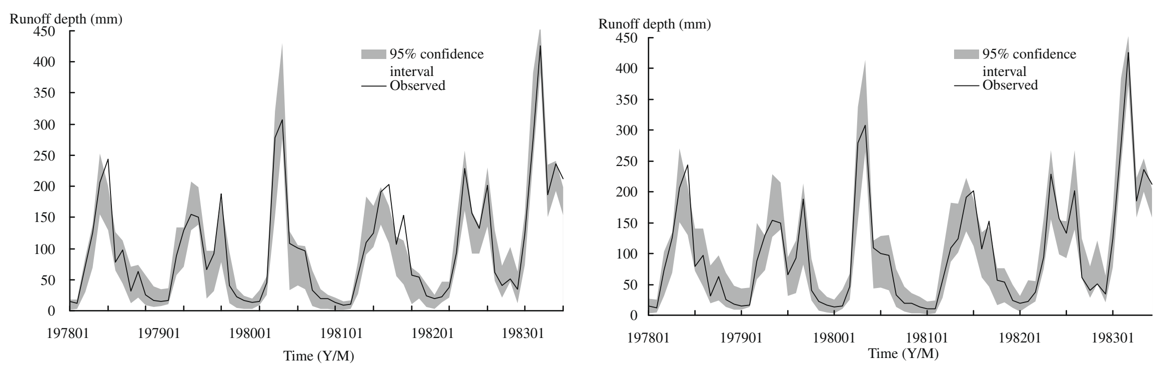

Fig. 4. Comparison of 95% confidence interval of monthly runoff due to parameter uncertainty calculated by GLUE method (left) and Bayesian method (right).

confidence intervals are calculated. A smaller ARIL value and at the same time a larger percentage represent a better performance.

The intervals at the 95% confidence level were considered. The ARIL value and the percentage of the observation coverage are 1.012% and 91% for the GLUE method, while the same statistics for the Bayesian method are 1.169% and 89%, respectively. These statistics of the two methods are quite similar, as can be expected since the summary properties of posterior distribution for each parameter are similar for the two methods (Table 2). This means the WASMOD model's structure well represents the hydrological behavior of the catchment.

The sensitivity of GLUE to the threshold value of the likelihood function

In the GLUE method the behavioral parameter sets are selected by choosing a cutoff threshold on the likelihood function or as a percentage of the best solutions from the sample. This feature together with the subjectivity involved in the definition of the likelihood function and the huge number of necessary model simulations are usually considered as one of the main drawbacks of the GLUE technique. Previous studies (e.g., Montanari, 2005; Blasone et al., 2008b) have highlighted that resulting parameter uncertainty and model simulation results are sensitive to the choice of threshold values. The effect of threshold values on the model results is dependent, among others, on the models, likelihood function, sample size and threshold values. In this study we illustrate how sensitive our model is to the choice of the threshold value when using the GLUE method with the Nash–Sutcliffe efficiency as a likelihood function. The illustration is done by calculating the posterior parameter cumulative distributions and model uncertainty with thresholds of 0.8 and 0.6, respectively. For comparison purposes, the posterior parameter cumulative distributions calculated using the Bayesian method are also shown in the same plot (Fig. 5). It is seen that when the threshold changed from 0.8 to 0.6, more parameter sets were retained as behavioral parameters in GLUE and the resulting posterior parameter cumulative distributions are much flatter than the other two. The 95% confidence intervals of model uncertainty due to parameter uncertainty

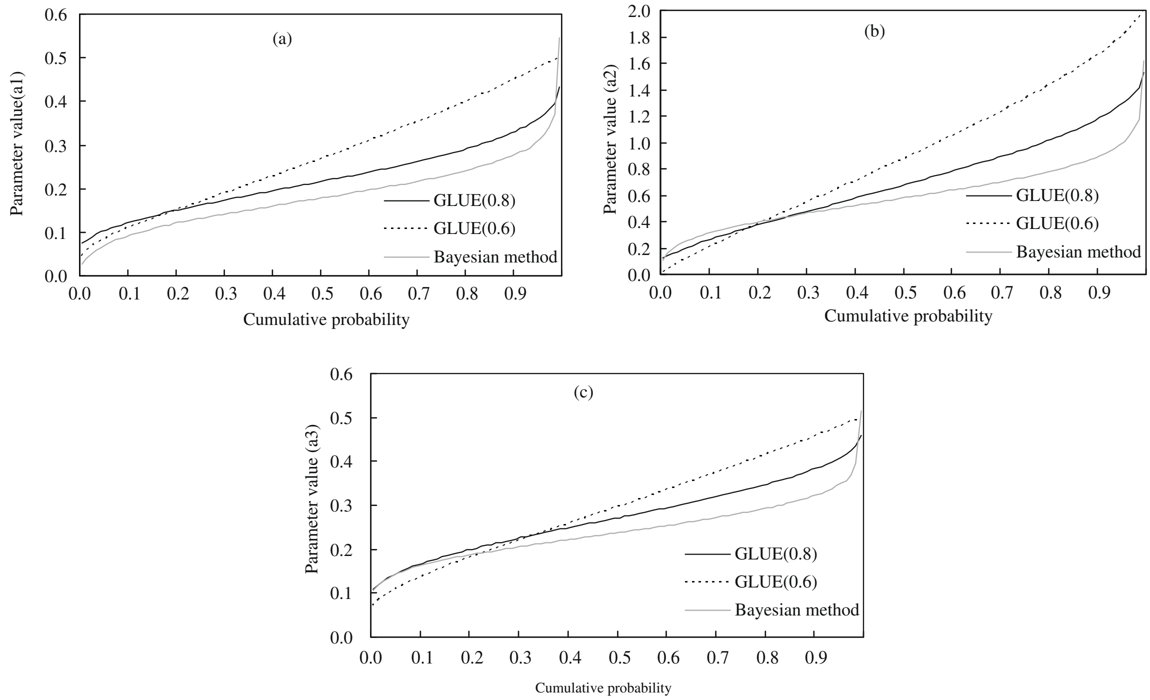

Fig. 5. Comparison of parameter cumulative probability by Bayesian method and GLUE method with thresholds of 0.8 and 0.6.

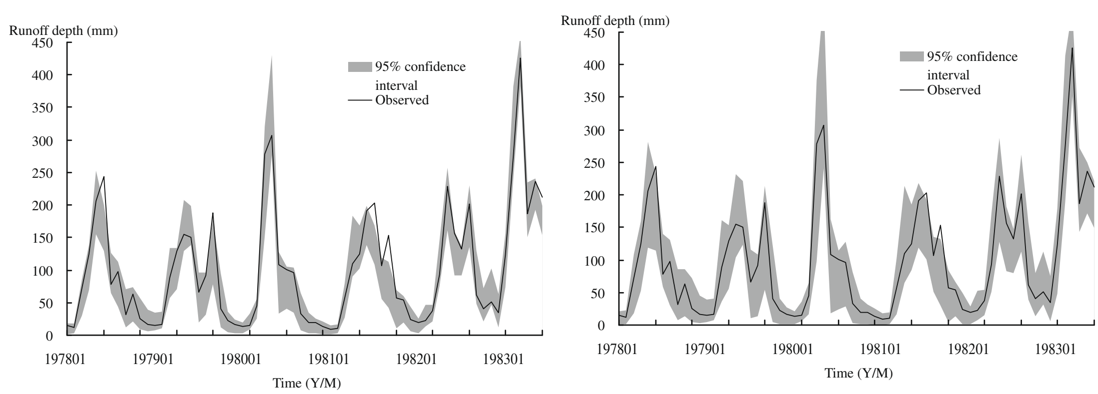

Fig. 6. Comparison of 95% confidence interval of monthly runoff due to parameter uncertainty calculated by GLUE method with thresholds of 0.8 (left) and 0.6 (right).

with threshold values of 0.8 and 0.6 are compared in Fig. 6. The ARIL values and the percentage of the observation coverage are 1.012% and 91% for the GLUE method with cutoff threshold value of 0.8, while the same statistics for the threshold value of 0.6 are 1.402% and 97%, respectively. This means that both the width of the confidence interval and the percentage coverage of the observation values are sensitive to the choice of the threshold values. There is, therefore, a need for further investigation to determine how the threshold value should be chosen in order to provide stabilized (may be difficult) and comparable results with each other in application of the GLUE method. This drawback of GLUE is related, to some extent, to the subjectivity in choosing the likelihood function. To reduce the subjectivity in the GLUE method, attempts have been made in the literature including Blasone et al. (2008b) who used an adaptive Markov Chain Monte Carlo scheme (the Shuffled Complex Evolution Metroplis (SCEM-UA) algorithm) to improve the computational efficiency of GLUE and proposed an alternative strategy to determine the value of the cutoff threshold based on an appropriate coverage of the resulting uncertainty bounds.

Discussion and conclusion

The GLUE and formal Bayesian methods are among the most popular techniques for uncertainty quantification of hydrological models. Formal Bayesian methods make strong assumptions about the nature of modeling errors, while GLUE tries to deal with uncertainty estimation by using an informal likelihood measure to avoid over conditioning. Recently, a strong debate has arisen in the hydrologic literature over the two methods concerning their theory (mathematics) and results. It is difficult to make a definite conclusion on the debate, because there is so much difference in the methods. At the same time progress has been made in improving the efficiency of the GLUE method (e.g., Blasone et al., 2008a,b; Xiong and O'Connor, 2008; Vrugt et al., 2008).

In this study parameter uncertainty and modeling uncertainty for the WASMOD model computed using GLUE and the Bayesian method are evaluated and compared for Shuntian basin, China, with the aim to examine their computational efficiency and discuss differences in their results. The following conclusions are drawn:

(1) there are slight differences in the parameter uncertainty values of the two methods, and parameter samples generated by the Bayesian method are less scattered than those by the GLUE method; (2) as for the simulated discharge uncertainty, the confidence intervals calculated by the Bayesian method and the GLUE method are very similar when the threshold value used in GLUE to determine the behavioral parameter sets is high enough. It seems that in our case study the actual shape of the posterior density function of parameters has a relatively limited effect on the final result; and (3) the posterior distribution of parameters and the 95% confidence interval of simulated discharge change significantly in the GLUE method as the threshold value changes; the lower the threshold value, the wider the confidence interval of model uncertainty. It should be noted here that the generality of results and conclusions of the study need to be verified through the application of other models in other regions.

Acknowledgments

The authors would like to thank the financial support of the Programme of Introducing Talents of Discipline to Universities – the 111 Project of Hohai University (B08048), the Outstanding Overseas Chinese Scholars Fund from CAS (The Chinese Academy of Sciences). We thank Prof Yongqin David Chen of the Chinese University of Hong Kong for providing the hydrological data through the Project No. CUHK4627/05H of the Research Grants Council of the Hong Kong Special Administrative Region, China. The authors would like to express their thanks to the two anonymous reviewers for their constructive and valuable comments, which improved the quality of the paper greatly.

References

Bates, B.C., Campbell, E.P., 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall–runoff modeling. Water Resources Research 37, 937–947.

Beven, K.J., 1993. Prophecy, reality and uncertainty in distributed hydrological modelling. Advances in Water Resources 16, 41–51.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 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., Freer, J., Hankin, B., Schulz, K., 2000. The use of generalised likelihood measures for uncertainty estimation in high order models of environmental systems. In: Fitzgerald, W.J. et al. (Eds.), Nonlinear and Nonstationary Signal Processing. Cambridge Univ. Press, New York, pp. 115–151.

Beven, K., Simth, 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, 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, 15–32.

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, 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. doi:10.1016/j.advwatres.2007.12.003.

Carpenter, T.M., Georgakakos, K.P., 2004. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulations of a distributed hydrological model. Journal of Hydrology 298, 202–221.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 49, 327–335.

Chowdhury, S., Sharma, A., 2007. Mitigating parameter bias in hydrological modelling due to uncertainty in covariates. Journal of Hydrology 340, 197–204.

Engeland, K., Gottschalk, L., 2002. Bayesian estimation of parameters in a regional hydrological model. Hydrological Earth System Sciences 6 (5), 883–898.

Engeland, K., Xu, C.-Y., Gottschalk, Lars, 2005. Assessing Uncertainties in a conceptual water balance model using Bayesian methodology. Hydrological Sciences Journal 50 (1), 45–63.

Freer, J., Beven, K.J., 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.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7, 457–472.

Georgakakos, K.P., Seo, D.J., Gupta, H., et al., 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. Journal of Hydrology 298, 222–241.

Gilks, W.R., Roberts, G.O., Sahu, S.K., 1998. Adaptive Markov chain Monte Carlo through regeneration. Journal of the American statistical association 93 (443), 1045–1054.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society Series B 69 (2), 243–268.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Hersbach, H., 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. Weather and Forecasting 15 (5), 559–570.

Jiang, T., Chen, Y.D., Xu, C.-Y., Chen, X., Singh, V.P., 2007. Development and testing of a simple physically-based distributed rainfall–runoff model for storm runoff simulation in humid forested basin. Journal of Hydrology 336, 334–346.

Jin, X.-L., Xu, C.-Y., Zhang, Q., Chen, Y.D., 2009. Regionalization study of a conceptual hydrological model in Dongjiang Basin, South China. Quaternary International 208, 129–137.

Kavetski, D., Franks, S., Kuczera, G., 2002. Confronting input uncertainty in environmental modelling. In: Gupta, H.V., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Calibration of Watershed Models. AGU Water Science and Applications Series, Duan, pp. 49–68.

Kavetski, D., Kuczera, G., Franks, S.W., 2006a. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. Water Resources Research 42, W03407.

Kavetski, D., Kuczera, G., Franks, S.W., 2006b. Bayesian analysis of input uncertainty in hydrological modeling: 2. Application. Water Resources Research 42, W03408.

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

Lindsay, D., Cox, S., 2005. Effective probability forecasting for time series data using standard machine learning techniques, In: Singh, S., et al. (Eds.), Pattern Recognition and Data Mining, vol. 3686. ICAPR 2005, LNCS, pp. 35–44.

Liu, Z.Y., 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, 347–364.

Mantovan, P., Todini, E., 2006. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. Journal of Hydrology 330, 368–381.

Marshall, L., Nott, D., Sharma, A., 2007. Towards dynamic catchment modelling: a Bayesian hierarchical modelling framework. Hydrological Processes 21, 847–861.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. The Journal of Chemical Physics 21, 1087–1092.

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.

Refsgaard, J.C., Storm, B., 1996. Construction calibration and validation of hydrological models. In: Abbott, M.B., Refsgaard, J.C. (Eds.), Distributed Hydrological Modelling, Water Science and Technology Library, vol. 22. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 41–54.

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, W00B06. doi:10.1029/2008WR006822.

Todini, E., 2008. A model conditional processor to assess predictive uncertainty in flood forecasting. International Journal of River Basin Management 6 (2), 123–137.

Toth, Z., Talagrand, O., Candille, G., Zhu, Y., 2003. Probability and ensemble forecasts. In: Jolliffe, I.T., Stephenson, D.B. (Eds.), Forecast Verification: a Practitioners's Guide in Atmospheric Science. John Wiley & Sons, Chichester, UK, pp. 37–164.

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), 1201. doi:10.1029/2002WR001642.

Vrugt, J.A., ter Braak, C.J.F., Gupta, H.V., Robinson, B.A., 2008. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? Stochastic Environmental Research and Risk Assessment 44. doi:10.1007/s00477-008-0274-.

Xiong, L-H., O'Connor, K.M., 2008. An empirical method to improve the prediction limits of the GLUE methodology in rainfall–runoff modelling. Journal of Hydrology 349, 115–124.

Xu, C.-Y., 2001. Statistical analysis of parameters and residuals of a conceptual water balance model – methodology and case study. Water Resources Management 15, 75–92.

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. Chapter 17.

Xu, C.-Y., Seibert, J., Halldin, S., 1996. Regional water balance modelling in the NOPEX area: development and application of monthly water balance models. Journal of Hydrology 180, 211–236.

Xu, C.-Y., Tunemar, L., Chen, Y.D., Singh, V.P., 2006. Evaluation of seasonal and spatial variations of conceptual hydrological model sensitivity to precipitation data errors. Journal of Hydrology 324, 80–93.

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, 167–182.

Yang, J., Reichert, P., Abbaspour, K.C., Xua, J., Yang, H., 2008. Comparing uncertainty analysis techniques for a SWAT application to the Chaohe basin in China. Journal of Hydrology 358, 1–23.