Uncertainty estimates by Bayesian method with likelihood of AR (1) plus Normal model and AR (1) plus Multi-Normal model in different time-scales hydrological models

Lu Li\textsuperscript{a,⇑}, Chong-Yu Xu\textsuperscript{a,b}, Jun Xia\textsuperscript{c}, Kolbjørn Engeland\textsuperscript{d}, Paolo Reggiani\textsuperscript{e,f}

\textsuperscript{a}Department of Geosciences, University of Oslo, P.O. Box 1047, Blindern, 0316 Oslo, Norway
\textsuperscript{b}Department of Earth Sciences, Uppsala University, Villaavägen 16, SE-752 36 Uppsala, Sweden
\textsuperscript{c}Key Laboratory of Water Cycle & Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
\textsuperscript{d}SINTEF Energy Research, 7465 Trondheim, Norway
\textsuperscript{e}Section of Hydraulic Structures and Probabilistic Design, Delft University of Technology, P.O. Box 5048, 2600GA Delft, The Netherlands
\textsuperscript{f}Deltares, P.O. Box 177, 2600MH Delft, The Netherlands

1. Introduction

The Bayesian method is widely used for uncertainty assessment in hydrologic modelling. During the last two decades several applications to various models and catchments have been pursued in the literature (Bates and Campbell, 2001; Duan et al., 2007; Engeland and Gottschalk, 2002; Li et al., 2010b; Marshall et al., 2007; Kavetski et al., 2006a,b; Kuczera et al., 2006; Kuczera and Parent, 1998; Krzysztofowicz and Kelly, 2000; Liu et al., 2005; Thiemann et al., 2001; Vrugt et al., 2003). The Bayesian method requires a specification of a likelihood function. For it to provide unbiased and reliable results several conditions need to be satisfied. The likelihood function provides a probability distribution of all simulation errors. It is therefore necessary to specify a correct distribution for the simulation errors and a possible mutual correlation structure. The most common approach is to transform the data for them to become normally and identically distributed (constant variance).

Bayesian revision is widely used in hydrological model uncertainty assessment. With respect to model calibration, parameter estimation and analysis of uncertainty sources, various regression and probabilistic approaches have been used in different models calibrated for either daily or monthly time step. None of these applications however includes a comparison of uncertainty analysis in hydrological models with respect to the time periods, at which the models are operated. This study pursues a comprehensive inter-comparison and evaluation of uncertainty assessments by Bayesian revision using the Metropolis Hasting (MH) algorithm with the hydrological model WASMOD with daily and monthly time step. In the daily step model three likelihood functions are used in combination with Bayesian revision: (i) the AR (1) plus Normal time period independent model (Model 1), (ii) the AR (1) plus Multi-Normal model (Model 2), and (iii) the AR (1) plus Normal time period dependent model (Model 3). In addition an index called the percentage of observations bracketed by the Unit Confidence Interval (\textit{PUCI}) was used for uncertainty evaluation. The results reveal that it is more important to consider the autocorrelation in daily WASMOD rather than monthly WASMOD. Firstly, the resulting goodness of fit of the daily model vs. observations as measured by the Nash–Sutcliffe efficiency value is comparable with that calculated by the optimization algorithm in monthly WASMOD. Secondly, the AR (1) model is not sufficiently adequate to estimate the distribution of residuals in daily WASMOD since \textit{PUCI} shows that Model 2 outperforms Model 1. Furthermore, the maximum Nash–Sutcliffe efficiency value of Model 2 is the largest. Thirdly, Model 3 performs best over the entire flow range, while Model 2 outperforms Model 3 for high flows. This shows that additional statistical parameters reflect the statistical characters of the residuals more efficiently and accurately. Fourthly, by considering the difference in terms of application and computational efficiency it becomes evident that Model 3 performs best for daily WASMOD. Model 2 on the other hand is superior for daily time step WASMOD if the auto-correlation of parameters is considered.

Keywords:
Uncertainty assessment
Hydrological models
Bayesian methods
Multi-Normal
Time scales

Article history:
Received 6 October 2010
Received in revised form 1 May 2011
Accepted 30 May 2011
Available online 16 June 2011

This manuscript was handled by Dr. A. Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Corresponding author.
E-mail address: lu.li@geo.uio.no (L. Li).
objective functions (Jin et al., 2010; Yang et al., 2008); (2) indices for goodness of fit at the level of the posterior distribution to evaluate the performance of uncertainty analysis methods: (1) Kuczera et al., 2006). Recently, some criteria have been proposed error and/or model structure error jointly or separately
Engeland et al., 2006; Yapo et al., 1998). Others consider the input sources lumped into parameter uncertainty such as the GLUE
Montanari and Brath, 2004; Reggiani and Weerts, 2008; Todini, 2008), because hydrological processes have a “memory effect”, the model residuals are often inter-correlated. In this context the discrete-time and continuous-time autoregressive error models have been used universally nowadays (Yang et al., 2007b), amongst which, the first order auto-regressive model (AR (1) model) is the most widely used. The latter is employed in combination with the transformation method to make the residuals satisfy the statistical assumptions of identical and independent distribution (Engeland et al., 2010; Krzysztofowicz, 2002; Yang et al., 2007a). Mantovan and Todini (2006) proposed a Multi-Normal AR (1) model as likelihood function by considering the autocorrelation of errors.

All uncertainty assessment methods account for different sources of uncertainty. Some methods use a statistical model of the residuals to address all the uncertainty sources (Engeland and Gottschalk, 2002; Jin et al., 2010; Krzysztofowicz, 2002; Montanari and Grossi, 2008). Others consider all uncertainty sources lumped into parameter uncertainty such as the GLUE method and the multi-objective method (Beven and Freer, 2001; Engeland et al., 2006; Yapo et al., 1998). Others consider the input error and/or model structure error jointly or separately (Chowdhury and Sharma, 2007; Kavetski et al., 2006a, 2006b; Kuczera et al., 2006). Recently, some criteria have been proposed to evaluate the performance of uncertainty analysis methods: (1) indices for goodness of fit at the level of the posterior distribution (efficiency) based on the Nash–Sutcliffe (NS) coefficient or other objective functions (Jin et al., 2010; Yang et al., 2008); (2) indices used to compare the width of the model uncertainty intervals (resolution) mainly based on 95% confidence interval of discharge (95CI), which include the width of 95CI, the average distance between the upper and the lower limits of 95% confidence intervals, 95PPU (d-factor), (Li et al., 2009; Xiong et al., 2009; Yang et al., 2008); and (3) indices to compare the distribution of the uncertainty estimates (reliability) which include the percentage of observations bracketed by the 95CI, the Continuous Rank Probability Score (CRPS). Engeland et al. (2010) proposed a hierarchy of all the criteria which are reliability, resolution and efficiency.

According to Todini (2007, 2011) all these studies treats the problem of “uncertainty”, which can be distinguished into “Validation Uncertainty” and “Predictive Uncertainty”. The “Validation Uncertainty” is the probability that the simulated value (water level, discharge, water volume, etc.) will be less than or equal to a prescribed value given prior knowledge, the historical information and the observed value, which refers to running the model in historical mode based on observations. Numerous approaches for quantifying the “Validation Uncertainty” have been proposed, including the Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Freer, 2001; Blasone et al., 2008a,b; Vogel et al., 2008; Xiong and O’Connor, 2008) and formal Bayesian methods (Montanari and Brath, 2004; Vuergt et al., 2003; Engeland et al., 2005; Yang et al., 2008; Jin et al., 2010; Li et al., 2010b). In which, the uncertainty has been estimated by 95% confidence interval (95CI) (Engeland et al., 2005; Li et al., 2009, 2010b) or 95% prediction uncertainty band (95PPU) (Abbaspour et al., 2007; Yang et al., 2008) calculated at the 2.5% and 97.5% levels of the cumulative distribution function of the model output variables. Krzysztofowicz (1999, 2001) defines “Predictive Uncertainty” as expressing the uncertainty on what may happen at a future stage conditional on the forecast of model, which was considered as a known value, based on the all available information, like input, parameter and model structure. It refers to running the model in forecast mode. After this, Todini (2007) redefined the predictive uncertainty that it is based on the posterior density of parameters and the conditional probability density of a predict and conditional on the covariates and a set of parameters. According to this, three approaches for the assessment of predictive uncertainty have been motioned (Todini, 2011), which are Hydrological Uncertainty Processor (HUP) (Krzysztofowicz, 1999), the Bayesian Model Averaging (BMA) (Raftery et al., 2003, 2005) and the Model Conditional Processor (MCP) (Todini, 2008; Coccia and Todini, 2010). This definition of predictive uncertainty is restricted to hydrologic forecasting. It might, however, argued that the predictive uncertainty might be defined for other applications that include predictions, e.g. streamflow in an ungauged catchment (conditioned on streamflow in a nearby catchment), or streamflow for a period of missing data (conditioned on observations outside this period). For this reason the difference between the “Predictive Uncertainty” and “Validation Uncertainty” as defined by Todini might be more related to the problem addressed (forecasting/non-forecasting) and not a general Bayesian definition of predictive uncertainty. In both types of uncertainty analyses, the Bayesian method requires a specification of a likelihood function. In order to provide unbiased and reliable results several conditions need to be satisfied. The likelihood function provides a probability distribution of all simulation errors. It is therefore necessary to specify a correct distribution for the simulation errors and a possible mutual correlation structure. The most common approach is to transform the data for them to become normally and identically distributed (constant variance).

Hydrological models at monthly time-step have been widely utilized for long-range streamflow forecasting, for assessing the long-term climatic change impact on water resources, and for regional water resources management (Xu and Vandewiele, 1995; Chen et al., 2007). Whereas daily or shorter time step hydrological models are required for real-time flood forecasting, for generation of synthetic sequences of hydrologic data for facility design, for studying the potential impacts of changes in land use or climate, and for interdisciplinary researches. Therefore, both monthly and daily hydrological modeling approaches are essential in water cycle and water resources researches. With respect to model calibration, parameter estimation, uncertainty sources analyses, various regression and probabilistic approaches have been employed in hydrological applications. Some include monthly water balance hydrological models (Benke et al., 2008; Engeland et al., 2005; Jin et al., 2010; Li et al., 2010b), others are daily conceptual hydrological models or physically-based distributed hydrological models (Blasone et al., 2008a; Engeland and Gottschalk, 2002; Li et al., 2010a; Liu et al., 2005; Looper et al., 2009; Wang et al., 2009; Yang et al., 2008, 2007b). However, none of the above applications seems to include comparisons of uncertainty analyses in time-variant hydrological models. It is common knowledge that the hydrological response variables are dependent on the time scale of model simulation. However, the question remains how model time scales impact on the performance of the uncertainty analysis methods and how models of similar structure operated with different time steps affect the uncertainty assessment via Bayesian methods. The objective of this paper is to attempt to fill this gap. This study performs a comprehensive comparison and evaluation of uncertainty estimates by Bayesian methods using the Metropolis Hasting (MH) algorithm in two validated conceptual hydrological models (WASMOD) with monthly and daily time.
steps. It aims at explaining the differences between the Bayesian method applied to hydrological models of similar concept and structure but with different time steps. The Box–Cox transformation in this study was used to obtain variables which are closer to a normal distribution. In daily WASMOD, two likelihood functions were used to remove the time dependence of residuals in the Bayesian method: (i) the AR (1) plus Normal model and (ii) the AR (1) plus Multi-Normal model. We also tested models, where the statistical parameters were dependent and/or independent on time, respectively. The Average Relative Interval Length (ARIL) defined by Jin et al. (2010), the percentage of observations bracketed by the Confidence Interval (PCI), and a new index called the Percentage of observations bracketed by the Unit Confidence Interval (PUCI) were used in this study for evaluation of model performance. Furthermore, the 95% confidence interval of daily discharge in low flow, medium flow and high flow were selected for comparison. Finally, the paper tries to give some advice to users on application of Bayesian method in validation uncertainty in time-variant hydrological model applications.

2. Study area and hydrological model

2.1. Study area

The selected study area is the Jiuzhou basin (Fig. 1) a tributary of the Dongjiang River in southern China. The river flows from the north-east to the south-west estuary with a drainage area of 385 km² (upstream area of Jiuzhou gauge station). The landscape is characterized by hills and plains. The Jiuzhou basin is located in a Monsoon region in a tropical and sub-tropical humid climate with a mean annual temperature of about 21 °C and only occasional incidents of winter daily air temperature dropping below 0 °C in the mountainous areas of the upper catchment part. Precipitation is generated mainly from two types of storms: frontal type and typhoon-type rainfalls. Eighty percent of the annual precipitation occurs during the wet season from April to September and the mean annual precipitation for the period of 1978–1988 is 1802 mm. The mean annual runoff is 993.4 mm or 55% of the annual precipitation. Precipitation data from Heduobu, Tangwei and Jiuzhou stations for the period of 1978–1988 are used for the study as the main input data. The potential evapotranspiration was calculated through the Penman–Monteith formula using meteorological data of the Heyuan meteorological station. All data have gone through quality control in earlier studies (Jin et al., 2010).

2.2. Wasmod

WASMOD is a well-tested simple conceptual water balance model with multiple time scales (Xu, 2002). The input data to the model include precipitation, potential evapotranspiration. Model output data are fast flow, slow flow, actual evapotranspiration and soil moisture. The model can be set up with different spatial and temporal scales varying from global to catchment and from daily to monthly (Gong et al., 2009; Jin et al., 2010; Widen-Nilsson et al., 2009, 2007; Xu, 2002). For this study a monthly and a daily version of the model at the catchment scale are used. The main
model equations are shown in Table 1. Snowmelt is not considered in the simulation because of the rare appearance of snow in the Jiuzhou River basin in the South of China. There are three parameters in monthly WASMOD and four parameters in daily WASMOD, with one additional routing parameter with respect to the former. Furthermore, parameter $a_2$ and $a_3$ have multiplied $10^3$ and $10^4$ respectively, aiming to be scaled into $[0, 1]$ in this paper.

### 3. Methods

#### 3.1. Residual analysis

The residuals $\xi_t$ represent the difference of the observed random variable $y_t$ and the simulated random variables $y_t^M$: $\xi_t = y_t - y_t^M$. Where $t$ is an index for time, $M$ indicates that the variable is simulated, $y$ is the variable in original space and $\eta$ is the transformed variable.

In this study, Box–Cox transformation method is used, which can be expressed as:

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

where $\lambda$ is a transformation parameter. It is often fixed, e.g. at the value of 0.3 in Yang et al. (2007) or 0.2 (Engeland et al., 2010). The square-root transformation and log-transformation are special cases of the Box–Cox transformation (Engeland and Gottschalk, 2002; Engeland et al., 2005; Jin et al., 2010; Langsrud et al., 1998; Wang et al., 2009).

The Lilliefors test (Lilliefors, 1967) was used to verify the normality of the distribution of the residuals of the Box–Cox transformed variable with $\lambda$ fixed at different values for the monthly and daily models. The test statistics is called LSTAT and it should be as small as possible. Calculations indicated that the Box–Cox transformation yields residuals closest to normal for values of $\lambda = 0.6$ in daily and $\lambda = 0.2$ in monthly WASMOD.

It was also necessary to verify whether $\xi_t$ were stochastically independent. In case of stochastic dependence, an AR (1) model needs to be fitted to the residuals as follows:

$$\xi_t - b = a(\xi_{t-1} - b) + \epsilon_t$$  

(3)

There is no need to use AR (1) model in monthly WASMOD, since the monthly residuals were not significantly correlated.

On the contrary, daily residuals were more heteroscedastic and autocorrelated. Using the Box–Cox transformation to reduce the heteroscedasticity of residuals increased their degree of autocorrelation. We therefore used the AR (1) model to make the Box–Cox transformed residuals independent. However, the AR (1) innovations were still auto-correlated. We therefore introduced AR (1) plus Multi-Normal model to account for all autocorrelations. This method has been used before in hydrological applications (Mantovan and Todini, 2006; Romanowicz et al., 1994).

Fig. 2 shows that the variances of AR (1) innovations from daily residuals have different values in different seasons. We therefore tested two statistical models: (i) in Model 1 the statistic parameters were constant and (ii) in Model 3 the parameters were allowed to depend on the season. Two simulation sub-periods were defined: a wet period from April to September and a dry period from October to March.

#### 3.2. Statistical models for the prediction errors

##### 3.2.1. Statistical model for monthly WASMOD

The assumption that there is neither bias nor autocorrelation in monthly residuals results in the following likelihood function for monthly WASMOD:

$$L(\theta, \varnothing; \xi) = \prod_{t=1}^{r} \frac{1}{\sigma} q(\xi_t^\lambda)$$

(4)

where $\varnothing$ represents the hydrological model parameters, $\omega$ represents the statistical parameters including in this case only the
standard deviation of the residuals $\sigma, \xi = \{\xi_1, \ldots, \xi_T\}$ is the residuals, $t$ is the time index, and $q$ is the normal density operator.

### 3.2.2. Statistical models for daily WASMOD

(a) The AR (1) plus Normal model in which statistical parameters depend on periods (Model 1)

The likelihood function for the AR (1) plus Normal model for the prediction errors in daily WASMOD were:

$$L(\omega, \theta | \xi) = \frac{1}{\sigma} q_{\sigma_0}(\xi_0) \prod_{t=1}^{T} \frac{1}{\sigma_t} q_{\sigma_t}(\xi_t - b - a(\xi_{t-1} - b))$$  \hspace{1cm} (5)

where $\theta$ represents the hydrological model parameters, $\omega$ represents the statistical parameters including an autoregressive term $a$, a bias term $b$ and a standard deviation of the residuals $\sigma, \xi_0$ is the initial residual value, $t$ is the time index, and $q$ is the normal density operator.

(b) The AR (1) plus Multi-Normal model (Model 2)

The likelihood function is built for time steps ranging from 1 to $T$ ($t = n$) given the AR (1) plus Multi-Normal model.

$$L(\omega, \theta | \xi) = \frac{\exp(-\frac{1}{2}(\xi - \mu)^T \Sigma^{-1}(\xi - \mu))}{(2\pi)^{n/2}(|\Sigma|)^{1/2}}$$  \hspace{1cm} (6)

where $\Sigma$ is the covariance matrix that can be approximated with:

$$\Sigma = \sigma_i^2 \begin{bmatrix} 1 & r_1 & r_1^2 & \cdots & r_1^{T-2} & r_1^{T-1} \\ r_1 & 1 & r_1 & \cdots & r_1^{T-3} & r_1^{T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_1^{T-2} & r_1^{T-3} & r_1^{T-4} & \cdots & 1 & r_1 \\ r_1^{T-1} & r_1^{T-2} & r_1^{T-3} & \cdots & r_1 & 1 \end{bmatrix}$$  \hspace{1cm} (7)

where

$$\sigma_i^2 = \text{var}[\xi_t]$$  \hspace{1cm} (8)

$$r_1 = \frac{\sum_{t=1}^{T-1} (\xi_t - \xi_{t-1})^2}{\sum_{t=1}^{T} (\xi_t - \xi_{t-1})^2}$$  \hspace{1cm} (9)

(c) The AR (1) plus Normal model in which statistical parameters depend on periods (Model 3)

The likelihood function for the daily WASMOD for which the statistical parameters depend on the periods is:

$$L(\omega, \theta | \xi) = \frac{1}{\sigma_0} q_{\sigma_0}(\xi_0) \prod_{t=1}^{T} \frac{1}{\sigma_t} q_{\sigma_t}(\xi_t - b - a(\xi_{t-1} - b))$$  \hspace{1cm} (10)

where all statistical parameters ($b, a, \sigma_t$) are labelled according to two periods: the wet period (April to September) and the dry period (rest months).

$$\omega_t \{b, a, \sigma_t\} = \begin{cases} \omega_{W} \{b_W, a_W, \sigma_W\} & t \in \{\text{April to September}\} \\ \omega_{D} \{b_D, a_D, \sigma_D\} & \text{others} \end{cases}$$  \hspace{1cm} (11)

### 3.3. Prior density

There is no information about the distribution of parameters. Parameter $\sigma$ is an unknown model standard deviation and with Jeffreys’ uninformative prior it is proportional to $\sigma^{-1}$ (Bernardo and Smith, 1994; Yang et al., 2007b). Besides, the prior probability densities of other parameters are generally taken as non-informative multi-uniform distribution in hydrological applications (Engeland et al., 2005; Jin et al., 2010; Liu et al., 2005; Yang et al., 2008).

The prior densities and intervals of all the parameters in the two hydrological models are shown in Table 2. The posterior density $\pi(\varphi | \xi)$ is:

$$\pi(\varphi | \xi) \propto L(\varphi | \xi) \cdot f(\varphi)$$  \hspace{1cm} (12)

where $f(\varphi)$ is the prior density, $\varphi = \{\omega, \theta\}$ and $\xi = \{\xi_1, \ldots, \xi_T\}$. The likelihood $L$ is given in Eqs. (4), (5), (6), and (10) above.

### 3.4. Metropolis Hasting algorithm

The Markov chain Monte Carlo method known as the Metropolis Hasting (MH) algorithm was used to sample from the posterior distributions. The Markov chain is initiated from a random starting value, whereby a new sample is derived from the proposal distribution. To avoid long “burn in” time, the Markov chain is initiated in correspondence of an approximated maximum Nash–Sutcliffe value calculated with the help of an optimization algorithm named VA05A (Hopper, 1978). A new sample was suggested from the proposal distribution. The new sample was accepted according to an acceptance probability. If not, the chain kept the old sample. The necessary steps of the Metropolis Hasting method have been explained in detail in the literature (Chib and Greenberg, 1995; Engeland and Gottschalk, 2002; Kuczera and Parent, 1998).

Here the proposal distributions $d$ were assumed as normal in the Random walk M–H Chain:

$$d(\omega) \sim N(\omega^0, \sigma_w)$$  \hspace{1cm} (13)

$$d(\theta) \sim N(\theta^0, \sigma_\theta)$$  \hspace{1cm} (14)

where $\omega = \{a, b, \sigma\}$ are the statistical parameters, $\theta$ are the hydrologic parameters, $L$ index the current state of the chain, $\sigma_w$ and $\sigma_\theta$ are the covariance matrixes of the proposal distributions. The normal distribution is not appropriate for $\sigma > 0$. However, in this study, the standard deviation of $\sigma$ was much smaller than the mean of $\sigma$, which means that the normal distribution density functions are mostly inside the feasible region. Other statistical parameters and model parameters in this paper satisfied the same assumption. The algorithm works best if the proposal density matches the target distribution which is generally unknown. The tuning of the variance was performed by calculating the acceptance rate. The ideal

### Table 2

<table>
<thead>
<tr>
<th>Hydrological models</th>
<th>Monthly WASMOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>$a_1, a_2, a_3$</td>
</tr>
<tr>
<td>Prior distributions</td>
<td>$U[0, 1], U[0, 1], U[0, 1]$</td>
</tr>
<tr>
<td>Hydrological models</td>
<td>Daily WASMOD</td>
</tr>
<tr>
<td>Parameter</td>
<td>$a_1, a_2, a_3, a_4$</td>
</tr>
<tr>
<td>Prior distributions</td>
<td>$U[0, 1], U[0, 1], U[0, 1], U[0, 1]$</td>
</tr>
<tr>
<td>$\sigma^{ab}$</td>
<td>$\propto \sigma^{-1}$</td>
</tr>
</tbody>
</table>

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

* $\propto \sigma^{-1}$ means the prior density of the parameter at value $\sigma$ is proportional to $\sigma^{-1}$. 
acceptance rate is approximately 40–50% for one parameter updating, decreasing to approximately 20–30% for multi-parameters updating in a one block (Chib and Greenberg, 1995; Engeland and Gottschalk, 2002). Furthermore, the Scale Reduction Score $\sqrt{R}$ (Gelman and Rubin, 1992) was used to check whether the MC chains converged.

3.5. Uncertainty confidence intervals estimation

The uncertainty intervals for streamflow due to parameter uncertainty were calculated using all the MH-samples of the hydrologic parameters. The samples were then sorted for every day in order to get credibility intervals. The empirical distribution $P_h(y < y_i)$ based on the rank $i$ was used to calculate the probability using the sorted sample $y_i$.

Table 3

<table>
<thead>
<tr>
<th>Method</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_1$</td>
</tr>
<tr>
<td>Min</td>
<td>0.783</td>
</tr>
<tr>
<td>Max</td>
<td>0.872</td>
</tr>
<tr>
<td>Average</td>
<td>0.830</td>
</tr>
<tr>
<td>Variance</td>
<td>1.59E-04</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.134</td>
</tr>
<tr>
<td>Estimated</td>
<td>0.829</td>
</tr>
<tr>
<td>2.50% Quantile</td>
<td>0.804</td>
</tr>
<tr>
<td>5% Quantile</td>
<td>0.809</td>
</tr>
<tr>
<td>50% Quantile</td>
<td>0.830</td>
</tr>
<tr>
<td>95% Quantile</td>
<td>0.850</td>
</tr>
<tr>
<td>97.50% Quantile</td>
<td>0.854</td>
</tr>
<tr>
<td>MNS</td>
<td>0.856</td>
</tr>
</tbody>
</table>

MNS: Maximum Nash–Sutcliffe.

Table 4

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>5% Quantile</th>
<th>50% Quantile</th>
<th>95% Quantile</th>
<th>MNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.967</td>
<td>0.993</td>
<td>0.981</td>
<td>1.83E-05</td>
<td>-0.164</td>
<td>0.974</td>
<td>0.981</td>
<td>0.988</td>
<td>0.805</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.112</td>
<td>0.177</td>
<td>0.143</td>
<td>1.26E-04</td>
<td>0.074</td>
<td>0.125</td>
<td>0.143</td>
<td>0.162</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.033</td>
<td>0.062</td>
<td>0.045</td>
<td>1.84E-05</td>
<td>0.42</td>
<td>0.038</td>
<td>0.044</td>
<td>0.052</td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.34</td>
<td>0.443</td>
<td>0.393</td>
<td>2.18E-04</td>
<td>-0.022</td>
<td>0.368</td>
<td>0.393</td>
<td>0.416</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.955</td>
<td>0.988</td>
<td>0.974</td>
<td>3.81E-05</td>
<td>-0.156</td>
<td>0.963</td>
<td>0.974</td>
<td>0.984</td>
<td>0.831</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.129</td>
<td>0.204</td>
<td>0.17</td>
<td>1.90E-04</td>
<td>-0.206</td>
<td>0.145</td>
<td>0.17</td>
<td>0.192</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.053</td>
<td>0.088</td>
<td>0.07</td>
<td>4.21E-05</td>
<td>-0.136</td>
<td>0.059</td>
<td>0.071</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.364</td>
<td>0.459</td>
<td>0.411</td>
<td>2.77E-04</td>
<td>-0.038</td>
<td>0.383</td>
<td>0.411</td>
<td>0.437</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.999</td>
<td>0.983</td>
<td>0.972</td>
<td>1.72E-05</td>
<td>-0.186</td>
<td>0.965</td>
<td>0.972</td>
<td>0.979</td>
<td>0.821</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.159</td>
<td>0.217</td>
<td>0.184</td>
<td>1.18E-04</td>
<td>0.715</td>
<td>0.17</td>
<td>0.182</td>
<td>0.206</td>
<td></td>
</tr>
<tr>
<td>$a_3$</td>
<td>0.057</td>
<td>0.088</td>
<td>0.073</td>
<td>2.55E-05</td>
<td>-0.393</td>
<td>0.064</td>
<td>0.074</td>
<td>0.081</td>
<td></td>
</tr>
<tr>
<td>$a_4$</td>
<td>0.329</td>
<td>0.462</td>
<td>0.394</td>
<td>3.91E-04</td>
<td>0.186</td>
<td>0.363</td>
<td>0.394</td>
<td>0.429</td>
<td></td>
</tr>
</tbody>
</table>

MNS: Maximum Nash–Sutcliffe.

$P_h(y < y_i) = i/n$ (15)

where $n$ is the number of sample simulations; $P_h(y < y_i)$ is the quantile of discharge $y_i$.

The 95% confidence intervals due to parameter uncertainty and model uncertainty for discharge were calculated by adding the model residuals in the form of a normally distributed random error with zero mean and variance $\sigma^2$ to each of the sample simulations that were available for each time step (Eq.(16)). For the daily model the bias and autoregressive terms were included as well (Eq.(17)). The inverse of the Box–Cox transformation was then applied.

$y_{i,t} = \eta_{i,t} + \text{norm}(0, \sigma_i)$ (16)

$y_{i,t} = \eta_{i,t} - b + a \cdot (\eta_{i,t-1} - \eta_{i,t-1} - b) + \text{norm}(0, \sigma_i)$ (17)

$y_{i,t} = h^{-1}(\eta_{i,t})$ (18)

where $i$ is the sample index, $t$ is the time index, $\text{norm}$ is the random value extracted from a normal distribution with zero mean and variance $\sigma^2$, and $h^{-1}$ represents the inverse operation of the Box–Cox transformation. Then the 5% percentile and 95% percentile of discharge due to parameter uncertainty and model uncertainty are derived by sorting new discharge values $y_{i,t}$ were sorted at each time step (Eq. (18)). And Eq. (15) was used to obtain the empirical distributions.

3.6. Evaluation of uncertainty interval estimates

Three indices were used to evaluate the resolution, reliability and efficiency of the uncertainty interval estimates: (1) the Average Relative Interval Length (ARIL) by Jin et al. (2010), (2) the percentage of observations bracketed by the Confidence Interval (PCI)
ARIL measures the resolution of the predictive distributions:

\[
ARIL_p = \frac{1}{n} \sum Limit_{\text{upper},t,p} \cdot Limit_{\text{lower},t,p} \tag{19}
\]

Limit_{upper,t,p} and Limit_{lower,t,p} are the upper and lower boundary values of the p confidence interval, n is the number of time steps, \( R_{\text{obs},t} \) is the observed discharge. ARIL should be as small as possible and was plotted as a function of p.

The PCI measures the reliability of the predictive distributions and it is based on the count, the number of the observed data within a range of simulated intervals, which is a function of p.

\[
PCI_p = \frac{\sum N_{Q_{\text{in}},t,p}}{n} \tag{20}
\]

in which, \( i \) is the acceptable sample index, \( t \) is the time index, \( R_{\text{obs},t} \) is the observed discharge, \( R_{\text{sim},t} \) is the simulated discharge, and \( \bar{R}_{\text{obs}} \) is the average value of \( R_{\text{obs},t} \). PCI was plotted as a function of p and should be close to the 45° diagonal line.

The maximum Nash–Sutcliffe value (MNS) measures the efficiency of the model predictions.

\[
MNS = \max \left\{ N_{S_i} \right\} \tag{21}
\]

\[
N_{S_i} = \frac{\sum_{i=1}^{T} (R_{\text{obs},t} - R_{\text{sim},t})^2}{\sum_{i=1}^{T} (R_{\text{obs},t} - \bar{R}_{\text{obs}})^2} \tag{22}
\]

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

However, the previous research (Li et al., 2010b) shows that it is not adequate to judge the uncertainty results by only ARIL or PCI, because they vary simultaneously. To facilitate a more direct uncertainty assessment for 95% confidence interval of discharge,

Table 5
The Correlation between the parameters for monthly WASMOD estimated by Bayesian method.

<table>
<thead>
<tr>
<th>Correlation</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td>1.000</td>
<td>-0.003</td>
<td>-0.029</td>
</tr>
<tr>
<td>( a_2 )</td>
<td>-0.003</td>
<td>1.000</td>
<td>0.725</td>
</tr>
<tr>
<td>( a_3 )</td>
<td>-0.029</td>
<td>0.725</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Fig. 4. Histograms of parameters of daily WASMOD derived by Metropolis Hasting method using: (a) Model 1; (b) Model 2 and (c) Model 3.
this paper proposes a new index based on ARIL and PCI, called the Percentage of observations bracketed by the Unit Confidence Interval (PUCI):

\[
PUCI = \frac{1.0 - \text{Abs}(PCI - 0.95))}{ARIL}
\] (23)

From Eq. (23), we can see that the PUCI is ranges from zero to infinity, and the upper boundary is not clear, which is a weak point of the index. It’s only used in 95% confidence interval evaluation. In fact the larger the PUCI the lower the uncertainty of 95% confidence interval of discharge is.

4. Results and discussion

4.1. Parameter estimates and parameter uncertainty

Five parallel simulation chains are used in the MH-algorithm with 5000 iterations each. The acceptance rate is ranging between 20% and 30%. The total number of samples of each parameter is 25,000. The Scale Reduction Score \(\sqrt{R}\) for all the parameters is approximately 1.0, indicating convergence of the iterative procedure for a chain.

Fig. 3 shows the marginal posterior parameter densities calculated by Bayesian method for monthly WASMOD. Sharp and peaky distributions show well identifiable parameters, while flat distributions indicate a larger parameter uncertainty. Fig. 3 shows that all parameters have well-defined posterior distributions, from which parameter estimates can be clearly inferred as modal values. The marginal posterior distributions for all parameters are nearly symmetric as indicated by the skewness coefficients in Table 3 and also visible in Fig. 3. Besides, it is evident that the interval of parameter \(a_1\) is very narrow and parameter \(a_3\) has the most skewed distribution.

Correlations between the parameters of monthly WASMOD estimated from Bayesian samples are represented in Table 5. The results show that the correlation coefficients range between \(-0.003\) and \(0.725\). The correlation coefficients of all parameters are quite small except between parameter \(a_2\) and \(a_3\) with a value of 0.725. It shows that the slow flow and fast flow are highly correlated since both of them are related to soil moisture content. For daily WASMOD, the marginal posterior parameter densities calculated by Models 1–3 are shown in Fig. 4. The following can be deduced from the Figure: (1) by comparing the posterior distributions of parameters derived by the three models, those estimated by Model 1 are slightly different; the variances and

Table 6
The Correlation between the parameters for daily WASMOD estimated by Bayesian method.

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a_1)</td>
<td>(a_2)</td>
<td>(a_3)</td>
<td>(a_4)</td>
<td>(a_1)</td>
<td>(a_2)</td>
</tr>
<tr>
<td>(a_1)</td>
<td>1</td>
<td>0.468</td>
<td>-0.617</td>
<td>-0.283</td>
<td>1</td>
<td>0.338</td>
</tr>
<tr>
<td>(a_2)</td>
<td>0.468</td>
<td>1</td>
<td>0.189</td>
<td>0.259</td>
<td>0.338</td>
<td>1</td>
</tr>
<tr>
<td>(a_3)</td>
<td>-0.617</td>
<td>0.189</td>
<td>1</td>
<td>0.327</td>
<td>-0.380</td>
<td>0.259</td>
</tr>
<tr>
<td>(a_4)</td>
<td>-0.283</td>
<td>0.259</td>
<td>0.327</td>
<td>1</td>
<td>-0.272</td>
<td>0.148</td>
</tr>
</tbody>
</table>

Fig. 5. The 95% confidence intervals of monthly discharge (1979/1–1988/12) due to parameter and model uncertainty (grey bands), simulated discharge at the maximum of the posterior distribution (solid) and observed discharge (dots).

Fig. 6. The Average Relative Interval Length (ARIL) and the percentage of observations bracketed by the Confidence Interval (PCI) due to different confidence interval value (Probability) for 1979–1988 monthly runoff in Jiuzhou river basin.
effective parameter space obtained from Model 3 is more similar to that from Model 1 than Model 2 and (2) the posterior distributions obtained from Model 1 are somehow sharper and narrower than those obtained from Model 2 and Model 3, indicating that a slightly more optimal parameter set has been identified.

This is confirmed by the variances of parameter samples given in Table 4, which reports posterior summary statistics for each parameter estimated by three likelihood functions for daily WASMOD. It shows several interesting features: (1) the variance of parameter samples estimated by Model 1 is smaller than those from Model 2 and Model 3; (2) the interval length for parameter \( a_4 \) is nearly the same for the three methods, while for the other three parameters, the samples by Model 2 and Model 3 have slightly larger intervals than those by Model 1; (3) the marginal posterior distributions for all parameters except for \( a_3 \) of Model 1 and \( a_2 \) of Model 3, are nearly symmetric as demonstrated by the skewness coefficients and (4) the maximum Nash-Sutcliffe (MNS) values of Model 1, Model 2 and Model 3 are 0.805, 0.831 and 0.821, respectively. Obviously, the MNS from Model 1 is the smallest. It indicates that Model 1 has low accuracy although it has the narrowest interval of posterior distribution for parameters. It means that residuals autocorrelation in daily WASMOD is more important, which needs to be considered for uncertainty assessment purposes.

Table 6 shows the correlations between the parameters samples estimated from the three Models. We see that compared to the monthly WASMOD, the average correlation between the parameters has increased slightly, but is still in the same range. And
3.3. Confidence interval of discharge due to parameters and model uncertainty

ARIL of 95% confidence interval due to parameter and model uncertainty derived from Model 2 is narrower than those from Model 1. Furthermore, the PCI obtained by Model 2 is nearly the same as that of Model 1; (2) ARIL of 95% confidence interval due to parameter and model uncertainty derived from Model 3 is the smallest with a value of 1.556, or less than 60% of 2.512 and 2.642 from Model 2 and Model 1, respectively. Meanwhile, the PCI of Model 3 is 92.9% which is quite close to 96.2% of Model 2 and 96.4% of Model 1. The explanation lies in the fact that the statistical parameters of Model 3 depend on simulation periods which makes the interval of dry period discharge to be much narrower and the result thus more reliable and less uncertain and (3) ARIL of 95% confidence interval due to parameter uncertainty is much narrower than those due to both of parameter uncertainty and model uncertainty for all statistical models in monthly and daily WASMOD. Above results reveal that the parameter uncertainty is not the most important key in uncertainty estimates for both monthly and daily hydrological models.

However, from results of Table 7, it is difficult to identify which likelihood function is the best for daily WASMOD when considering both ARIL and PCI as indices for confidence interval evaluation. For instance Model 1 has the biggest PCI value while Model 3 has the smallest ARIL value. In order to settle this problem and make the comparison easier and clearer, a new index as defined in Eq. (23) is used for comparison and the results are shown in Table 8. In which the values of PUCI based on low, medium and high flows due to 95% confidence interval for 1978–1982 daily runoffs in the Jiuzhou basin are compared for the three likelihood functions. It can be clearly seen that (1) Model 3 has the largest PUCI of 0.629 and Model 1 has the smallest PUCI of 0.373 for all flows. Consequently Model 3 is the best for daily time steps, highlighting the importance for the time-dependent statistical parameters; (2) the PUCI of Model 2 is the largest based on the high flow, implying that Model 2 is less uncertain and better in high flow estimate than the other two models and (3) an inter-comparison of Model 1 with Model 2 without time dependence shows in terms of PUCI values that Model 2 is superior to Model 1 over the whole range of flows. It reflects that Model 2 which uses more auto-correlation steps over the time periods outperforms Model 1.

4.3. Comparison of application and computational efficiency

Table 9 shows the comparison of application and computational efficiency of statistical models in WASMOD with different time-scales. From the running time, we can see that Model 2 is

### Table 7
The comparison of uncertainty measures between two hydrological models for the 95% confidence interval.

<table>
<thead>
<tr>
<th>Method</th>
<th>ARIL</th>
<th>PCI (%)</th>
<th>ARIL</th>
<th>PCI (%)</th>
<th>ARIL</th>
<th>PCI (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly WASMOD (79–88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bayesian_P</td>
<td>0.250</td>
<td>29.2</td>
<td>0.451</td>
<td>28.2</td>
<td>0.433</td>
<td>26.6</td>
</tr>
<tr>
<td>Bayesian_PM</td>
<td>1.516</td>
<td>94.2</td>
<td>2.642</td>
<td>96.4</td>
<td>2.512</td>
<td>96.2</td>
</tr>
<tr>
<td>Bayesian_PM</td>
<td>1.516</td>
<td>94.2</td>
<td>2.642</td>
<td>96.4</td>
<td>2.512</td>
<td>96.2</td>
</tr>
</tbody>
</table>

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.

### Table 8
The PUCI based on low, medium and high flows due to 95% confidence interval for 1978–1982 daily runoffs in Jiuzhou river basin.

<table>
<thead>
<tr>
<th>Method</th>
<th>Low Flows</th>
<th>Medium Flows</th>
<th>High Flows</th>
<th>All Flows</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARIL</td>
<td>PCI (%)</td>
<td>ARIL</td>
<td>PCI (%)</td>
</tr>
<tr>
<td>Model 1</td>
<td>0.255</td>
<td>0.428</td>
<td>0.805</td>
<td>0.373</td>
</tr>
<tr>
<td>Model 2</td>
<td>0.263</td>
<td>0.442</td>
<td>0.813</td>
<td>0.393</td>
</tr>
<tr>
<td>Model 3</td>
<td>0.610</td>
<td>0.607</td>
<td>0.675</td>
<td>0.629</td>
</tr>
</tbody>
</table>

now the largest correlation is between a1 and a3 and not a2 and a3 as in the monthly WASMOD.

### Table 9
Comparison of application and computational efficiency (Computer with Intel(R) Core(TM) 2 Duo CPU T8300 @ 2.40 GHz and 2.39 GHz, 1.96 GB of RAM).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
<td>Model 2</td>
</tr>
<tr>
<td>Difficulty of implement</td>
<td>Very easy</td>
<td>Little complicated</td>
</tr>
<tr>
<td>Number of runs × number of chains</td>
<td>5000 × 5</td>
<td>5000 × 5</td>
</tr>
<tr>
<td>Time</td>
<td>1 min</td>
<td>141.6 min</td>
</tr>
</tbody>
</table>

ARIL of 95% confidence interval due to parameter and model uncertainty derived from Model 2 is narrower than those from Model 1. Furthermore, the PCI obtained by Model 2 is nearly the same as that of Model 1; (2) ARIL of 95% confidence interval due to parameter and model uncertainty derived from Model 3 is the smallest with a value of 1.556, or less than 60% of 2.512 and 2.642 from Model 2 and Model 1, respectively. Meanwhile, the PCI of Model 3 is 92.9% which is quite close to 96.2% of Model 2 and 96.4% of Model 1. The explanation lies in the fact that the statistical parameters of Model 3 depend on simulation periods which makes the interval of dry period discharge to be much narrower and the result thus more reliable and less uncertain and (3) ARIL of 95% confidence interval due to parameter uncertainty is much narrower than those due to both of parameter uncertainty and model uncertainty for all statistical models in monthly and daily WASMOD. Above results reveal that the parameter uncertainty is not the most important key in uncertainty estimates for both monthly and daily hydrological models.
computationally more demanding than Model 1 and Model 3. Implementation of model 1 is very easy compared to Model 2. This is due to the fact that Model 2 considers the autocorrelation of whole sample periods, which implies hundreds or thousands of matrix calculations for the more complex likelihood function and subsequent processing. For monthly WASMOD, the computational efficiency is much higher than that of daily WASMOD. For daily WASMOD, although it is more efficient for Model 2 to find the correct posterior distribution of parameters than for Model 1 and Model 3, Model 2 is computationally more expensive than Models 1 and Model 3, which is certainly the main disadvantage of using this model in the Bayesian method for uncertainty assessment.

5. Conclusions

This study performed a comprehensive comparison and evaluation of uncertainty estimation through the Bayesian method by applying the Metropolis Hasting (MH) algorithm to two well-tested conceptual hydrological models (WASMOD) operated with monthly and daily time step. Several likelihood functions have been constructed to supply the uncertainty estimates in monthly and daily hydrological models. The Box–Cox transformation was used on the observed and simulated flows in all likelihood functions. For monthly WASMOD a statistical Gaussian error model was constructed for the residuals after the transformation. For daily WASMOD, the AR (1) plus Multi-Normal model was used after the transformation (Model 2). Furthermore, Model 1 and Model 3 compared the results derived by two likelihood functions whose statistical parameters dependent on time and independent on time, respectively. A novel index called Percentage of observations bracketed by the Unit Confidence Interval (PUCI) has been proposed for uncertainty evaluation of discharges. The following conclusions are drawn from the results of this study:

The autocorrelations of residuals in monthly WASMOD and daily WASMOD are different. In monthly WASMOD, it is not important to consider the autocorrelation, whereas in daily WASMOD it is necessary.

In monthly WASMOD, the posterior distributions of parameters were sharp and peaky which is associated with well identifiable model parameter sets. Besides, the goodness of model fit, as measured by the maximum Nash–Sutcliffe efficiency value, is comparable to the value calculated by the optimization algorithm.

The AR (1) model is not adequate enough to estimate the distribution of residuals in daily WASMOD. The values of PUCI show that Model 2 outperforms Model 1, although the posterior distributions of parameters derived by Model 1 are narrower than those obtained by the Model 2. Furthermore, the maximum Nash–Sutcliffe efficiency value obtained by Model 2 is 0.831, which is much higher than 0.805 achieved by Model 1. We conclude that the estimated parameters derived from Model 2 are more accurate than that of Model 1.

For daily WASMOD, the values of PUCI show that Model 3 performs the best over the entire flow range, while Model 2 is the best for high flows. Moreover, the MNS derived from Model 3 is 0.821, which is slightly smaller than that of Model 2 which is in turn larger than that of Model 1. This is because Model 3 uses more statistical parameters which reflect more appropriately the statistical properties of the residuals than Model 1. In fact Model 3 performs superior on 95% confidence interval of discharge estimates than Model 1.

Considering the difference in application and computational efficiency, we can see that a simple Gaussian error model is suitable for monthly WASMOD since there is no auto-correlation in transformed monthly residuals, while the value of PUCI shows that Model 3 is the best for daily WASMOD. But if considering the uncertainty of high flows and accuracy of estimated parameters, Model 2 is superior compared to Model 1 and Model 3 for daily WASMOD.

Despite that Model 2 outperforms in high flows for daily WASMOD, the PUCI of Model 2 is smaller than that of Model 3 for all flows, which shows that Model 2 is also sub-optimal for daily WASMOD. It is better for statistical models to split the simulation period. Two periods (dry and wet period) for the statistic parameters are probably not enough, which means that additional time sections need to be considered for Bayesian analyses in daily WASMOD. Furthermore, some low flows and high flows do not fit the normal distribution after the Box–Cox transformation and AR (1) modelling in daily WASMOD, which implies that further efforts are required to improve the formulation of likelihood functions used in hydrological applications since the assumption of normal distribution of the residuals in daily WASMOD over the entire flow range is inappropriate. Finally, one of the major challenges is to find an appropriate likelihood function for the error distribution when extreme flows are outside the normal distribution. In view of the computational effort required by Model 2, another important issue is to find an improved way to account for error auto-correlation in daily hydrological models. We observe that although the study has led to interesting findings, the generality of the results requires using additional hydrological models and study regions in future research.

Acknowledgments

This study was supported by Key Project of the Natural Science Foundation of China (No. 40730632), the Norwegian Research Council (RCN) for sponsoring project number 171783 (FRIMUF) and the Environment and Development Programme (FRIMUF) of the Research Council of Norway, project 190159/V10 (SoCoCA). The authors would like to thank Mario L.V. Martina for his advice and help in this study. We also appreciate the comments and suggestions from two reviewers and the associate editor, Prof. Ezio Todini. Their comments greatly helped to improve the quality of this paper.

References


