Development and comparison in uncertainty assessment based Bayesian modularization method in hydrological modeling

Lu Li a,⇑, Chong-Yu Xu a,⇑, Kolbjørn Engeland c

⇑ Corresponding author. Tel.: +47 22856678; fax: +47 22854215.
E-mail address: lu.li@geo.uio.no (L. Li).

a Department of Geosciences, University of Oslo, P.O. Box 1047, 0316 Oslo, Norway
b Department of Earth Sciences, Uppsala University, Sweden
c SINTEF Energy Research, 7465 Trondheim, Norway

SUMMARY

With respect to model calibration, parameter estimation and analysis of uncertainty sources, various regression and probabilistic approaches are used in hydrological modeling. A family of Bayesian methods, which incorporates different sources of information into a single analysis through Bayes' theorem, is widely used for uncertainty assessment. However, none of these approaches can well treat the impact of high flows in hydrological modeling. This study proposes a Bayesian modularization uncertainty assessment approach in which the highest streamflow observations are treated as suspect information that should not influence the inference of the main bulk of the model parameters. This study includes a comprehensive comparison and evaluation of uncertainty assessments by our new Bayesian modularization method and standard Bayesian methods using the Metropolis-Hastings (MH) algorithm with the daily hydrological model WASMOD. Three likelihood functions were used in combination with standard Bayesian method: the AR(1) plus Normal model independent of time (Model 1), the AR(1) plus Normal model dependent on time (Model 2) and the AR(1) plus Multi-normal model (Model 3). The results reveal that the Bayesian modularization method provides the most accurate streamflow estimates measured by the Nash–Sutcliffe efficiency and provide the best in uncertainty estimates for low, medium and entire flows compared to standard Bayesian methods. The study thus provides a new approach for reducing the impact of high flows on the discharge uncertainty assessment of hydrological models via Bayesian method.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Conceptual hydrological models are widely used tools for calculating the runoff dynamics and the water balance at various scales and regions (e.g. Xu et al., 1996, 2012; Widén-Nilsson et al., 2007, 2009; Kizza et al., in press; Zhang et al., 2012), for predicting hydrological impact of climate change (e.g. Veijalainen and Vehviläinen, 2008; Chiu et al., 2010; Jung et al., 2012; Wetherald and Manabe, 2002; Yang et al., 2012; Yuan et al., 2012), for estimating impact of land use changes (e.g. Niehoff et al., 2002; Petchprayoon et al., 2010; Rose and Peters, 2001; Wagener, 2007; Jiang et al., 2012), for simulating runoff in ungauged basins (e.g. Van Diewie and Elias, 1995; Xu, 1999; Kokkonen et al., 2003), and for flood forecasting (e.g. Kitanidis and Bras, 1980; Krzysztofowicz, 1999, 2002; Jasper et al., 2002; Nester et al., in press; Faulkner et al., 2012). Since the observations have unavoidable errors, computer models are not a perfect representation of reality, and model parameters are usually not uniquely identifi-

Available online 13 February 2013
Accepted 3 February 2013
Received in revised form 13 December 2012
Received 4 October 2012
Received in revised form 13 December 2012
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
High flows
Uncertainty assessment
Modularization Bayesian
Hydrological model
WASMOD

Article info

Article history:
Available online 13 February 2013
Accepted 3 February 2013
Received in revised form 13 December 2012
Received 4 October 2012

Article info

Article history:
Received 4 October 2012
Received in revised form 13 December 2012
Accepted 3 February 2013
Available online 13 February 2013
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
High flows
Uncertainty assessment
Modularization Bayesian
Hydrological model
WASMOD

With respect to model calibration, parameter estimation and analysis of uncertainty sources, various regression and probabilistic approaches are used in hydrological modeling. A family of Bayesian methods, which incorporates different sources of information into a single analysis through Bayes' theorem, is widely used for uncertainty assessment. However, none of these approaches can well treat the impact of high flows in hydrological modeling. This study proposes a Bayesian modularization uncertainty assessment approach in which the highest streamflow observations are treated as suspect information that should not influence the inference of the main bulk of the model parameters. This study includes a comprehensive comparison and evaluation of uncertainty assessments by our new Bayesian modularization method and standard Bayesian methods using the Metropolis-Hastings (MH) algorithm with the daily hydrological model WASMOD. Three likelihood functions were used in combination with standard Bayesian method: the AR(1) plus Normal model independent of time (Model 1), the AR(1) plus Normal model dependent on time (Model 2) and the AR(1) plus Multi-normal model (Model 3). The results reveal that the Bayesian modularization method provides the most accurate streamflow estimates measured by the Nash–Sutcliffe efficiency and provide the best in uncertainty estimates for low, medium and entire flows compared to standard Bayesian methods. The study thus provides a new approach for reducing the impact of high flows on the discharge uncertainty assessment of hydrological models via Bayesian method.

© 2013 Elsevier B.V. All rights reserved.

Without reference to the specific content of the image, the text is about the development and comparison of Bayesian modularization methods in hydrological modeling. It discusses the need for uncertainty assessment and the limitations of traditional methods. The study proposes a new approach that considers high flows as suspect information and evaluates its effectiveness compared to standard Bayesian methods.
distribution, the common approach is to assume the errors are normally and identically distributed. Then a data transformation is often needed to satisfy the assumption (Jin et al., 2010; Maranzano and Krzysztofowicz, 2004; Montanari and Brath, 2004; Todini, 2008). The discrete-time and continuous-time autoregressive error models (Engeland et al., 2010; Krzysztofowicz, 2002; Mantovan and Todini, 2006; Yang et al., 2007a) are commonly used.

Various regression and probabilistic approaches have been employed in hydrological applications for model calibration, parameter estimation, and analysis of uncertainty sources (Li et al., 2010b; Liu et al., 2005; Looper et al., 2009; Wang et al., 2009; Yang et al., 2008, 2007b). Some methods build a statistical model of the residuals to consider all uncertainty sources (Engeland and Gottschalk, 2002; Jin et al., 2010; Krzysztofowicz, 2002; Montanari and Grossi, 2008). Some others use parameter uncertainty to address all uncertainty sources, like the GLUE method (Beven and Freer, 2001). While the rest analyze the input error and/or model structure error jointly or separately (Chowdhury and Sharma, 2007; Kavetski et al., 2006a, 2006b; Kuczera et al., 2006). However, the problem with all the probabilistic approaches is that it is difficult to derive error models whose assumptions are fulfilled by the data. None of the above methods aim to improve the assessment of the uncertainty by considering different levels of flows, i.e., low, medium, and high flows. The success of the Bayesian method depends on the realism in the formulation of the likelihood function. It is, however, difficult to find a proper likelihood function that is suitable for all flows (low, medium, and high). One important question that has not yet been investigated is how high flows affect the uncertainty intervals of discharge and parameter estimates in the Bayesian method for hydrological modeling.

Modeling high streamflows is challenging due to quality of both data and models. The quality of high streamflow observations depends on the quality of the rating curve, and will in many cases be based on a few data points or even an extrapolation (Guererroa et al., 2012). In this case, high flows easily become outliers. In statistical analysis, the outlier’s problem is a complex one. Many studies of Bayesian analysis have attempted to treat the problems regarding outliers, influence data, error diagnostics or suspect data, i.e. assigning prior probabilities to different possible explanations of the outliers; deletion or down weighting is taken as a result of finding outliers (Barnett and Lewis, 1978; Hawkins, 1980; Cook and Weisberg, 1982; Chaloner and Brant, 1988). Kuczera et al. (2006) characterized model error by using storm-dependent parameters and pointed out that there is a need to deal with outliers. McMillan et al. (2010) designed a method to quantify uncertainty in river discharge measurements caused by stage-discharge rating curve uncertainty. In their study the discharge error distribution which is nonstationary in time, was parameterized. The results showed that flood flows’ uncertainty is high due to the limited number of observations, up to ±23% of the median discharge and ignoring rating curve uncertainty could lead to significant underestimation of the uncertainty associated with the model flow predictions, particularly during flood events. It also pointed out those natural uncertainties, i.e. hydraulic geometry shifts during flood events, are more significant than measurement uncertainties. Furthermore for high streamflows, the results are sensitive to the numerical solution of the hydrological model, i.e. how are the differential equations integrated in time (Clark and Kavetski, 2010; Kavetski and Clark, 2010, 2011). It is also difficult to calibrate the model to high streamflows since small timing errors might result in large errors in streamflow values and timing. Many non-Bayesian calibration strategies try to solve this problem. Sorooshian and Dracup (1980) developed maximum likelihood criteria for various types of streamflow errors, including the heteroscedastic maximum likelihood estimator (HMLE), which deals specifically the type of errors which is streamflow-value dependent (i.e., large flows have higher errors). Other studies shows that the parameters could be calibrated to fit different parts of the hydrograph using the multi-objective method or flow transformation (e.g. Engeland et al., 2010; Fencia et al., 2007; Madsen, 2000; Yapo et al., 1998; Xu, 2001) or a fuzzy multi-objective method (Yu and Yang, 2000). These papers show that model calibration is non-unique since no unique calibration will be optimal with respect to all performance measures considered (Madsen et al., 2002) and model parameters should be time-invariant in a given catchment (Fencia et al., 2008). One solution to this problem is to follow Hostache et al. (2011) who calibrate two hydrological models, one for low flow and one for high flows, and then derived a forecasting chain based on these two models. To establish the two models they use two objective functions, one enhancing the model error with respect to low flow simulation, and the other enhancing model error with respect to high flows based on Fencia et al. (2007). Coccia and Todini (2010) modeled the predictive uncertainty of flood forecasting by using a joint truncated normal distribution, in order to improve adaptation to low and high flows. The truncated normal distributions (TNDs) for low flows and high flows were used to divide the entire normal domain into two sub-domains in their Model Conditional Processor (MCP) framework, which means that the MCP can be applied assuming that the joint distribution in the normal space is not unique but can be divided into two TNDs.

The aim of this paper is to improve the uncertainty assessment in hydrological modeling by using a Bayesian modularization method. In this method reliable and suspect information used in parameter inference are separated. In this study, the 2% highest flows were considered as suspect data. This application of modularization method therefore aims to reduce the impact from suspected high flows and get more reliable uncertainty estimates of discharge for different levels of flow magnitudes, especially for low and medium flows. The primary aim was achieved by performing a comprehensive comparison and evaluation of uncertainty estimates by the new Bayesian modularization method and standard Bayesian methods using the Metropolis-Hastings (MH) algorithm. The study is exemplified using a daily conceptual hydrological model (WASMOD) (Li et al., 2011; Xu, 2002). In the standard Bayesian methods, 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 Multinormal model. We also tested error models where the statistical parameters were dependent and independent on time, respectively. This study can be considered as a continuation of the study by Li et al. (2011).

2. Study area and data

2.1. Study area

The Jiuzhou catchment, a tributary of the Dongjiang River in southern China, was used to perform the study (Fig. 1). The catchment has a drainage area of 385 km² at Jiuzhou gauge station. It is characterized by hills and plains. The catchment is located in a Monsoon region in the tropical and sub-tropical humid climate with a mean annual temperature of about 21 °C and only few days of air temperature dropping below 0 °C in the mountainous areas. The average annual precipitation for the period of 1978–1988 is around 1802 mm, with 80% falling in April to September. The average annual runoff is 993.4 mm or half of the annual precipitation.

2.2. Data analysis

Daily Precipitation data of 1978–1982 were the main input to the model. The Penman–Monteith formula was used to calculate
the potential evaporation. All data have gone through quality control in earlier studies (Jin et al., 2010; Li et al., 2011). To make a proper statistic model for the residuals of discharge, the Box–Cox transformation method was used:

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

(1)

where $y$ is the variable in original space, $\eta$ is the transformed variable and $\lambda$ is a transformation parameter. Criteria from Lilliefors (1967) were used to verify the normality of the transformed variable. Different values of $\lambda$ were evaluated and the best value was found to be 0.6 (Fig. 2). The residuals, defined as the difference between observed and simulated discharge values, are often heteroscedastic and autocorrelated at daily time step. The residuals $\xi_t$ of the transformed observations $\eta_t(y_t)$ and simulations $\eta^M_t(y^M_t)$ are modeled as:

$$\xi_t = \eta_t(y_t) - \eta^M_t(y^M_t)$$

(2)

where $t$ represents the index of time, $M$ indicates that the variable is simulated by hydrological model. It was also necessary to verify whether $\xi_t$ were stochastically independent. In our case the residuals were auto-correlated. We therefore fitted a AR(1) model. Fig. 4 shows that the variances of AR(1) innovations from daily residuals have different values in different seasons. We therefore divided them into two periods, i.e. a wet period from March to September and a dry period from October to April.

3. Methodology

3.1. Standard Bayesian methods

Bayes’ theorem is expressed as follows:

$$\pi(\phi | \xi) = \frac{L(\phi | \xi) \cdot f(\phi)}{\int L(\phi | \xi) \cdot f(\phi) d\phi}$$

(3)

where the posterior density $\pi(\phi | \xi)$ of the model parameters $\phi = \{\theta, \omega\}$ conditioned on a sequence of observations $\xi = \{\xi_1, \ldots, \xi_n\}$ is derived from the prior density $f(\phi)$ and the likelihood function $L(\phi | \xi)$. $n$ is the number of observations, and the model parameters were specified to be $\phi = \{\theta, \omega\}$ in which $\theta$ represents the hydrological model parameters and $\omega$ represents statistical parameters. The
integral in the denominator will in our application not be evaluated since we will use a Metropolis–Hastings algorithm.

In view of the results of data analyzes mentioned above, we tested three statistical models, whose likelihood functions \( L(\phi|\xi) \) are described in the following sub-sections (Li et al., 2011).

3.1.1. AR(1) plus Normal model independent of time (Model 1)

The likelihood function of the AR(1) plus Normal model (Yang et al., 2007b) for the prediction errors were:

\[
L(\phi|\xi) = \frac{1}{\sigma_0} \prod_{t=1}^{n} q \left( \frac{\xi_t - (\xi_{t-1} - b)}{\sigma} \right)
\]

where \( \phi \) 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.

3.1.2. The AR(1) plus Normal model dependent on time (Model 2)

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

\[
L(\phi|\xi) = \frac{1}{\sigma_0} \prod_{t=1}^{n} q \left( \frac{\xi_t - (\xi_{t-1} - b)}{\sigma_t} \right)
\]

where \( \phi \) are the statistical parameters and which are labeled according to two periods: dry period, \( W \) (March to September) and dry period, \( D \) (October to February).

3.1.3. The AR(1) and Multi-normal model (Model 3)

The likelihood function (Mantovan and Todini, 2006) is built for time steps ranging from 1 to \( t \) in order to predict \( t + 1 \):

\[
L(\phi|\xi) = \exp \left( -\frac{1}{2}(\xi - b)^T \Sigma^{-1} (\xi - b) \right)
\]

where \( \Sigma \) is the covariance matrix that can be approximated with (Basilevsky, 1983, p221; Mantovan and Todini, 2006):

\[
\Sigma = \sigma_c^2 \begin{bmatrix}
1 & r_1 & r_1^2 & \cdots & r_1^{n-1} & r_1^n \\
r_1 & 1 & r_1 & \cdots & r_1^{n-2} & r_1^{n-1} \\
r_1^2 & r_1 & 1 & \cdots & r_1^{n-3} & r_1^{n-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
r_1^{n-2} & r_1^{n-3} & r_1^{n-4} & \cdots & 1 & r_1 \\
r_1^{n-1} & r_1^{n-2} & r_1^{n-3} & \cdots & r_1 & 1
\end{bmatrix}
\]

where

\[
\sigma_c^2 = \text{var}[\xi_t] \quad t = 1, \ldots, n
\]

\[
r_1 = \frac{\sum_{t=2}^{n} (\xi_t - \bar{\xi})(\xi_{t+1} - \bar{\xi})}{\sum_{t=1}^{n} (\xi_t - \bar{\xi})^2}
\]

in which \( \sigma_c^2 > 0, r_1 \) is the first-order autocorrelation coefficient, \( \bar{\xi} \) is the mean of residuals.

3.2. Bayesian modularization method

Bayesian analysis incorporates different sources of information into a single analysis through Bayes’ theorem. This might be a problem if part of information has flaws. Then the other sources of information might be overly influenced, which impact directly on the results of uncertainty assessment. Liu et al. (2009) give a methodological suggestion that is called modularization for dealing with this problem in computer models. The modularization methods suggest keeping inference for different components or modules in the model totally or partly separated so that suspect information will be contained within one module and not influence all parts of the model.

In hydrological modeling the streamflow observations might be more or less suspect depending on the flow value. In this study we attempted to use the modularization concept, and we assumed that implausibly large flows are suspect information. This is partially supported by Fig. 2 which shows that only the residuals between -3 and 3 fit the normal distribution line. We therefore, in a first step, excluded implausibly large streamflow observations in the estimation of hydrological and statistic parameters estimation. In a second step the statistical parameters were re-estimated using both reliable and suspect information. We will describe this Bayesian modularization approach in more details in the following subsections.

3.2.1. Step one – inference using trusted information

The first step was to identify the trusted information and use these data for inference in a separate module. High flows are probably more suspect compared to low and medium flows, e.g. extremely sensitive to the numerical solution of the hydrological model; difficult to find a statistical model to address its property and in some condition the extrapolation of statistics can result in significant errors for small flows. In this study, we classify the runoff into three groups, i.e. high flows, medium flows and low flows, which are accounting for 2% of total flows, respectively. We found that 71 points accounting for around 2% of total points are out of normality line, in which 49 out-boundary spots are high flows accounting for 69% of total points and none are low flows (Fig. 3). So in this study, we assumed that all low and medium flows are trusted. We divide the entire data set into two sample groups: 2% highest flows was considered to be suspect sample group and the rest points are belonging to the trusted group, i.e. high flows (8%), all medium flows (70%) and all low flows (20%). For the trusted sample, the autocorrelation of residuals has been checked which is almost the same as before. It indicates that the elimination of some high flows did not break the condition of AR(1) model. Both hydrological and statistical parameters were estimated within the trusted sample. In this likelihood function we only considered those flows of \( \eta < \eta_c \):

Fig. 3. Plot of residuals vs. observed runoffs of daily WASMOMD at Jiuizhou station (1979–1988): (1) red dash divided low flows and medium flows and purple dash divided medium flows and high flows; (2) two black lines represents the normal region of residuals (−3, 3) according to Fig. 2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
\[
L(\phi | \zeta) = \frac{1}{\sigma_0} q \left( \frac{\zeta_0}{\sigma_0} \right) \cdot \prod_{t=1}^{m} \frac{1}{\sigma_t} q \left( \frac{\zeta_t - a_t (\zeta_{t-1} - b_{t-1})}{\sigma_t} \right)
\]

where \( m \) is the time dimension without suspected high flows, which means that for all \( t = 1, \ldots, m \), we have \( \eta_t \leq \eta_a^u \); \( \eta_a^u \) represents threshold value of high flows; \( \eta_a^u \) and \( \eta_a^l \) are the transformed observed variables and the transformed simulated variables at the initial time. The same as Model 2, Bayesian theorem updating of the prior distributions of the statistical parameters given in Table 1 are obtained separately for the two periods, i.e. wet period and dry period, for Bayesian modularization method. Other notations are as before.

In this case, the 2% highest flows that are considered as outliers, were deleted in Bayesian analysis of step one. For some cases that down weighing can be taken instead of deletion in Bayesian analysis if there is some knowledge or confidence about suspect data.

3.2.2. Step two – inclusion of suspect information

In the second step, the estimated values of the hydrological parameters from step one were used to provide streamflow realization. The suspect data, i.e. 2% highest flows were used together with the trusted data for inference of the statistical parameters, while the statistical parameters from Step two were used for uncertainty analysis of the trusted flows, i.e. low, medium and part of high flows, while the statistical parameters from Step two were used for the uncertainty analysis of suspect flows, i.e. 2% highest flows.

3.3. MCMC simulations

The Metropolis-Hastings (MH) algorithm (Hastings, 1970), a Markov Chain Monte Carlo (MCMC) methodology, was used to get the posterior distributions of the parameters (Chib and Greenberg, 1995; Engel and Gottschalk, 2002; Kuczera and Parent, 1998; Li et al., 2009a). The prior densities of all parameters in the daily WASMOD are shown in Table 1. Furthermore, the Scale Reduction Score \( \sqrt{R} \) (Gelman and Rubin, 1992) was used to check whether the Monte Carlo chains converged.

3.4. Estimation and evaluation of uncertainty intervals

All the MH-samples of the hydrologic parameters were used to calculate the uncertainty intervals for discharge due to parameter uncertainty, whereas the total predictive distribution was calculated by adding the model residuals to each of the sample simulations (Engeland et al., 2005; Li et al., 2011).

We wanted the predictive distribution for the streamflow to be narrow and reliable. An ideal uncertainty analysis technique would lead to a 95% probability band that is as narrow as possible while still being a correct estimate (i.e. on average 95% of the observations are inside this interval). The sharpness was measured by the Average Relative Interval Length (ARIL) (Gelman and Rubin, 1992) was used to check whether the empirical distribution is identical to the predicted distribution, that is, in the case of a perfect deterministic forecast (Hersbach, 2000; Yang et al., 2008). The PCI was proposed by Li et al. (2011) for uncertainty assessment of 95% confidence interval of discharge based on ARIL and PCI.

\[
PCI = (1.0 - \text{Abs}(PCI - 0.95)) /\text{ARIL}
\]

The PCI ranges from zero to infinity, and the upper boundary is not clear (Li et al., 2011). The larger the PCI the lower the uncertainty of 95% confidence interval of discharge is. Finally to be in line with tradition in hydrology the maximum Nash–Sutcliffe value (MNS) (Nash and Sutcliffe, 1970) is used for measuring efficiency. The MNS measures the efficiency of the model predictions, which should be as close to 1 as possible.

4. Results and discussion

The prior distributions specified in Table 1 were used. The MH-algorithm was used to obtain 5 independent chains with 5000 iterations each giving totally 25000 samples. The acceptance rate was between 20% and 30%. The Scale Reduction Scores \( \sqrt{R} \) for all the parameters have been limited between 0.9 and 1.1 to guarantee the convergence of the iterative procedure for a chain.
4.1. Parameter estimates

Fig. 5 shows the marginal posterior parameter densities calculated by Bayesian modularization method and standard Bayesian methods: the AR(1) plus Normal model independent of time (Model 1), the AR(1) plus Normal model dependent on time (Model 2) and the AR(1) plus Multi-normal model (Model 3). The x-axis scales of parameters densities of Bayesian modularization method

![Histograms of hydrological parameters of daily WASMOD derived by: (a) Model 1; (b) Model 2; (c) Model 3; (d) Bayesian modularization. (Note the scale difference in the x-axis.)](image)

Table 2

<table>
<thead>
<tr>
<th>Methods</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Bayesian Modularization</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARL</td>
<td>PCI (%)</td>
<td>ARL</td>
<td>PCI (%)</td>
</tr>
<tr>
<td>Bayesian _P</td>
<td>0.451</td>
<td>28.2</td>
<td>0.28</td>
<td>12.1</td>
</tr>
<tr>
<td>Bayesian _PM</td>
<td>2.489</td>
<td>96.4</td>
<td>1.963</td>
<td>94.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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.
are different from that of other three models, which are much wider. It shows that the posterior distributions derived by standard Bayesian methods are narrower and sharper than those obtained by Bayesian modularization method, which means less uncertainty in parameters. Besides, the posterior distribution of parameter $a_1$ in Bayesian modularization method is not symmetrical. All these indicate that the uncertain range of hydrological parameters is larger derived by Bayesian modularization method than others. It is probably because there are mainly low and medium flows to estimate hydrological parameters in step one of Bayesian modularization method. So there is less data to estimate the posterior distribution of hydrological parameters in Bayesian modularization method which leads to a larger estimation variance.

### 4.2. Reliability

The PCI counts the ratio of the observed data within a range of intervals $p$, which can be plotted as a function of $p$ and should be close to the 45 degree diagonal line, i.e. the 95% PCI should be as close to 0.95 as possible. The difference of 95% confidence intervals between all statistical models can be seen in Table 2. From the table, we can find that PCI of Bayesian modularization method is 92.3% which is close to 94.8% of Model 2, while Model 1 and Model 3 get the same result of 96.4%. Furthermore, PCI based on low, medium, high and all flows for the 95% confidence interval are shown in Table 3, which indicates (1) all models overestimated low flows and underestimated high flows to some extent; (2) Model 1 and Model 3 over-estimated PCI for all flows and medium flows, while which were slightly under-estimated by Bayesian modularization method; and (3) most of the observed data has been covered by the 95% confidence intervals for all flows, which can also be seen more clearly from Fig. 7 which presents the plots of change of the percentage of observations bracketed by the Confidence Interval (PCI) for different confidence interval values for daily WASMOD. It indicates that (1) PCI estimated by Model 1, Model 2 and Model 3 are more similar compared with the results from Bayesian modularization method; (2) for low, medium and all flows, Model 1, Model 2 and Model 3 obviously over-estimated the intervals; (3) for high flows, all models underestimated 95% confidence interval; and (4) the PCI derived from Bayesian modularization method is most stable compared with those from other three models for low, medium and all flows, which means Bayesian modularization method performs the best in Reliability.
Fig. 7. Proportion of observation inside confidence intervals (PCI) for Model 1, Model 2, Model 3 and Model 4 (Bayesian modularization) of daily runoffs in Jiuzhou river basin for (a) low flows, (b) medium flow, (c) high flows and (d) all flows. High flows are 10% highest observed flows, low flows are 20% lowest flows and medium flows are the rest flows.

Fig. 8. Average Relative Interval Length (ARIL) for Model 1, Model 2, Model 3 and Model 4 (Bayesian modularization) of daily runoffs in Jiuzhou river basin for (a) low flows, (b) medium flow, (c) high flows and (d) all flows. High flows are 10% highest observed flows, low flows are 20% lowest flows and medium flows are the rest flows.
4.3. Sharpness

The sharpness measured by $ARIL$ should be as small as possible. The 95% confidence intervals of discharge from 1982 estimated by all statistical models due to parameter and model uncertainty for daily WASMOD are plotted in Fig. 6 for illustrative purpose. It shows that the confidence interval derived by Bayesian modularization method is much narrower than that from the standard Bayesian methods, which indicates that the uncertainty of discharge estimated by this Bayesian modularization method is reduced. While comparing the three standard Bayesian methods in Fig. 6a–c, it indicates that the confidence intervals in low flow periods derived by Model 2 are narrower than those derived by Model 3 and Model 1. Fig. 8 shows the plots of $ARIL$ from Models 1–3 and Bayesian modularization method for all flow classes. In this figure we see that the confidence interval from the Bayesian modularization method is the sharpest for low, medium, high and all flows. For high flows, $ARIL$ estimated by Model 1 and Model 3 are almost identical, which are slightly less sharp than that from Model 2. The difference of 95% confidence intervals between all statistical models can be seen in Table 2. From the table, it is confirmed that $ARIL$ of 95% confidence interval due to parameter and model uncertainty derived from Bayesian modularization method is the smallest with a value of 1.150, followed by Model 2 of 1.963, and Model 1 and Model 3 of 2.489 and 2.362, respectively.

4.4. Measures combining sharpness and reliability

From results of Table 2, it is not yet clear enough to identify which statistical model is the best when considering both $ARIL$ and $PCI$ as the indices for confidence interval evaluation. For instance Model 1 has the biggest $PCI$ value while Bayesian modularization method has the smallest $ARIL$ value for 95% confidence interval. Therefore, two combined measures, $PUCI$ and $CRPS$ are calculated and their values are shown in Table 3 for 95% confidence intervals of all flow classes. It can be clearly seen that Bayesian modularization method has the largest $PUCI$ values for low, medium, high and all flows. Besides, the $CRPS$ results show that for low and medium flows accounting for 90% of total flows, the Bayesian modularization has the smallest value which indicates better perform in the credibility intervals estimates, while Model 1 performs slightly better in $CRPS$ for high flow.

4.5. Efficiency

The maximum Nash–Sutcliffe ($MNS$) values of Models 1–3 and Bayesian modularization method are shown in Table 3, which are 0.805, 0.823, 0.827 and 0.834, respectively. Again the MNS from Bayesian modularization method is the best. It indicates that Bayesian modularization method has higher accuracy although it has the widest interval of posterior distribution for parameters.

5. Conclusions

This study performed a comprehensive comparison and evaluation of uncertainty estimation between the proposed Bayesian modularization method and standard Bayesian methods by applying the Metropolis–Hastings (MH) algorithm to the well-tested daily conceptual hydrological model (WASMOD). The parameters uncertainty, reliability, sharpness and efficiency of uncertainty intervals have been compared. From results of this study the following conclusions are drawn:

- In terms of the $PCI$ criterion, the uncertainty interval derived from the Bayesian modularization method is most reliable. Other three Bayesian models performed equally well.
- In terms of the $ARIL$ criterion, the uncertainty interval derived by the Bayesian modularization method is the sharpest.
- The maximum Nash–Sutcliffe efficiency value obtained by the Bayesian modularization method is the highest.
- According to the results of $PUCI$, Bayesian modularization method outperforms other standard Bayesian methods over the entire flow range.
- The values $CRPS$ show that the Bayesian modularization method performs best in low and medium flows which is accounting for 90% of total flows.

The study show that the Bayesian modularization method performs better in terms of most criteria and flow classes than the standard Bayesian methods as exemplified by the daily WASMOD in the Jiuzhou river basin. However, it is worth of noting that the Bayesian modularization method still under-estimated the uncertainty interval of high flows in this study. The reason is probably that the normal distribution assumption for high flows is not good enough. One of the major challenges is therefore to find an appropriate likelihood function for the error distribution for different flow classes. Although this study has led to interesting findings, a topic for future research is to assess the generality of the results by testing the method in other areas using other models.

Acknowledgements

We thank two anonymous reviewers for their valuable manuscript comments. This paper is financially supported by China Ministry of Water Resources’ special funds for scientific research on public causes (No. 200901042), 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), and the 111 Program of Introducing Talents of Hydrology and Water Resources Discipline to Hohai University. We would like to thank Prof Yongqin David Chen of the Chinese University of Hong Kong for providing the hydrological data.

References


