Article
Chinese Science Bulletin

Geography

November 2013 Vol.58 No.31: 3796–3805
doi: 10.1007/s11434-013-5903-2

Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting

SANG YanFang[1,2*], SHANG LunYu[2], WANG ZhongGen[1], LIU ChangMing[1] & YANG ManGen[3]

[1] *Key Laboratory of Water Cycle & Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;*
[2] *Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, Chinese Academy of Sciences, Lanzhou 730000, China;*
[3] *School of Geographical Sciences, Fujian Normal University, Fuzhou 350007, China*

Received October 17, 2012; accepted March 26, 2013; published online May 30, 2013

Wavelet regression (WR) models are used commonly for hydrologic time series forecasting, but they could not consider uncertainty evaluation. In this paper the AM-MCMC (adaptive Metropolis-Markov chain Monte Carlo) algorithm was employed to wavelet regressive modeling processes, and a model called AM-MCMC-WR was proposed for hydrologic time series forecasting. The AM-MCMC algorithm is used to estimate parameters' uncertainty in WR model, based on which probabilistic forecasting of hydrologic time series can be done. Results of two runoff data at the Huaihe River watershed indicate the identical performances of AM-MCMC-WR and WR models in gaining optimal forecasting result, but they perform better than linear regression models. Differing from the WR model, probabilistic forecasting results can be gained by the proposed model, and uncertainty can be described using proper credible interval. In summary, parameters in WR models generally follow normal probability distribution; series' correlation characters determine the optimal parameters values, and further determine the uncertain degrees and sensitivities of parameters; more uncertain parameters would lead to more uncertain forecasting results and hard predictability of hydrologic time series.

hydrologic time series forecasting, wavelet, regression model, Bayesian theory, probabilistic forecasting, predictability

Citation: Sang Y F, Shang L Y, Wang Z G, et al. Bayesian-combined wavelet regressive modeling for hydrologic time series forecasting. Chin Sci Bull, 2013, 58: 3796–3805, doi: 10.1007/s11434-013-5903-2

Hydrologic time series forecasting is to reveal the future hydrologic regimes and further guide practical water activities [1,2]. The issue has received tremendous attention presently. Black-box models with two advantages as low quantitative demands of data and simple formulation are used commonly [3]. Linear regression (LR) models are the typical black-box models and developed extensively [4], but they are based on the stationarity and linearity assumptions. Artificial neural network (ANN) is another important type of black-box models. An ANN model can learn complicated nonlinear relationships [5]. However, the neural network structure is difficult to be determined, even if using trial-to-error procedure [6]. In addition, many parameters and neural network structure lack reliable physical basis, which are very important for hydrologic time series forecasting [7].

Observed hydrologic series in nature usually show nonstationary characteristics under multi-temporal scales. Wavelet analysis (WA) can elaborate the characteristics of a series in temporal and frequency domains, so it is suitable for handling the nonstationary characters of hydrologic series [8,9]. The combination of wavelet analysis with black-box models is a prevalent approach to conduct hydrologic time series forecasting, and various studies demonstrated the effectiveness of this practice [10,11]. However, hydrologic processes like any other natural processes have uncertainty [12,13], and forecasting result with a single optimal value is not convincing

*Corresponding author (email: sangyf@igsnrr.ac.cn; sunsangyf@gmail.com)

[14]. Hydrologic time series forecasting is to estimate an uncertain future hydrologic event [15]. Do probabilistic forecasting and then evaluate uncertainty would be more effective approach, since it enables decision-makers and the public to make decisions by considering uncertainty explicitly.

Generally, the uncertain factors influencing hydrologic forecasting include data uncertainty [16,17], model structure uncertainty [18], and parameter uncertainty [19]. When applying certain black-box model to hydrologic time series forecasting, they usually assume that the statistical properties of series are temporal persistent, so parameters remain constant over time when they are determined. This is often unreasonable, because climate and other factors impact hydrologic dynamics and may cause the changes of series' statistical properties [14]; moreover, it is also caused by the short or even missing length of series. Therefore, parameter is a key factor in hydrologic time series forecasting by black-box models. Bayesian theory is an universal theoretical framework for probabilistic forecasting by combining with deterministic models [20], and it can quantitatively describe the uncertainty of the object studied using known information [21,22]. By exploiting the advantage of Bayesian theories, this study is to propose a Bayesian-combined wavelet regression model for hydrologic time series forecasting, with uncertainty evaluation taken into account.

1 The AM-MCMC-WR model proposed

1.1 AM-MCMC-WR modeling processes

By combining the AM-MCMC (Adaptive Metropolis-Markov chain Monte Carlo) algorithm with wavelet regressive modeling, an improved model for hydrologic time series forecasting along with uncertainty evaluation is proposed, called AM-MCMC-WR. AM-MCMC is employed to estimate parameters uncertainty in WR model, based on which probabilistic forecasting can be conducted. The AM-MCMC algorithm will be described together with wavelet regressive modeling processes in the following. The modeling processes by AM-MCMC-WR are depicted in Figure 1, which contain four main steps: wavelet decomposition of series, determination of proper wavelet regression model, AM-MCMC sampling process of parameters, and probabilistic forecasting. The specific steps are described as

(1) For the hydrologic series $f(t)$ analyzed, choose suitable wavelet and decomposition level to decompose it into a set of sub-signals.

(2) Identify the deterministic components of series $f(t)$ by conducting significance testing of DWT, and combine them as $f_W(t)$. The result is used as the input data of wavelet regression model.

(3) Select proper linear regression model to fit the series $f_W(t)$:

$$f(t+\tau)=\sum_{i=1}^{p}\alpha_i f_W(t-i+1)\,, \tag{1}$$

where $f(t+\tau)$ is the value to be forecasted with the forecasting period of τ, $f_W(t\text{-}i\text{+}1)$ is the past hydrologic value, and α_i is the ith regressive parameter to be estimated with the total number of p.

(4) Estimate the value of parameter α_i (i=1, 2, …, p), and take the result as its initial sample value.

(5) Choose proper prior PDF $\pi(\alpha_i)$ of parameter α_i, and then initialization, j=0.

(6) Compute the covariance matrix C_i, and generate new parameter sample $\alpha_i^*\sim(\alpha_{i,j},C_i)$:

$$C_i=\begin{cases}C_0 & j\leqslant j_0,\\ s_d COV(\alpha_1,\alpha_2,...,\alpha_p)+s_d\varepsilon I_d & j\geqslant j_0,\end{cases} \tag{2}$$

where j_0 is initial sampling number, S_d is a scale factor, ε is a small positive number which ensure C_i be a nonsingular matrix, C_0 is the initial covariance matrix, and I_d is a unit matrix.

(7) Compute the acceptance probability of the new parameter α_i^* using eq. (3):

$$\rho=\min\left\{1,\frac{p(x/\alpha_i^*)\pi(\alpha_i^*)}{p(x/\alpha_{i,j})\pi(\alpha_{i,j})}\right\}. \tag{3}$$

(8) Generate a random number following uniform distribution as $u\sim U(0,1)$. If $\rho>u$, accept $\alpha_{i,j+1}=\alpha_i^*$; or else, accept $\alpha_{i,j+1}=\alpha_{i,j}$. Then $j=j+1$.

(9) Analyze each parameter's sampling results, and judge whether sampling results converge or not. If not, increase sampling number and do the same analyses in the steps of (6)–(8), until becoming stable.

(10) Forecast the hydrologic value with certain forecasting period using parameters' sampling results, and gain the probabilistic forecasting result (i.e. probability distribution). Finally, evaluate the uncertainty of forecasting results by estimating the credible intervals at certain confidence level.

1.2 Several key problems about AM-MCMC-WR modeling

When applying the above proposed model for hydrologic time series forecasting, some key problems should be carefully considered:

Wavelet decomposition of series. Accurate decomposition of series is a key problem in wavelet regressive modeling processes. Wavelet and decomposition level choices are two key problems in discrete wavelet decomposition. Sang [23] gave a practical guide to discrete wavelet decomposition. By comparing energy function of hydrologic series with the reference energy function, he proposed the methods for wavelet and decomposition level choice, and significance testing of DWT. It is used here for accurate decomposition of hydrologic series, as the basis of wavelet regressive modeling.

Value range of parameter. Estimating value ranges of

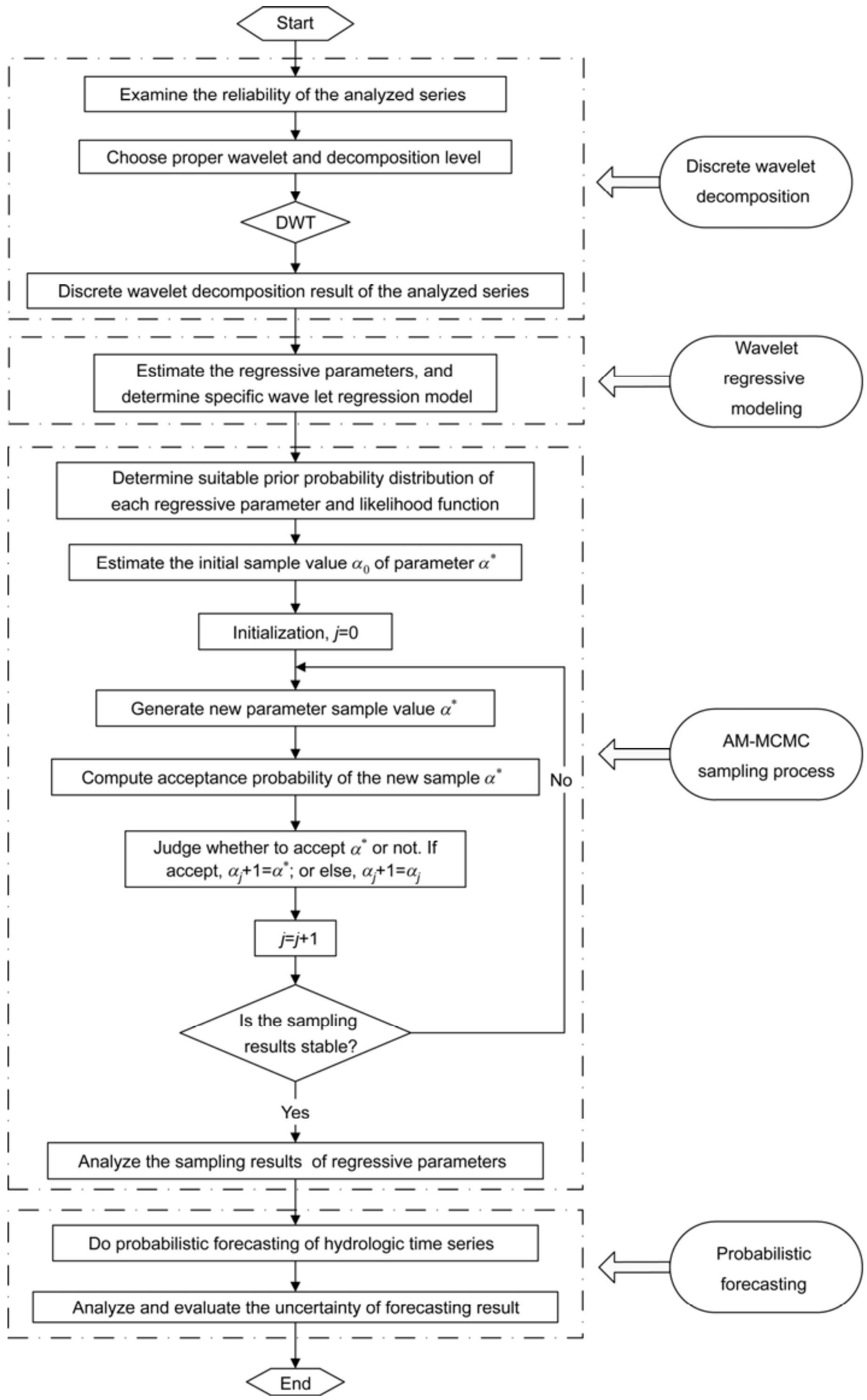

Figure 1 The processes of hydrologic time series forecasting along with uncertainty evaluation by the AM-MCMC-WR model proposed.

parameters is very important for determining proper prior PDF of parameter and for accurately describing parameter's uncertainty. Physical properties of hydrologic variables should be considered for estimating value ranges of parameters, for instance, parameter values must ensure the runoff and precipitation values being not smaller than zero.

Prior PDF of parameter. Selection of parameter's prior PDF is crucial for obtaining reliable sampling results. Both uniform and normal probability distributions are used in practice. If using the former, it is very important but difficult to determine parameter's value range. If using normal probability distribution, mean values of parameters can be estimated by proper method, and standard deviation can be valued suitable ratio of mean values, for instance, the ratio is 0.5 for two examples in "Case study". The normal probability distribution is recommended here, because it can avoid the estimation errors of parameters' value ranges.

Initial sample value of parameter. Determination of parameter's initial sample value is another very important problem which influences the efficiency of AM-MCMC algorithm [24]. For accelerating the convergence sampling rate, in this paper the parameter value estimated by the least-square method, but not random value as done in normal practices, is used as the initial sample value of parameter.

Likelihood function. Likelihood function reflects model's performance. The sampling results obtained by AM-MCMC would not converge to the posterior PDF if using unreasonable likelihood function. For the wavelet regressive modeling processes, the likelihood function in eq. (4) is used in this paper:

$$L=\left(\sigma_{\varepsilon}^{2}\right)^{-N}=\left((x-\hat{x})^{2}\right)^{-N}, \tag{4}$$

where x is the observed data and $\hat{x}$ is the modeled data; N is a parameter determining the weight of the analyzed parameter in model; the bigger the N is, the bigger weight of the analyzed parameter is.

2 Case study

2.1 Data

Two monthly runoff series measured at the Xixian (denoted as RS1) and Bengbu (denoted as RS2) hydrologic stations in the Huaihe River watershed are used in this study (Figure 2). Drainage areas of the two sites are 10190 km^2 and 121300 km^2 respectively. The two observed series have the same length of 41 years (492 months) from 1961 to 2001. After examining the reliability of data, the first 30-year data are chosen for calibration and the remaining 11 years are chosen for verification.

Statistical characters of the two series are presented in Table 1. Because the Bengbu station locates at the downstream and the Xixian station locates at the upstream, the runoff magnitude of the former is much bigger. The two runoff series show similar scattered variations, but RS1 series shows bigger positive skewness; RS2 series shows more obvious autocorrelation than RS1. In the calibration period of RS1 series, runoff data fall within the range of 3.27–1560.00 m^3/s, while those in verification period fall within the range of 4.16–1030.00 m^3/s, being smaller than the former. The same results can be found for RS2 series. According to these results, it is thought that extreme runoff values in verification data sets can be accurately modeled as long as nice models are established using the calibration data sets.

2.2 Discrete wavelet decomposition results

The two series were decomposed by significance testing of DWT using the "db8" wavelet. Energy functions of the two series are compared with the reference energy function, and wavelet decomposition results are displayed in Figure 3. The sub-signal under "D" levels is reconstructed by detail

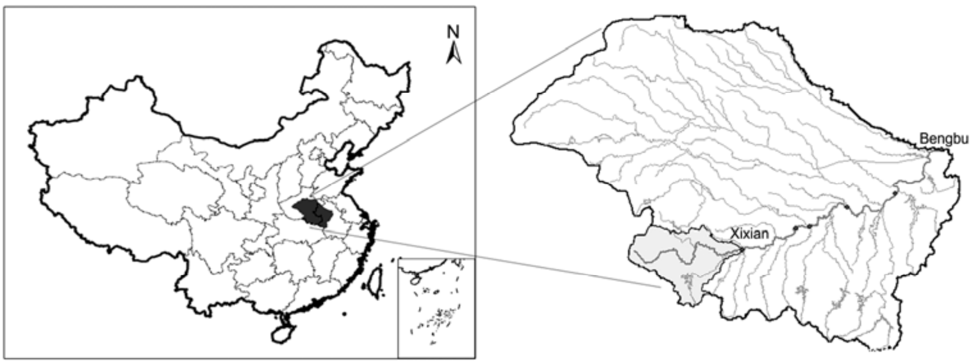

Figure 2 The locations of the Xixian and Bengbu hydrological stations in the Huaihe River watershed.

Table 1 Statistical characters of monthly runoff data measured at the Xixian and Bengbu hydrologic stations[a)]

Stations	Data set	Statistical characters						
		x_{mean} (m^3/s)	x_{min} (m^3/s)	x_{max} (m^3/s)	C_v	C_s	R_1	R_2
Xixian (RS1)	calibration	119.07	3.27	1560.00	1.43	3.70	0.32	0.14
	verification	103.31	4.16	1030.00	1.54	3.16	0.38	0.21
	whole	114.84	3.27	1560.00	1.46	3.59	0.34	0.16
Bengbu (RS2)	calibration	848.62	0.00	6130.00	1.21	1.93	0.65	0.32
	verification	654.24	0.00	5870.00	1.59	2.73	0.62	0.30
	whole	796.47	0.00	6130.00	1.30	2.13	0.65	0.32

a) The x_{mean}, x_{min}, x_{max}, C_v, C_s, R_1 and R_2 denote the mean, minimum, maximum, coefficient of variation, coefficient of skewness, lag-1 and lag-2 autocorrelation coefficients, respectively.

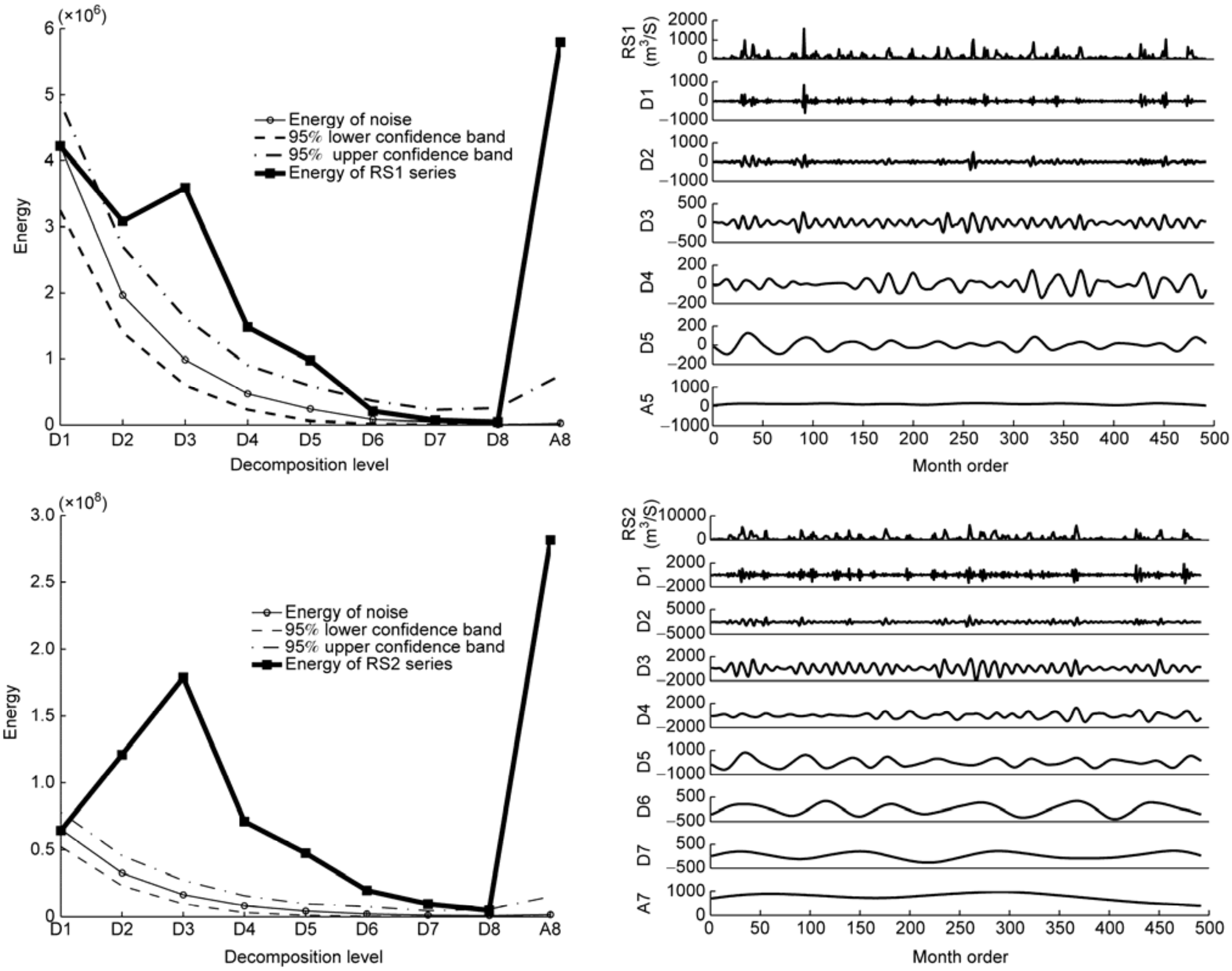

Figure 3 Energy functions (left) and wavelet decomposition results (right) of RS1 and RS2 series.

wavelet coefficients, and the sub-signal under "A" level is reconstructed by approximate wavelet coefficients. The sub-signals of RS1 series under D1, D6, D7 and D8 have the energies falling within 95% confidence interval, so they are thought as noise, and the others are the deterministic components for modeling RS1 series. The sub-signals of RS2 series under D1 and D8 are thought as noise, and the others are the deterministic components for modeling RS2 series.

2.3 Modeling processes

The two runoff series are analyzed by AM-MCMC-WR, LR and WR models for comparison. When modeling RS1 and RS2 series by LR model, both the input and output data are original data. When modeling RS1 series by WR model, the input data is the sum of sub-signals under D2, D3, D4, D5 and A8 (denoted as $f_{W1}(t)$), and the output data is original data; when modeling RS2 series by WR model, the input data is the sum of sub-signals under D2, D3, D4, D5, D6, D7 and A8 (denoted as $f_{W2}(t)$), and the output data is original data. The input and output data of two series by AM-MCMC-WR model are the same as those in WR modeling processes.

Both one-month- and three-month-ahead modeling are conducted. The input vectors to LR and WR models are determined by analyzing series' partial correlations. As shown in Figure 4, one-month-lag data are used as the input vectors of LR and WR models for RS1 series, and two month-lag data are used as the input vectors of LR and WR models for RS2 series. Parameters of LR and WR models are estimated by the least square method.

In the AM-MCMC sampling process of parameters, the initial sampling number is set as 200, and the total sampling number is set as 10000 for one-month-ahead forecasting but 30000 for three-month-ahead forecasting, mainly to ensure stable sampling results; ε equals 10^{-4}. S_d equals $2.4^2/d$, where d=1 for WR_1 model and d=2 for WR_2 model. Normal PDF is determined as the prior PDF of parameter, the mean of its normal PDF is valued as that estimated by the least square method, and the standard derivation is valued as 0.5 times of the mean value. The elements in matrix C_0 equal 30% of the standard deviation of the corresponding parameters.

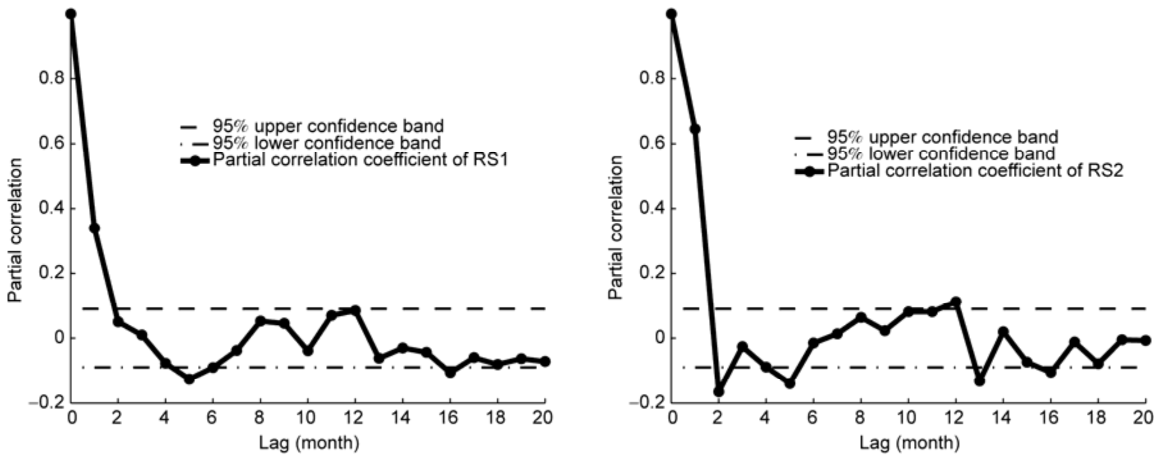

Figure 4 Partial autocorrelation functions of RS1 and RS2 series with the 95% confidence interval.

Three indexes, RMSE (root mean square error), AARE (average absolute relative error) and R^2 (coefficient of determination), are used to evaluate the forecasting results:

$$\mathrm{RMSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[f'(i)-f(i)\right]^2},$$
$$\mathrm{AARE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{f'(i)-f(i)}{f(i)}\right|\times 100\%, \tag{5}$$
$$R^2=1-\frac{\sum_{i=1}^{n}(f(i)-f'(i))^2}{\sum_{i=1}^{n}(f(i)-\overline{f(i)})^2},$$

where $f(i)$ is observed data with the total number of n and mean of $\overline{f(i)}$, and $f'(i)$ is modeled data.

2.4 Forecasting results

The LR and WR models for conducting one-month-ahead forecasting of RS1 and RS2 series, denoted as $\mathrm{LR}_1(1)$, $\mathrm{LR}_2(1)$, $\mathrm{WR}_1(1)$ and $\mathrm{WR}_2(1)$, are determined as eq. (6):

$$\begin{aligned}&\mathrm{LR}_1(1): f(t)=0.546\times f(t-1),\\&\mathrm{LR}_2(1): f(t)=0.867\times f(t-1)-0.092\times f(t-2),\\&\mathrm{WR}_1(1): f(t)=0.835\times f_{W1}(t-1),\\&\mathrm{WR}_2(1): f(t)=1.411\times f_{W2}(t-1)-0.596\times f_{W2}(t-1).\end{aligned}\tag{6}$$

The prior PDF of the parameter in $\mathrm{WR}_1(1)$ model is assigned as $\alpha_1\sim N(0.835, 0.418)$. The prior PDF of two parameters in $\mathrm{WR}_2(1)$ model is assigned as: $\alpha_1\sim N(1.411, 0.706)$, $\alpha_2\sim N(-0.596, 0.298)$. After the sampling number equals 4000 and till up to 10000, x_{mean} and C_v (coefficient of variance) of the sampling results of parameters become stable, indicating the convergent sampling results. Statistical characters of parameters' sampling results are shown in Table 2, which indicates that (1) The x_{mean} of parameters' posterior PDFs are very close to those in eq. (6), so optimal parameter results estimated by AM-MCMC are reliable, and sampling results are reliable. C_v values of all posterior PDFs differ with 0.5 prior determined, showing the prior and posterior PDF differences of parameters; (2) the x_{mean} and x_{mode} values of the posterior PDF of each parameter are similar, so it is thought that these parameters follow normal probability distribution; and (3) scatter degrees vary with the parameters analyzed. Bigger absolute optimal parameter value has smaller scatter degree, and *vice versa*. Scatter degree to some extents reflects the uncertainty and sensitivity of parameter. If certain input vector in WR model occupies an important proportion of original series, its corresponding parameter would have small scatter degree (i.e. C_v value) and small uncertainty, which means that the parameter has high sensitivity; however, if certain input vector in WR model occupies an insignificant proportion, its corresponding parameter would have big scatter degree and big uncertainty, and is not sensitive to the final forecasting result.

Based on parameters' sampling results, one-month-ahead probabilistic forecasting results at the verification period are obtained. The results are depicted in Figure 5 and evaluated in Table 3. Several conclusions follow: (1) Because x_{mean} and x_{mode} values of parameters' posterior PDFs are similar with those estimated by the least-square method, the optimal forecasting results by AM-MCMC-WR and WR models are also similar. (2) Three models used show different performances. The LR model is to directly analyze original series, so the results are influenced by noise and insignificant components in original series. However, when firstly separating the deterministic components, the above influences can be effectively overcome, and forecasting results can be improved. Scatterplots in Figure 5 show that forecasting results by AM-MCMC-WR and WR models are more close to the exact line compared to those by LR model. (3) Both optimal and probabilistic forecasting results of RS1

Table 2 Statistical analyses of the sampling results of parameters when conducting one-month-ahead forecasting by the AM-MCMC-WR model proposed

Model	Parameter	Estimated value by the leased square method	Characteristic statistics of sampling result					
			x_{max}	x_{min}	x_{mean}	x_{pv}	C_υ	C_s
WR1	α_1	0.835	2.868	−1.184	0.789	0.649	0.615	0.065
WR2	α_1	1.411	2.776	−0.168	1.293	1.385	0.308	−0.001
	α_2	−0.596	0.676	−2.235	−0.605	−0.554	0.681	−0.182

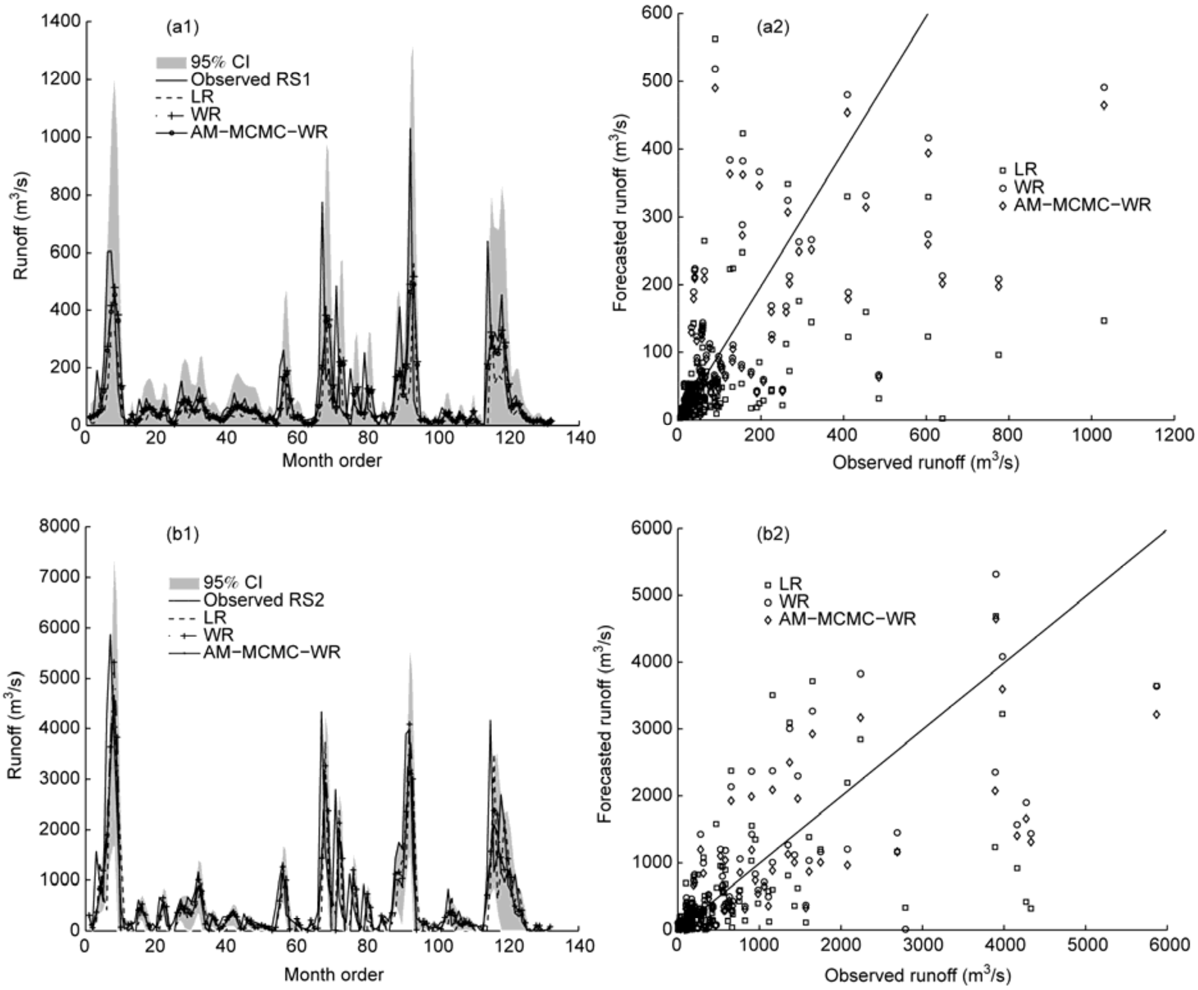

Figure 5 The results of one-month-ahead forecasting of RS1 (a1) and RS2 (b1) series in the verification period, and scatterplots of RS1 (a2) and RS2 (b2) series by LR, WR, and AM-MCMC-WR models.

Table 3 Comparison of the LR, WR and AM-MCMC-WR models in one-month-, and three-month-ahead forecasting of the RS1 and RS2 series

Forecasting period	Forecasted series	Model used	Index (calibration)			Index (verification)		
			RMSE	AARE	R^2	RMSE	AARE	R^2
One-month-ahead	RS1	LR	77.58	0.454	0.810	148.59	0.613	0.385
		WR	73.90	0.452	0.823	119.68	0.498	0.658
		AM-MCMC-WR	70.28	0.413	0.823	119.48	0.466	0.658
	RS2	LR	521.55	0.404	0.869	836.41	0.544	0.624
		WR	465.71	0.336	0.905	720.79	0.402	0.742
		AM-MCMC-WR	484.17	0.337	0.895	702.56	0.365	0.741
Three-month-ahead	RS1	LR	113.92	0.585	0.641	146.40	0.622	0.384
		WR	91.32	0.493	0.754	129.61	0.524	0.657
		AM-MCMC-WR	84.72	0.494	0.754	138.09	0.522	0.657
	RS2	LR	786.18	0.527	0.623	831.97	0.631	0.620
		WR	686.07	0.435	0.792	757.00	0.441	0.736
		AM-MCMC-WR	667.01	0.432	0.798	807.25	0.447	0.739

and RS2 series can be obtained by the AM-MCMC-WR model proposed. Figure 5 indicates that almost all original data (especially extreme runoff values) in verification period fall within the 95% credible interval (CI) estimated by the AM-MCMC-WR model, so it is thought that uncertainty is well estimated. However, conventional LR and WR models cannot do this. (4) Forecasting results of RS2 series are more accurate than those of RS1 series by any models, and uncertainty of the forecasting results of RS1 series is more obvious. It is mainly due to their different correlation characters. Since all the three models are based on series' correlation, those hydrologic series with good correlation would have easy predictability. For those hydrologic series with bad correlation, they would be difficult to be accurately modeled, and the results would have big uncertainty.

The LR and WR models for conducting three-month-ahead forecasting of RS1 and RS2 series, denoted as $LR_1(3)$, $LR_2(3)$, $WR_1(3)$, $WR_2(3)$, are determined as eq. (7):

$$\begin{aligned}
&LR_1(3): f(t) = 0.366 \times f(t-3),\\
&LR_2(3): f(t) = 0.474 \times f(t-3) - 0.023 \times f(t-4),\\
&WR_1(3): f(t) = 0.405 \times f_{W1}(t-3),\\
&WR_2(3): f(t) = 0.659 \times f_{W2}(t-3) - 0.210 \times f_{W2}(t-4).
\end{aligned} \tag{7}$$

In three-month-ahead forecasting, AM-MCMC sampling results of parameters are shown in Table 4. The three-month-ahead probabilistic forecasting results are presented in Figure 6. Compared with one-month-ahead forecasting results, we can find that (1) The sampling number ensuring stable sampling results should be up to 20000 in three-month-ahead forecasting. It is mainly due to more obvious parameter uncertainties along with forecasting period increase. (2) The understandings of parameters' uncertainties and sensitivities obtained here are similar as those in one-month-ahead forecasting. (3) Compared with Figure 5, Figure 6 shows that along with the increase of forecasting period, forecasting results of RS1 and RS2 series by any models become worse. It is due to the weak correlations of input and output data sets in long-step forecasting by regression models. Therefore, it is concluded that the performances of regression models become worse with the forecasting period increasing. Although the optimal forecasting results become worse, the 95% credible interval can also cover almost all original data, so uncertainty can be quantitatively estimated.

3 Discussion

Research on hydrologic time series forecasting has great significance in solving practical water activities. When applying conventional linear regression or wavelet regression models to hydrologic time series forecasting, the result is single value and cannot effectively consider uncertainty. In this paper, an AM-MCMC-WR model was proposed by employing the AM-MCMC algorithm to wavelet regressive modeling processes. Results of two examples verified its better performance compared with LR and WR models. Some understandings follow:

In the wavelet regressive modeling processes, absolute optimal value of parameter determines its uncertainty degree. Those parameters with bigger absolute optimal values would have weaker uncertainty and higher sensitivity, but the parameters with smaller absolute optimal values would have more obvious uncertainty and lower sensitivity. It is due to the correlations between input and output data sets in regression model. For the hydrologic series with good correlation characters, its dominant deterministic components (i.e. the input vector of WR model) would occupy an important proportion of original series, and the corresponding parameter would have big absolute value, small scatter degree and uncertainty, but high sensitivity.

Uncertainties and sensitivities of parameters in regression model determine the uncertainty of probabilistic forecasting result. When parameters have weak uncertainties and high sensitivities, the probabilistic forecasting result would have small uncertainty, whereas if parameters have small sensitivities and obvious uncertainties, forecasting results would have big uncertainties, and the estimated credible interval with certain confidence level would be much big. It indicates the uncertainty consistence between parameters and hydrologic time series forecasting.

Those hydrologic time series with good correlation characters are easier to be modeled than those with weak correlation characters when using linear regression models. It is determined by the essence of linear regression models, also including wavelet regression model and the AM-MCMC-WR model proposed. Generally, those hydrologic series with good correlation characters would correspond to the parameters with big absolute value, smaller scatter degree and uncertainty, and high sensitivity, by which accurate result of hydrologic time series forecasting with small uncertainty can be obtained, that is, the hydrologic series has

Table 4 Statistical analyses of the sampling results of parameters when conducting three-month-ahead forecasting by the AM-MCMC-WR model proposed

Model	Parameter	Estimated value by leased square method	Characteristic statistics of sampling result					
			x_{max}	x_{min}	x_{mean}	x_{pv}	C_v	C_s
WR1	α_1	0.405	3.194	−2.172	0.263	0.229	2.021	0.124
WR2	α_1	0.659	2.504	−1.470	0.568	0.467	0.784	0.009
	α_2	−0.210	1.598	−2.169	−0.271	−0.201	1.600	−0.021

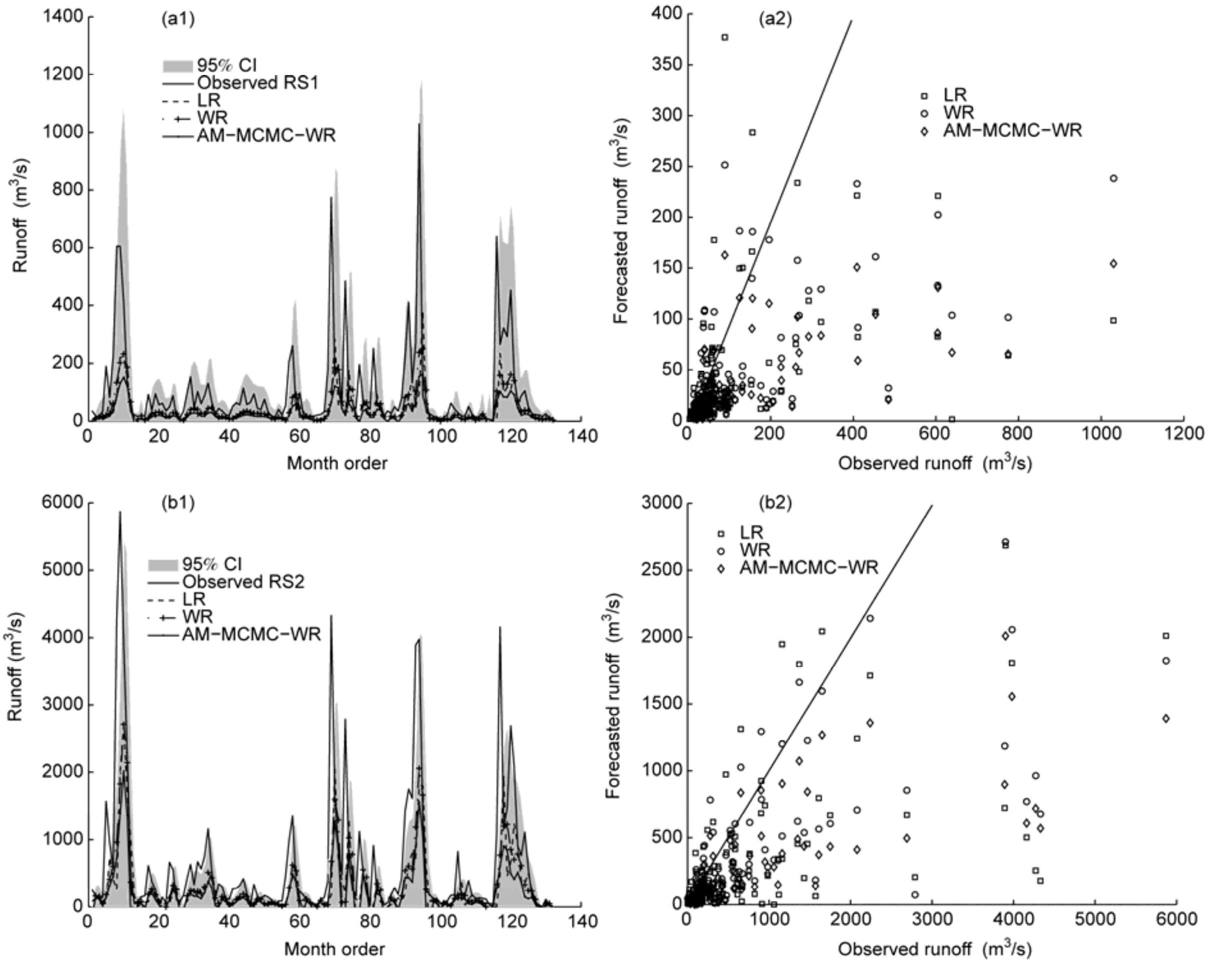

Figure 6 The results of three-month-ahead forecasting of RS1 (a1) and RS2 (b1) series in the verification period, and scatterplots of RS1 (a2) and RS2 (b2) series by LR, WR, and AM-MCMC-WR models.

easy predictability. However, those hydrologic series with bad correlation characters would have hard predictability when using linear regression models.

Hydrologic series usually show non-stationary variations at multi-temporal scales. When applying linear regression model to directly analyze raw hydrologic series, accurate forecasting results cannot be gained due to the impacts of noise and insignificant components in original series. However, when firstly separating the dominant deterministic components by wavelet analysis, the above influences can be overcome, and hydrologic time series forecasting results can be improved. Because the AM-MCMC-WR model is based on wavelet regressive modeling process, it has the identical performance as WR models in obtaining optimal forecasting results. Here, it is concluded that wavelet analysis can helpfully guide and improve the accuracy of hydrologic time series forecasting.

4 Conclusions

The AM-MCMC algorithm and the wavelet regression model were used together, and an AM-MCMC-WR model for hydrologic time series forecasting along with uncertainty evaluation was proposed. By applying to the Xixian and Bengbu hydrological stations in the Huaihe River watershed for one-month- and three-month-ahead forecasting, the performance of the proposed model was investigated. The results indicated that uncertainties of parameters and hydrologic time series forecasting are objective existences. Compared with conventional linear regression and wavelet regression models which are to estimate the optimal forecasting result, the proposed AM-MCMC-WR model is to do probabilistic forecasting, so uncertainty can be quantitatively evaluated using proper credible interval. In practical application, several detailed issues should be carefully considered when using the proposed model, such as accurate wavelet decomposition of series, and determination of proper wavelet regression model. Only after carefully solving these issues, could reasonable result of hydrologic time series forecasting be obtained.

Although the AM-MCMC-WR model proposed can do probabilistic forecasting, it has difficulty in accurately forecasting hydrologic extreme values, as clearly shown in the results of two examples. For RS1 and RS2 series, although the value ranges of the calibration data set can contain those of the verification data sets, accurate forecasting results of runoff extreme values cannot be obtained by any of the three models used. Factually, forecasting of hydrologic extreme values is a difficult task [3,25]. In the theories of stochastic hydrology [26,27], deterministic components and extreme values in hydrologic series are generated by different mechanisms [28]. Hydrologic extreme values usually show pure random characters and thus could not be accurately forecasted by any deterministic models. To improve the forecasting result, it may be more feasible and desirable by separating hydrologic extreme values first and then describing them by proper statistical models.

Finally, it should be pointed out that this study mainly studied parameter uncertainty, but data uncertainty and model uncertainty should be further studied. By doing this, uncertainty of hydrologic time series forecasting can be more comprehensively taken into consideration.

The authors gratefully acknowledged the valuable review comments and suggestions given by the editor, Li Zhang, and the anonymous reviewers. The authors also thanked Liu Feifei for her assistance in preparation of the manuscript. This work was supported by the National Natural Science Foundation of China (*41201036*) *and the Opening Fund of Key Laboratory for Land Surface Process and Climate Change in Cold and Arid Regions, Chinese Academy of Sciences* (*LPCC201203*).

1 Cheng C T, Xie J X, Chau K W, et al. A new indirect multi-step-ahead forecasting model for a long-term hydrologic forecasting. J Hydrol, 2008, 361: 118–130

2 Tiwari M K, Chatterjee C. Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). J Hydrol, 2010, 382: 20–33

3 Jain A, Kumar A M. Hybrid neural network models for hydrologic time series forecasting. Appl Soft Comput, 2007, (7): 585–592

4 Kwon H H, Lall U, Khalil A F. Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. Water Resour Res, 2006, 42: W11404

5 Luk K C, Ball J E, Sharma A. A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. J Hydrol, 2000, 227: 56–65

6 Kisi O. Wavelet regression model for short-term streamflow forecasting. J Hydrol, 2010, 389: 344–353

7 Zhang B, Govindaraju R S. Forecasting of watershed runoff using Bayesian concepts and modular neural networks. Water Resour Res, 2000, 36: 753–762

8 Labat D. Recent advances in wavelet analyses: Part 1. A review of concepts. J Hydrol, 2005, 314: 275–288

9 Sang Y F. A review on the applications of wavelet transform in hydrologic time series analysis. Atmos Res, 2013, 122: 8–15

10 Wang W S, Ding J. Wavelet network model and its application to the predication of hydrology. Nat Sci, 2003, 1: 67–71

11 Nourani V, Mogaddam A A, Nadiri A O. An ANN-based model for spatiotemporal groundwater level forecasting. Hydrol Process, 2008, 22: 5054–5066

12 Arabi M, Govindaraju R S, Hantush M M. A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. J Hydrol, 2007, 333: 459–471

13 Geza M, Poeter E P, McCray J E. Quantifying predictive uncertainty for a mountain-watershed model. J Hydrol, 2009, 376: 170–181

14 Krzysztofowicz R. The case for probabilistic forecasting in hydrology. J Hydrol, 2001, 249: 2–9

15 Kumar D N, Maity R. Bayesian dynamic modeling for nonstationary hydroclimatic time series forecasting along with uncertainty quantification. Hydrol Process, 2008, 22: 3488–3499

16 Biemans H, Hutjes R W A, Kabat P, et al. Effects of precipitation uncertainty on discharge calculations for Main River Basins. J Hydromet, 2009, 10: 1011–1025

17 Bormann H. Sensitivity of a soil-vegetation-atmosphere-transfer scheme to input data resolution and data classification. J Hydrol, 2008, 351: 154–169

18 Tung W H, Mays L W. Risk models for flood levee design. Water Resour Res, 1981, 17: 833–841

19 Kaheil Y H, Gill M K, McKee M, et al. A new Bayesian recursive technique for parameter estimation. Water Resour Res, 2006, 42: W08423

20 Congdon P. Bayesian Statistical Modeling. West Sussex: Wiley Press, 2001

21 Reggiani P, Weerts A H. A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine. J Hydrol, 2008, 356: 56–69

22 Sang Y F, Wang D, Wu J C. Probabilistic forecast and uncertainty assessment of hydrologic design values using Bayesian theories. Hum Ecol Risk Assess, 2010, 16: 1184–1207

23 Sang Y F. A practical guide to discrete wavelet decomposition of hydrologic time series. Water Resour Manag, 2012, 26: 3345–3365

24 Vrugt J A, Gupta H V, Bouten W, et al. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour Res, 2003, 39: W1201

25 Minns A W, Hall M J. Artificial neural networks as rainfall-runoff models. Hydrol Sci J, 1996, 41: 399–417

26 Yevjevich V. Stochastic Process in Hydrology. Colorado: Water Resources Publications, 1972

27 Sang Y F, Wang Z G, Liu C M. Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis. J Hydrol, 2012, 424: 154–164

28 Kuczera G. Uncorrelated measurement error in flood frequency inference. Water Resour Res, 1992, 28: 183–188