

# Improved Wavelet Modeling Framework for Hydrologic Time Series Forecasting

Yan-Fang Sang

Received: 22 October 2012 / Accepted: 19 February 2013 /  
Published online: 3 March 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** The combination of wavelet analysis with black-box models presently is a prevalent approach to conduct hydrologic time series forecasting, but the results are impacted by wavelet decomposition of series, and uncertainty cannot be evaluated. In this paper, the method for discrete wavelet decomposition of series was developed, and an improved wavelet modeling framework, WMF for short, was proposed for hydrologic time series forecasting. It is to first separate different deterministic components and remove noise in original series by discrete wavelet decomposition; then, forecast the former and quantitatively describe noise's random characters; at last, add them up and obtain the final forecasting result. Forecasting of deterministic components is to obtain deterministic forecasting results, and noise analysis is to estimate uncertainty. Results of four hydrologic cases indicate the better performance of the proposed WMF compared with those black-box models without series decomposition. Because of having reliable hydrologic basis, showing high effectiveness in accuracy, eligible rate and forecasting period, and being capable of uncertainty evaluation, the proposed WMF can improve the results of hydrologic time series forecasting.

**Keywords** Hydrologic time series forecasting · Wavelet · Black-box model · Noise · Statistical analysis · Uncertainty

---

Y.-F. Sang (✉)

Key Laboratory of Water Cycle & Related Land Surface Processes,  
Institute of Geographic Sciences and Natural Resources Research,  
Chinese Academy of Sciences, Beijing 100101, China  
e-mail: sangyf@igsnr.ac.cn

Yan.-Fang. Sang

e-mail: sunsangyf@gmail.com

Y.-F. Sang

Key Laboratory for Land Surface Process and Climate Change in Cold and Arid Regions,  
Chinese Academy of Sciences, Lanzhou 730000, China

## 1 Introduction

Hydrologic forecasting is to reveal the future hydrologic regimes, and further provide useful information for practical water activities (Tiwari and Chatterjee 2010; Dutta et al. 2012). In nature, hydrologic processes show highly complex characteristics due to the influences of many and always interrelated physical factors (Labat et al. 2000). Moreover, climatic change (Hanson et al. 2004) and human activities (Carsten et al. 2008) add the complexities of hydrologic processes by changing land surface conditions. Although great efforts were devoted by researchers, enough understanding of hydrologic processes has not been gained presently. The models used currently cannot always meet practical needs enough due to their limited applicable ranges and defects, causing the difficulty in accurately hydrologic forecasting (Tiwari and Chatterjee 2010).

Generally, the models used for hydrologic forecasting can be divided into two types: mathematical–physical models and black-box models (BBMs). The former mainly use a series of mathematical equations to describe physical hydrologic processes, so they usually need a large amount of data for calibration and validation purposes, and are computationally extensive (Arora 2002). However, the need of enough data cannot be always met, especially in developing countries where recorded data lengths are usually short. Comparatively, BBMs are effective alternatives in many practical situations because of having two advantages as low quantitative data demands and simple formulation (Jain and Kumar 2007). Hydrologic forecasting by black-box models is usually called hydrologic time series forecasting, and it is mainly considered in this study.

For overcoming the black-box defects of BBMs, many methods are often combined with BBMs to conduct hydrologic series forecasting. Present studies and practical applications (reviewed in Section 2) indicate the better performance of hybrid models rather than single model. Among the former the combination of wavelet analysis with black-box models has become a prevalent approach to forecast monthly precipitation (Nourani et al. 2009), monthly runoff (Kisi 2008, 2009b), shallow groundwater level and daily discharge (Wang and Ding 2003), rainfall-runoff (Coulibaly and Burn 2004), droughts (Kim and Valdes 2003), monthly lake level (Kisi 2009a), and other hydrologic variables. Wavelet analysis can elaborate the localized characteristics of a series both in temporal and frequency domains (Percival and Walden 2000), so it can handle series' non-stationary characteristics and then guide hydrologic time series forecasting (Nourani et al. 2008).

However, the effectiveness of wavelet models is impacted by several key but presently unsolved issues. The most critical one is accurate wavelet decomposition of series, as the basis of wavelet aided hydrologic time series forecasting. Previous studies usually chose the decomposition level for wavelet decomposition as  $\log_{10}n$  ( $n$  is series' length) (Aussem et al. 1998; Wang and Ding 2003). It may be unreasonable because of just depending on series' length but cannot considering series' composition (Nourani et al. 2011). Wavelet choice and wavelet threshold de-noising also impact the accuracy of wavelet decomposition (Sang et al. 2009a). Another key issue is uncertainty evaluation. Uncertainty is the objective existence in hydrologic forecasting (Krzysztofowicz 2001; Ghosh and Katkar 2012); previous studies of wavelet models mainly focused on improving the accuracy of forecasting result, but there are lacks of studies on uncertainty.

The main objective of this study is to propose an improved wavelet modeling framework for hydrologic time series forecasting. The methods for wavelet decomposition of series were first improved by discussing several key factors, by which different deterministic components and noise in hydrologic series can be separated. Based on wavelet decomposition of series, the improved wavelet modeling framework (WMF) was proposed. Finally, the

improved WMF was verified by analyzing four cases. The conclusions are presented in the final section.

## 2 Review on Black-Box Models

Black-box models use certain mathematical tools to reveal the correlations of hydrologic variables, but do not consider physical hydrologic processes. Serial-analysis techniques (SAT) are the typical black-box models and develop extensively, but they cannot accurately describe hydrologic series with nonlinear and non-stationary characteristics. Another typical kind of black-box models is the neural models which also have gained increasing applications in hydrology. Neural models are supposed to possess the ability of reproducing the nonlinear relationships between input explanatory variables and output forecasted variables (Jain and Kumar 2007). Many studies have proved that they can model the hydrologic processes by a proper three-layer neural network with input-, hidden- and output-layer, and can perform better than SATs (Luk et al. 2000). By summarizing previous studies and applications of black-box models in hydrology, some personal understandings and conclusions are gained as follows:

- (1) **Series decomposition should precede forecasting.** An important understanding is accurate result of hydrologic time series forecasting usually cannot be gained if directly analyzing raw hydrologic series by black-box models (Jain and Kumar 2007). It is mainly due to the multi-components of hydrologic series (Labat 2005; Sang et al. 2012). Therefore, a good knowledge of hydrologic series' composition should be first gained, or else the forecasting process would lack reliable hydrologic basis. However, most of black-box models used presently are not based on the composition of hydrologic series (Nourani et al. 2009). To improve hydrologic time series forecasting, accurate decomposition of series should precede the forecasting processes.
- (2) **Deterministic components and noise should be forecasted respectively.** Another major understanding is black-box models cannot satisfactorily forecast hydrologic extrema (Zhang and Govindaraju 2000). In the theories of stochastic hydrology (Yevjevich 1972), hydrologic series is basically composed of deterministic components and noise. The former are generated by certain deterministic physical mechanisms, and show deterministic characteristics as periods and trend; noise is generated by many random and uncertain factors, and shows pure random characters. Noise contaminates deterministic components and is closely relevant to hydrologic extrema (Kuczera 1992). Therefore, hydrologic extrema usually show pure random characters and cannot be accurately forecasted by deterministic models. To improve the result of hydrologic time series forecasting, deterministic components and noise should be first separated and then forecasted respectively.
- (3) **Uncertainty evaluation should be studied.** Previous studies of hydrologic time series forecasting mainly focused on improving the accuracy of forecasting result, but there is a lack of study on uncertainty evaluation. Hydrologic processes like any other natural processes have uncertainty (Arabi et al. 2007), so the forecasting result with a single optimal value is not convincing and "honest" (Sang et al. 2010). Uncertainty evaluation should be studied to obtain more reasonable forecasting results.
- (4) **Hybrid model performs better than single model.** Accurate result of hydrologic time series forecasting usually cannot be obtained by any model alone. To overcome the defects of single model, many new techniques have been applied in hydrologic time

series forecasting, such as information theory (Bagtzoglou and Hossain 2009), fuzzy theory (Gao and Er 2005; Jeong et al. 2012) and wavelet theory (Coulibaly and Burn 2004). Various analyses have demonstrated the better performances of hybrid models compared with single model, and it is mainly due to the complementary effects among different methods.

- (5) **Wavelet models should be developed.** Wavelet models is one kind of the popular and widely used hybrid models, since wavelet analysis can remove noise and identify different deterministic components in hydrologic series, based on which the process of hydrologic time series forecasting is more guidable. However, several key issues impact the effectiveness of wavelet models, including (a) accurate wavelet decomposition of series is impacted by many key factors (Sang 2012); (b) hydrologic extrema still cannot be accurately forecasted if directly analyzing raw series by wavelet models, as clearly shown in various cases (Kisi 2009b; Nourani et al. 2009; etc.); (c) uncertainty evaluation is still an open issue when using wavelet models. These issues will be studied in the following sections, mainly to improve wavelet aided hydrologic time series forecasting.

### 3 Wavelet Decomposition of Series

The issue of wavelet decomposition of series was studied by Sang (2012). He first established the reference energy function for discrete wavelet analysis by operating Monte-Carlo simulation; then, by comparing energy function of hydrologic series with the reference energy function, he proposed the methods for wavelet choice, decomposition level choice, wavelet threshold de-noising and significance testing of DWT; finally, he gave a step-by-step guide to discrete wavelet decomposition of hydrologic series.

The methods were briefly described as follows: (1) for the analyzed series  $f(t)$  with the length of  $n$ , choose proper wavelet and compute the decomposition level of  $\log_2 n$ , then apply dyadic DWT to it; (2) remove noise  $f_n(t)$  in  $f(t)$  by wavelet threshold de-noising (WTD):  $f(t) = f_d(t) + f_n(t)$ ; (3) determine the energy function of the de-noised series  $f_d(t)$  by computing the energy  $E_j = \sum_{t=1}^n (f_{dj}(t))^2$  of its sub-signal  $f_{dj}(t)$  under each level  $j$ ; (4) determine the energy function of noise and estimate proper confidence interval (e.g. 95 %) by operating Monte-Carlo simulation; (5) use the energy of the sub-signal  $f_{d1}(t)$  under the first level to rescale both the energy function of noise and confidence interval, and take the result as the reference energy function; (6) compare the energy function of  $f_{dj}(t)$  with the reference energy function, and certain  $f_{dj}(t)$  with the energy exceeding the confidence interval is taken as deterministic component; and (7) all deterministic components can be identified, and accurate wavelet decomposition of series  $f(t)$  is obtained finally. The method was used in this study for accurate wavelet decomposition of series, as the basis of wavelet aided hydrologic time series forecasting.

### 4 Improved Wavelet Modeling Framework Proposed

Hydrologic time series forecasting can be improved based on wavelet decomposition of series. As before stated, deterministic components and noise in original hydrologic series show obviously different characteristics, and they need to be analyzed by proper models respectively.

Deterministic components are analyzed by proper black-box model (such as SAT, neural model) to obtain the deterministic forecasting results. Three cases would be met in practice. First, forecasting of certain hydrologic variable by SAT models using its previous values; in this case, each deterministic components identified in original series need to be forecasted respectively. Second, forecasting of certain hydrologic variable by neural models using its previous values; in this case, deterministic components are used as the input data of neural network, and the number of deterministic components identified equals the input layer's neurons number. Third, forecasting of certain hydrologic variable by neural models using other explanatory variables, such as rainfall-runoff forecasting; in this case, each explanatory series need to be decomposed, and their total number equals the input layer's neurons number, and then the forecasted series need to be separated into deterministic component and noise, and the former is forecasted by the decomposed explanatory series.

Noise is the inevitable part of hydrologic data (Yevjevich 1972). They reflect the intrinsic uncertainty of hydrologic processes and generate hydrologic extrema. In hydrologic time series forecasting, it can be an effective approach to evaluate uncertainty: quantitatively describe the random characters of the noise separated from the forecasted series by proper statistical method, such as normal hypothesis testing, hydrologic frequency analysis, or the principle of maximum entropy (POME) theory (Singh 1998), and estimate the forecasting result of hydrologic extrema with proper guarantee rate (or confidence level). Specifically, the separated noise following skew distribution can be analyzed by hydrologic frequency analysis using proper probability density function, and the separated noise following normal distribution can be analyzed by normal hypothesis testing.

By summarizing the above results, an improved wavelet modeling framework, WMF for short, is proposed for hydrologic time series forecasting (Fig. 1). It includes four main steps: decomposition of series to separate noise and identify deterministic components; selection of appropriate model to forecast deterministic components; analysis of noise to evaluate uncertainty; and summarization of them to obtain the final forecasting result.

The specific steps of hydrologic time series forecasting by the improved WMF are explained as follows:

- (1) Choose proper wavelet and decomposition level, and then apply dyadic DWT to the analyzed series  $f(t)$ , by which wavelet coefficients of the series  $f(t)$  under each level are obtained;
- (2) Remove noise in  $f(t)$  by WTD. The result is expressed as:

$$f(t) = f_d(t) + f_n(t) \quad (1)$$

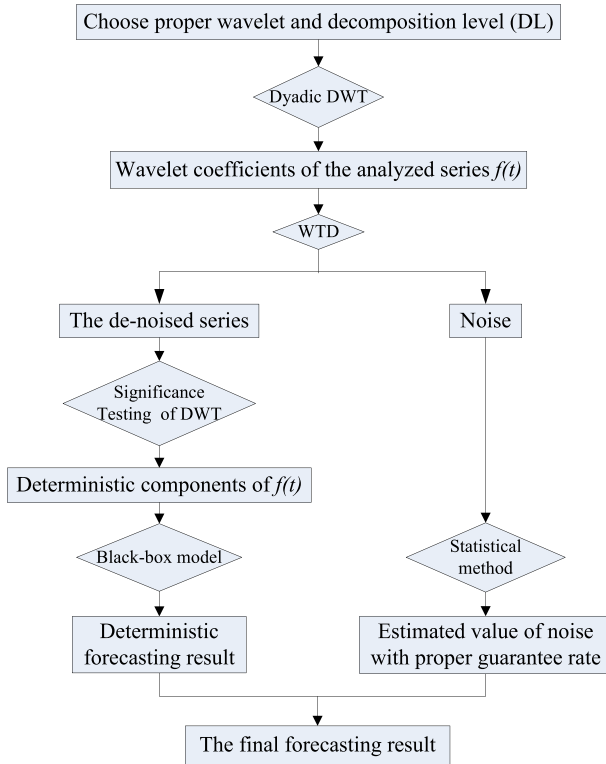
where  $f_d(t)$  and  $f_n(t)$  are the de-noised series and noise respectively;

- (3) Identify and separate deterministic components in series  $f_d(t)$  by conducting significance testing of DWT. The result is expressed as:

$$f(t) = f_d(t) + f_n(t) = \sum_{i=1}^l f_{di}(t) + f_n(t) \quad (2)$$

where  $f_{di}(t)$  is the  $i$ th deterministic component identified with the total number of  $l$ .

- (4) If the hydrologic series is forecasted using its previous values, it is decomposed according to the steps (1)–(3). If the hydrologic series is forecasted by other explanatory series, the latter are decomposed according to the steps (1) and (3), and the forecasted series is de-noised according to the steps (1) and (2).



**Fig. 1** Steps of hydrologic time series forecasting by the wavelet modeling framework (WMF) proposed. In Figure 1, “DWT” is discrete wavelet transform, and “WTD” is wavelet threshold de-noising

- (5) Select proper black-box model to forecast deterministic components in the forecasted series using the results in step (4), and obtain the deterministic forecasting result  $\hat{f}_d(t + \tau)$  with the forecasting period  $\tau$ .
- (6) Use proper statistical method to describe the random characters of noise data  $f_n(t)$ , and estimate the value  $\hat{f}_n(t)$  with proper guarantee rate (or confidence level).
- (7) Add up the two results, and obtain the final forecasting result  $\hat{f}(t + \tau)$  with proper guarantee rate:

$$\hat{f}(t + \tau) = \hat{f}_d(t + \tau) + \hat{f}_n(t) \tag{3}$$

## 5 Case Study

### 5.1 Data

Four cases were analyzed for verification purpose. These data were measured in different sampling rates and from different basins (Table 1). The case-I presents 20-year monthly runoff data measured at the Dashankou station in Northwest China, and it has two dominant periods of 6 and 12 months (Sang et al. 2009b). The case-II presents 41-year monthly precipitation data measured at the Nanjing weather station in East China, and it has a

**Table 1** Observed hydrologic data used in this paper

Case	Series	Type	Length	Hydrologic station	Location	Climate condition
Case-I	RS1	Monthly runoff	240 (1978–1997)	Dashankou	Xinjiang Province	Inland arid climate
Case-II	RS2	Monthly precipitation	492 (1961–2001)	Nanjing	Nanjing City	Subtropical monsoon climate
Case-III	RS3	Annual precipitation	41 (1961–2001)	Nantong	Nantong City	Subtropical monsoon climate
Case-IV	RS4	Daily coal mine discharge	125 (Jun 1 to Oct 3 in 2003)	Hanqiao	Xuzhou City	Temperate monsoon climate
	PS4	Daily rainfall	125 (Jun 1 to Oct 3 in 2003)	Hanqiao	Xuzhou City	Temperate monsoon climate

dominant period of 12 months due to the annual variability of hydrologic processes. The case-III presents 41-year annual precipitation data measured at the Nantong weather station at the outlet of the Changjiang River. The case-IV presents 125-day mine discharge data measured at the Hanqiao Coal Mine in the mid-east of China, and it is mainly determined by the local rainfall of PS4 series (Sang et al. 2009c).

## 5.2 Wavelet Decomposition Results

The wavelets used for analyzing these series were chosen according to de-noising results. The statistical characters of original series, de-noised series and separated noise, including mean ( $\bar{x}$ ), coefficient of variation ( $C_v$ ), coefficient of skewness ( $C_s$ ), lag-1 ( $r_1$ ) and lag-2 ( $r_2$ ) autocorrelation coefficients, were computed. Table 2 indicates that by using the chosen wavelets mean values of de-noised and original series are similar; but de-noised series'  $C_v$  values are smaller and  $r_1$  values are bigger than those of originals series; moreover, five separated noise series mainly show random characters by doing statistic hypothesis testing. Therefore, it is thought that de-noising results of these series by the chosen wavelets are reasonable.

Energy functions of these series are depicted in Fig. 2, and wavelet decomposition results are displayed in Fig. 3. RS1 series includes three deterministic components: the sub-signal under level 2 has the period of 6 months, the sum of sub-signals under levels 3–5 has the period of 12 months, and the residue is trend. RS2 series includes two deterministic components: the sub-signal under level 3 is the periodic component, and the residue is trend. RS3 series includes two deterministic components: the sub-signal under level 2 is the periodic component, and the residue is trend. For RS4 series, its sub-signals under the first two levels are composed of noise, but the residue is trend. PS4 series includes two deterministic components: the sum of sub-signals under the first six levels is the periodic component, and the residue is trend.

## 5.3 Forecasting Results by Different Models

Six factors were considered when operating forecasting: (1) data. Each data of forecasted series was divided into two parts (Table 3) for calibration and verification respectively; (2) edge effect. The point-symmetric extension method was used to handle double edges of series (Kharitonenko et al. 2002); (3) model. Five models, denoted as M1, M2, M3, M4 and M5, were used. Since back propagation neural network (BPNN) performs better than SAT models, all five models were based on the former. For any series, it is directly forecasted when using M1 model; de-noised but un-decomposed, and then forecasted when using M2 model; decomposed

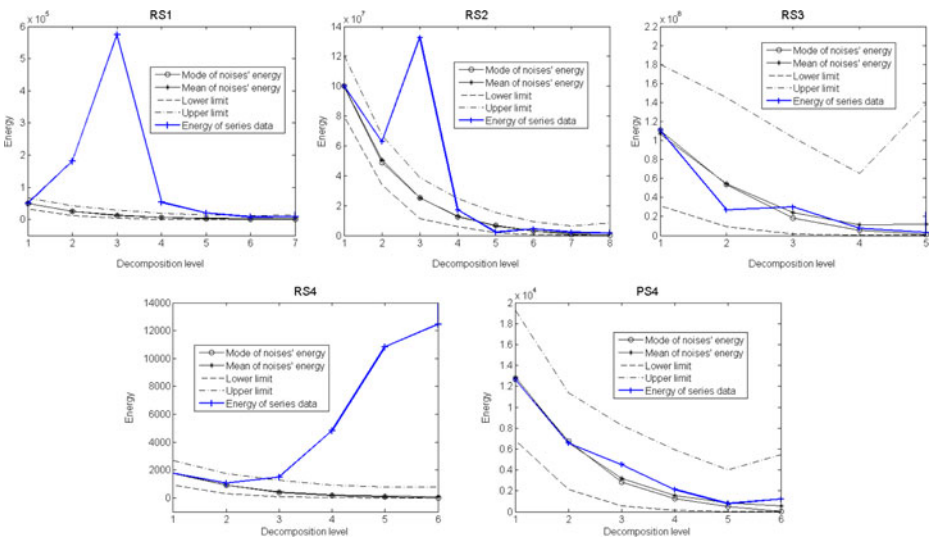


**Table 2** Statistical characters of the de-noising results of five series

Series	Type	Statistical characters					SNR	Chosen wavelet
		$\bar{X}$	$\sigma$	$C_s$	$r_1$	$r_2$		
RS1	X	101.55	61.24	0.99	0.73	0.35	23.03	db6
	S <sup>-</sup>	100.94	49.86	0.79	0.79	0.39		
	N	0.61	15.76	0.77	0.30	0.08		
RS2	X	86.85	81.10	2.21	0.29	0.11	1.52	coif5
	S <sup>-</sup>	86.84	57.82	1.04	0.84	0.52		
	N	0.01	53.59	0.88	-0.25	-0.31		
RS3	X	1054.80	222.01	0.38	-0.17	0.31	-2.54	db10
	S <sup>-</sup>	1055.10	146.50	0.07	0.70	0.15		
	N	-0.27	166.34	0.08	-0.22	0.14		
RS4	X	103.50	30.94	-0.32	0.96	0.93	39.59	sym5
	S <sup>-</sup>	101.41	30.19	-0.45	0.98	0.96		
	N	2.08	4.17	1.05	0.30	0.02		
PS4	X	6.97	15.42	3.15	0.12	0.09	15.20	sym5
	S <sup>-</sup>	6.82	11.25	3.48	0.17	0.08		
	N	0.15	5.26	1.40	0.13	0.07		

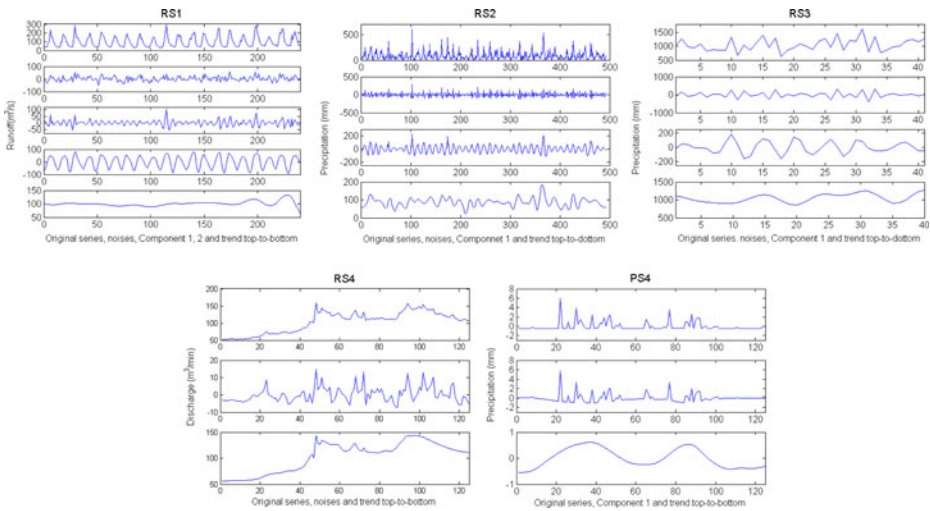
$X$ ,  $S^-$  and  $N$  are the original series, de-noised series and separated noise respectively. The statistical characters include mean ( $\bar{x}$ ), coefficient of variation ( $C_v$ ), coefficient of skewness ( $C_s$ ), lag-1 ( $r_1$ ) and lag-2 ( $r_2$ ) autocorrelation coefficient. SNR is signal-to-noise ratio and is calculated as  $-20 \log(\text{Var}(S^-)/\text{Var}(N))$ , in which “Var” means calculating variance

by the decomposition level of  $\log_{10}n$  but not de-noised, and then forecasted when using M3 model; decomposed by significance testing of DWT but not de-noised, and then forecasted when using M4 model; de-noised and decomposed, and then forecasted when using M5 model



**Fig. 2** Energy functions of the series used in this paper, and the reference energy function with 95 % confidence interval





**Fig. 3** Wavelet decomposition results of the series used in this paper

(i.e., the proposed WMF); (4) model structure. The number of input layer’s neurons was determined according to series’ correlations when using M1 and M2 models, but determined according to series’ decomposition results when using M3, M4, and M5 models. The number of hidden layer’s neurons was selected by the trial-error procedure; (5) forecasting period. Both 1-step and 6-step ahead forecasting were conducted; (6) results evaluation. Three indexes, *AARE* (average absolute relative error), *TS<sub>x</sub>* (threshold statistics) and *R<sup>2</sup>* (coefficient of determination), were used to evaluate the results:

$$\begin{aligned}
 AARE &= \frac{1}{N} \sum_{i=1}^N \left| \frac{f'(i) - f(i)}{f(i)} \right| \times 100\% \\
 TS_x &= \frac{n_x}{N} \times 100\% \\
 R^2 &= \frac{\sum_{i=1}^N (f(i) - f'(i))^2}{\sum_{i=1}^N (f(i) - f'(i))^2 + \sum_{i=1}^N (f'(i) - \overline{f'(i)})^2}
 \end{aligned}
 \tag{4}$$

where *f(i)* is observed data and *f'(i)* is forecasted data with the total number of *N*, *n<sub>x</sub>* is the number of data points with the *AARE* values being smaller than *x*%;  $\overline{f'(i)}$  is the mean of *f'(i)*.

**Table 3** Data used in the forecasting processes

Forecasted series	Total length	Training data	Validating data	Model	
				Input data	Output data
RS1	240 months (1978–1997)	192 (1978–1993)	48 (1994–1997)	Previous runoff	Future runoff
RS2	492 months (1961–2001)	372 (1961–1991)	120 (1992–2001)	Previous precipitation	Future precipitation
RS3	41 years (1961–2001)	31 (1961–1991)	10 (1992–2001)	Previous precipitation	Future precipitation
RS4	125 days (Jun 1 to Oct 3 in 2003)	95 (Jun 1 to Sep 3)	30 (Sep 4 to Oct 3)	Precipitation (PS4)	Mine discharge

The 1-step and 6-step ahead forecasting results of each case obtained by five models were evaluated, and the results are shown in Tables 4, and 5, in which the structures of BPNN are also presented.

The noise separated from RS1, RS2 and RS4 series have the  $C_s$  values bigger than 0.7 (Table 2), and they do not follow normal distribution by dong statistical hypothesis testing. Therefore, these noise series were described by the Pearson-III (P-III) distribution and then analyzed by hydrologic frequency analysis according to China's general practices. The noise in RS3 series with the  $C_s$  value of 0.08 follows normal distribution accosting to the statistical testing results, and it was analyzed by normal hypothesis testing. The estimated noise values with various guarantee rates are shown in Table 6.

The final 1-step and 6-step ahead forecasting results of four cases by the proposed WMF are depicted in Fig. 4. More details were discussed in the following section.

### 6 Results Discussion

This study improved the understanding about the relationship between hydrologic series' composition and hydrologic time series forecasting, and proposed an improved wavelet

**Table 4** Evaluation of the 1-step ahead forecasting results of four cases obtained by different models

Forecasted series	Model used	Model structure	Index (calibration)					Index (verification)				
			AARE	TS <sub>5</sub>	TS <sub>10</sub>	TS <sub>20</sub>	R <sup>2</sup>	AARE	TS <sub>5</sub>	TS <sub>10</sub>	TS <sub>20</sub>	R <sup>2</sup>
RS1	M1	6-9-1	0.16	0.08	0.15	0.47	0.75	0.28	0.00	0.00	0.12	0.58
	M2	6-9-1	0.10	0.35	0.59	0.77	0.88	0.20	0.16	0.25	0.52	0.79
	M3	3-8-1	0.13	0.18	0.33	0.54	0.83	0.25	0.02	0.06	0.18	0.58
	M4	6-11-1	0.12	0.24	0.38	0.52	0.85	0.24	0.08	0.15	0.32	0.65
	M5	3-5-1	<b>0.09</b>	<b>0.38</b>	<b>0.58</b>	<b>0.78</b>	<b>0.91</b>	<b>0.16</b>	<b>0.22</b>	<b>0.33</b>	<b>0.55</b>	<b>0.83</b>
RS2	M1	6-13-1	0.23	0.08	0.21	0.31	0.63	0.35	0.00	0.10	0.20	0.41
	M2	6-13-1	0.15	0.32	0.42	0.72	0.86	0.22	0.10	0.30	0.47	0.67
	M3	3-8-1	0.17	0.09	0.16	0.53	0.76	0.23	0.20	0.20	0.40	0.53
	M4	4-6-1	0.16	0.10	0.17	0.58	0.74	0.24	0.20	0.30	0.40	0.57
	M5	2-5-1	<b>0.11</b>	<b>0.35</b>	<b>0.57</b>	<b>0.70</b>	<b>0.88</b>	<b>0.18</b>	<b>0.20</b>	<b>0.30</b>	<b>0.50</b>	<b>0.73</b>
RS3	M1	4-4-1	0.12	0.37	0.47	0.63	0.85	0.24	0.02	0.05	0.13	0.57
	M2	4-4-1	0.08	0.43	0.66	0.74	0.94	0.10	0.11	0.19	0.45	0.76
	M3	2-5-1	0.12	0.30	0.60	0.73	0.92	0.17	0.08	0.13	0.32	0.63
	M4	3-6-1	0.11	0.32	0.53	0.77	0.93	0.15	0.07	0.20	0.37	0.67
	M5	2-4-1	<b>0.04</b>	<b>0.42</b>	<b>0.70</b>	<b>0.85</b>	<b>0.96</b>	<b>0.08</b>	<b>0.25</b>	<b>0.29</b>	<b>0.59</b>	<b>0.84</b>
RS4	M1	4-7-1	0.12	0.16	0.26	0.30	0.77	0.18	0.00	0.00	0.00	0.57
	M2	4-7-1	0.03	0.32	0.57	0.78	0.95	0.08	0.30	0.46	0.67	0.83
	M3	3-6-1	0.05	0.23	0.51	0.73	0.89	0.15	0.13	0.20	0.37	0.76
	M4	7-8-1	0.03	0.35	0.53	0.77	0.93	0.14	0.10	0.23	0.43	0.79
	M5	2-7-1	<b>0.03</b>	<b>0.55</b>	<b>0.85</b>	<b>0.91</b>	<b>0.95</b>	<b>0.05</b>	<b>0.50</b>	<b>0.87</b>	<b>0.95</b>	<b>0.91</b>

M1 is a single BPNN; M2 is a wavelet de-noising-based BPNN, M3 and M4 are wavelet decomposition-based BPNNs, while M5 is a wavelet de-noising and decomposition-based BPNN

**Table 5** Evaluation of the 6-step ahead forecasting results of four cases obtained by different models

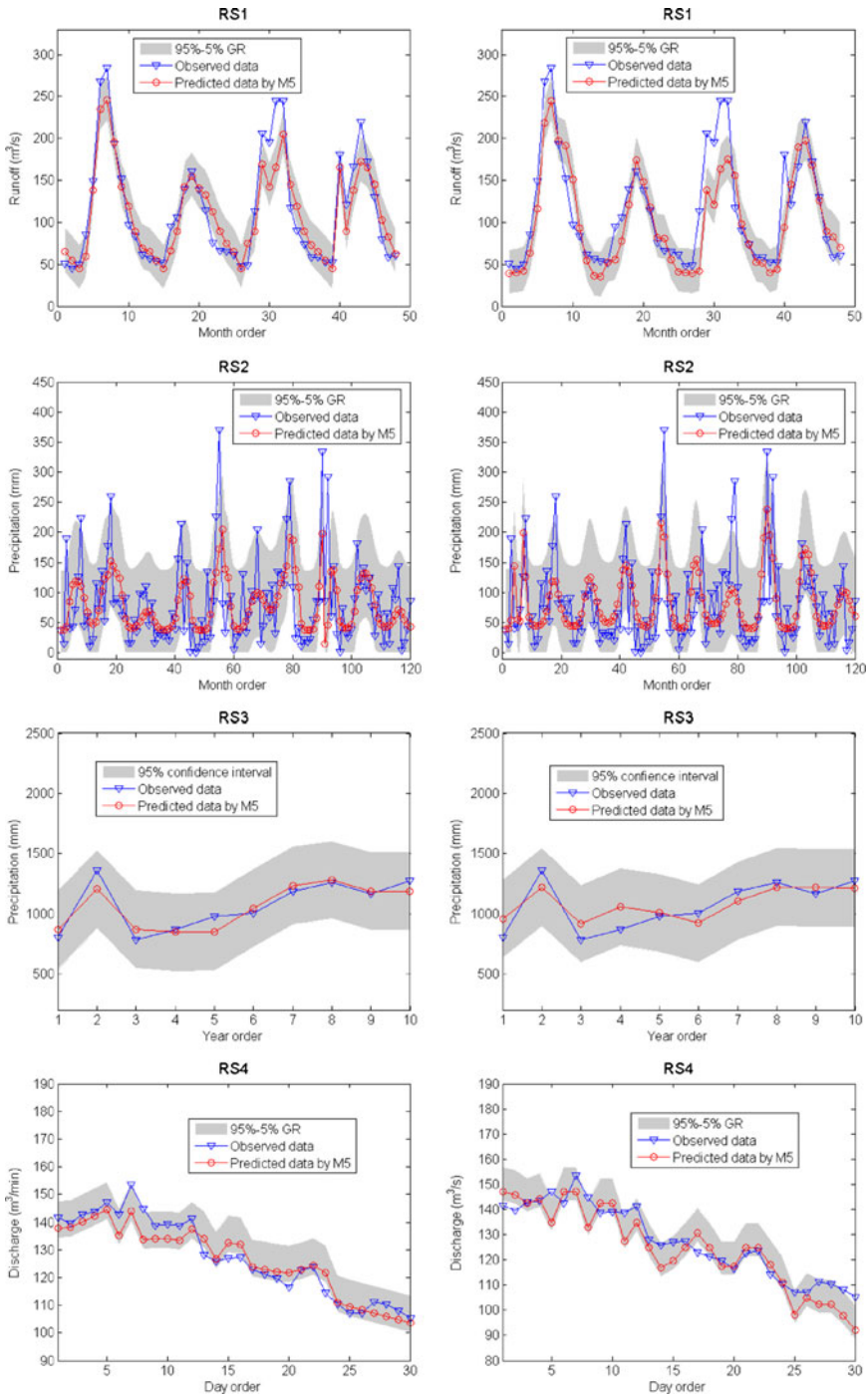
Forecasted series	Model used	Model structure	Index (calibration)					Index (verification)				
			AARE	TS <sub>5</sub>	TS <sub>10</sub>	TS <sub>20</sub>	R <sup>2</sup>	AARE	TS <sub>5</sub>	TS <sub>10</sub>	TS <sub>20</sub>	R <sup>2</sup>
RS1	M1	6-9-1	0.19	0.10	0.16	0.41	0.70	0.29	0.00	0.02	0.06	0.43
	M2	6-9-1	0.11	0.32	0.55	0.73	0.85	0.22	0.06	0.10	0.35	0.66
	M3	3-8-1	0.14	0.12	0.29	0.56	0.82	0.27	0.04	0.04	0.10	0.55
	M4	6-11-1	0.14	0.23	0.42	0.53	0.83	0.24	0.02	0.12	0.20	0.63
	M5	3-5-1	<b>0.10</b>	<b>0.25</b>	<b>0.59</b>	<b>0.77</b>	<b>0.88</b>	<b>0.19</b>	<b>0.16</b>	<b>0.33</b>	<b>0.53</b>	<b>0.75</b>
RS2	M1	6-13-1	0.24	0.12	0.21	0.33	0.62	0.37	0.00	0.00	0.10	0.36
	M2	6-13-1	0.15	0.27	0.41	0.67	0.82	0.23	0.07	0.24	0.32	0.51
	M3	3-8-1	0.18	0.07	0.18	0.35	0.69	0.34	0.10	0.17	0.30	0.45
	M4	4-6-1	0.17	0.07	0.17	0.49	0.75	0.31	0.10	0.20	0.29	0.48
	M5	2-5-1	<b>0.12</b>	<b>0.16</b>	<b>0.50</b>	<b>0.67</b>	<b>0.85</b>	<b>0.21</b>	<b>0.23</b>	<b>0.26</b>	<b>0.47</b>	<b>0.69</b>
RS3	M1	4-4-1	0.17	0.30	0.47	0.61	0.81	0.25	0.03	0.07	0.10	0.52
	M2	4-4-1	0.10	0.45	0.62	0.71	0.91	0.16	0.08	0.20	0.32	0.67
	M3	2-5-1	0.13	0.26	0.54	0.72	0.86	0.21	0.05	0.13	0.23	0.58
	M4	3-6-1	0.13	0.30	0.52	0.74	0.89	0.20	0.04	0.18	0.23	0.60
	M5	2-4-1	<b>0.06</b>	<b>0.48</b>	<b>0.62</b>	<b>0.78</b>	<b>0.93</b>	<b>0.10</b>	<b>0.18</b>	<b>0.27</b>	<b>0.53</b>	<b>0.77</b>
RS4	M1	4-7-1	0.10	0.16	0.23	0.33	0.74	0.22	0.05	0.07	0.17	0.54
	M2	4-7-1	0.06	0.34	0.56	0.72	0.89	0.09	0.30	0.43	0.51	0.75
	M3	3-6-1	0.13	0.29	0.51	0.67	0.81	0.16	0.03	0.17	0.40	0.63
	M4	7-8-1	0.11	0.31	0.56	0.71	0.88	0.14	0.13	0.20	0.41	0.69
	M5	2-7-1	<b>0.03</b>	<b>0.33</b>	<b>0.65</b>	<b>0.77</b>	<b>0.98</b>	<b>0.07</b>	<b>0.25</b>	<b>0.53</b>	<b>0.76</b>	<b>0.88</b>

modeling framework, by which deterministic forecasting result can be gained, and uncertainty can be evaluated conveniently. By comparing the forecasting results of four cases using various models, it can be found that:

**Table 6** Estimated values of noise with various guarantee rates by statistical methods\*

Noise	Method used	Estimated results of parameters			Estimated value of noise		
		$\bar{X}$	$\sigma$	C <sub>s</sub>	5 %	50 %	95 %
In RS1	Hydrologic frequency analysis (P-III)	0.61	15.76	0.77	29.50	-20.03	-24.23
In RS2	Hydrologic frequency analysis (P-III)	0.01	53.59	0.88	99.26	-8.51	-73.09
In RS3*	Normal hypothesis testing	-0.27	166.34		-11.08~10.54	-112.56~112.01	-326.30~325.69
In RS4	Hydrologic frequency analysis (P-III)	2.08	4.17	1.05	9.94	1.36	-3.31

\*The results about the noise in RS3 series are value ranges with various confidence levels. "P-III" is the Pearson-III probability distribution



**Fig. 4** 1-step (*left*) and 6-step (*right*) ahead forecasting results of four cases by different models. In Figure 4, M1 is a single BPNN; M2 is a wavelet de-noising-based BPNN, M3 and M4 are wavelet decomposition-based BPNNs, whereas M5 is a wavelet de-noising and decomposition-based BPNN. “GR” is the guarantee rate

- (1) Hydrologic time series forecasting can be improved by first removing noise in the forecasted series. Forecasting results of four cases are not good enough when using M1, M3 and M4 models, and it is due to the noise impacts. However, when first removing noise from original series, forecasting results can be improved by M2 and M5 models.
- (2) Accurate series decomposition can improve hydrologic time series forecasting. Among the five models used, M3, M4 and M5 models are based on wavelet decomposition of series, and their results are much better than those by M1 model. The results demonstrate that deterministic components in hydrologic series show different characteristics, so they should be first separated and then forecasted respectively. In addition, M3 model performs worse than M4 and M5 models due to the unreasonable wavelet decomposition result using the decomposition level of  $\log_{10}n$ .
- (3) Uncertainty can be evaluated by the proposed WMF. By estimating the noise values with certain guarantee rate (Table 6), uncertainty of forecasting results can be quantitatively evaluated. Figure 4 indicates that most of the observed data fall within the range estimated by the forecasting results with 5 % and 95 % guarantee rates (95 % confidence interval for case-III) using M5 model, thus the uncertainty of forecasting result was effectively evaluated. It is concluded that de-noising does not impact the accuracy of deterministic forecasting results, but provides a feasible approach for uncertainty evaluation.
- (4) Forecasting results become worse with forecasting period increasing. The 6-step ahead forecasting results by any models are worse than 1-step ahead forecasting results (Tables 4, and 5), but this worse is not so perceptible when using M5 model. Therefore, it is thought that hydrologic forecasting, especially those with longer forecasting periods, can be improved by the proposed WMF.
- (5) Forecasting results of four cases show difference. Among the four cases, results of case-IV are the best, those of case-I and case-III followed, but the results of case-II are the worst. Forecasting results of any cases by M5 model are more accurate than those by other four models. Thereby, it is thought the proposed WMF has wide applicability in hydrologic time series forecasting.
- (6) Performances of five models have big differences. Tables 4, and 5 clearly show the different performances of five models. Among them, the M1 model performs the worst, especially when the forecasted series is greatly impacted by noise and longer forecasting period is needed. The performances of M2, M3, M4 and M5 models are different due to their different wavelet results; the results by M5 and M2 model as a whole are better than those by M4 model, while the results by M3 model are the worst. On the whole, the performance of five models is M5, M2, M4, M3, and M1 from good to bad.
- (7) The proposed WMF performs better compared with other four models. Forecasting process by the proposed WMF is based on hydrologic series' composition and so has reliable hydrologic basis; moreover, the eligible rates of forecasting result are higher, and accurate forecasting result with longer forecasting period can also be obtained; besides, uncertainty can be estimated quantitatively, which can make the final forecasting result more reasonable.

## 7 Conclusions

In this paper, by first discussing several key issues on wavelet decomposition of series, an improved wavelet modeling framework was proposed for hydrologic time series forecasting. Analyses of four different cases verified its performance. By summarizing the study, it can

be found that identification and separation of multi-components (including deterministic components and noise) of hydrologic series is an effective approach to improve hydrologic series forecasting. Compared with conventional black-box models, the WMF proposed shows three advantages: accurate decomposition of series that makes the forecasting process more guidable, higher effectiveness in accuracy and eligible rate and longer forecasting period, and the ability of uncertainty evaluation.

Nonetheless, more attention should still be paid to three issues in practice. First, accurate de-noising result is the prerequisite when using the proposed WMF, so noise should be further studied and more effective de-noising methods are needed. Second, conventional methods should be improved. Since black-box models are used to forecast deterministic components in original series when using the proposed WMF, forecasting result is influenced by their defects, such as determination of neural models' structures and parameters estimation; to improve hydrologic time series forecasting, defects of conventional black-box models should be further solved. In addition, for wavelet transform with down-sampling we could get quite different wavelet coefficients even if only shifting the analyzed series a few sample points, so the bad behavior of down-sampling should be carefully considered in discrete wavelet analysis of series, and the use of orthonormal multi-wavelet shell can avoid this problem (Chen et al. 2003). Third, when using the proposed WMF, specific model should be established based on both WMF and the hydrologic activity studied, and several detailed issues, such as wavelet decomposition of series and selection of proper black-box models, should also be carefully analyzed.

**Acknowledgments** The author gratefully acknowledged the helpful review comments and suggestions given by the Editor-in-Chief, George P. Tsakiris, and the anonymous reviewers. The author also thanked Ms. Feifei Liu for her assistance in the preparation of the manuscript. This project was supported by the National Natural Science Foundation of China (NSFC) (No. 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).

## References

- Arabi M, Govindaraju RS, Hantush MM (2007) A probabilistic approach for analysis of uncertainty in the evaluation of watershed management practices. *J Hydrol* 333(2–4):459–471
- Arora VK (2002) The use of the aridity index to assess climate change effect on annual runoff. *J Hydrol* 265:164–177
- Aussem A, Campbell J, Murtagh F (1998) Wavelet-based feature extraction and decomposition strategies for financial forecasting. *J Comput Intell Financ* 6(2):5–12
- Bagtzoglou AC, Hossain F (2009) Radial basis function neural network for hydrologic inversion: an appraisal with classical and spatio-temporal geostatistical techniques in the context of site characterization. *Stoch Env Res Risk A* 23(7):933–945
- Carsten M, Morton C, Ralf K, Gunter M, Harry V, Frank W (2008) Modeling the water balance of a mesoscale catchment basin using remotely sensed land cover data. *J Hydrol* 353:322–334
- Chen GY, Bui TD, Krzyzak A (2003) Contour-based handwritten numeral recognition using multiwavelets and neural networks. *Pattern Recog* 36(7):1597–1604
- Coulibaly P, Burn DH (2004) Wavelet analysis of variability in annual Canadian streamflows. *Water Resour Res* 40:W03105
- Dutta D, Welsh WD, Vaze J, Kim SSH, Nicholls D (2012) A comparative evaluation of short-term streamflow forecasting using time series analysis and rainfall-runoff models in eWater source. *Water Resour Manage* 26(15):4397–4415
- Gao Y, Er MJ (2005) NARMAX time series model forecasting: feed-forward and recurrent fuzzy neural network approaches. *Fuzzy Sets Syst* 150(2):331–350

- Ghosh S, Katkar S (2012) Modeling uncertainty resulting from multiple downscaling methods in assessing hydrological impacts of climate change. *Water Resour Manage* 26(12):3559–3579
- Hanson RT, Newhouse MW, Dettinger MD (2004) A methodology to assess relations between climatic variability and variations in hydrologic time series in the southwestern United States. *J Hydrol* 287:252–269
- Jain A, Kumar AM (2007) Hybrid neural network models for hydrologic time series forecasting. *Appl Soft Comput* 7:585–592
- Jeong C, Shin JY, Kim T, Heo JH (2012) Monthly precipitation forecasting with a neuro-fuzzy model. *Water Resour Manage* 26(15):4467–4483
- Kharitonenko I, Zhang X, Twelves S (2002) A wavelet transform with point-symmetric extension at tile boundaries. *IEEE T Image Process* 11(12):1357–1364
- Kim TW, Valdes JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *J Hydrol Eng* 8(6):319–328
- Kisi O (2008) River flow forecasting and estimation using different artificial neural network techniques. *Hydrol Res* 39(1):27–40
- Kisi O (2009a) Neural network and wavelet conjunction model for modeling monthly level fluctuations in Turkey. *Hydrol Process* 23:2081–2092
- Kisi O (2009b) Wavelet regression model as an alternative to neural networks for monthly streamflow forecasting. *Hydrol Process* 23:3583–3597
- Krzysztofowicz R (2001) The case for probabilistic forecasting in hydrology. *J Hydrol* 249:2–9
- Kuczera G (1992) Uncorrelated measurement error in flood frequency inference. *Water Resour Res* 28(1):183–188
- Labat D (2005) Recent advances in wavelet analyses: Part 1. A review of concepts. *J Hydrol* 314:275–288
- Labat D, Ababou R, Mangin A (2000) Rainfall-runoff relations for karstic springs. Part II: continuous wavelet and discrete orthogonal multiresolution analyses. *J Hydrol* 238:149–178
- Luk KC, Ball JE, Sharma A (2000) A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting. *J Hydrol* 227(1):56–65
- Nourani V, Komasi M, Mano A (2009) A multivariate ANN-wavelet approach for rainfall-runoff modeling. *Water Resour Manage* 23:2877–2894
- Nourani V, Mogaddam AA, Nadiri AO (2008) An ANN-based model for spatiotemporal groundwater level forecasting. *Hydrol Process* 22:5054–5066
- Nourani V, Kisi O, Komasi M (2011) Two hybrid artificial intelligence approaches for modeling rainfall-runoff process. *J Hydrol* 402:41–59
- Percival DB, Walden AT (2000) *Wavelet methods for time series analysis*. Cambridge University Press, Cambridge
- Sang YF (2012) A practical guide to discrete wavelet decomposition of hydrologic time series. *Water Resour Manage* 26:3345–3365
- Sang YF, Wang D, Wu JC (2010) Probabilistic forecast and uncertainty assessment of hydrologic design values using Bayesian theories. *Hum Ecol Risk Assess* 16(5):1184–1207
- Sang YF, Wang D, Wu JC, Zhu QP, Wang L (2009a) Entropy-based wavelet de-noising method for time series analysis. *Entropy* 11(4):1123–1147
- Sang YF, Wang D, Wu JC, Zhu QP, Wang L (2009b) The relation between periods' identification and noises in hydrologic series data. *J Hydrol* 368:165–177
- Sang YF, Wu JC, Wang D, Ling CP (2009c) New model of groundwater simulation and forecasting based on wavelet de-noising. *Manag Groundwater Environ* pp 55–58
- Sang YF, Wang ZG, Liu CM (2012) Period identification in hydrologic time series using empirical mode decomposition and maximum entropy spectral analysis. *J Hydrol* 424:154–164
- Singh VP (1998) *Entropy-based parameter estimation in hydrology*. Kluwer Academic Publishers, Boston
- Tiwari MK, Chatterjee C (2010) Uncertainty assessment and ensemble flood forecasting using bootstrap based artificial neural networks (BANNs). *J Hydrol* 382(1–4):20–33
- Wang WS, Ding J (2003) Wavelet network model and its application to the predication of hydrology. *Nature Sci* 1(1):67–71
- Yevjevich V (1972) *Stochastic process in hydrology*. Water Resources Publications, Colorado
- Zhang B, Govindaraju RS (2000) Forecasting of watershed runoff using Bayesian concepts and modular neural networks. *Water Resour Res* 36(3):753–762