

Comparison of hybrid models for daily streamflow prediction in a forested basin

Xue Li, Jian Sha, You-meng Li and Zhong-Liang Wang

ABSTRACT

Accurate forecasting of daily streamflow is essential for water resource planning and management. As a typical non-stationary time series, it is difficult to avoid the effects of noise in the hydrological data. In this study, the wavelet threshold de-noising method was applied to pre-process daily flow data from a small forested basin. The key factors influencing the de-noising results, such as the mother wavelet type, decomposition level, and threshold functions, were examined and determined according to the signal to noise ratio and mean square error. Then, three mathematical techniques, including an optimized back-propagation neural network (BPNN), optimized support vector regression (SVR), and adaptive neuro-fuzzy inference system (ANFIS), were used to predict the daily streamflow based on raw data and wavelet de-noising data. The performance of the three models indicated that a wavelet de-noised time series could improve the forecasting accuracy. The SVR showed a better overall performance than BPNN and ANFIS during both the training and validating periods. However, the estimation of low flow and peak flow indicated that ANFIS performed best in the prediction of low flow and that SVR was slightly superior to the others for forecasting peak flow.

Key words | adaptive neuro fuzzy inference system, artificial neural network, daily flow forecasting discrete wavelet transform, support vector regression

Xue Li
Jian Sha
Zhong-Liang Wang (corresponding author)
Tianjin Key Laboratory of Water Resources and Environment,
Tianjin Normal University,
Tianjin 300387,
China
E-mail: wangzhongliang@vip.skleg.cn

You-meng Li
The School of Computer Software,
Tianjin University,
Tianjin 300350,
China

INTRODUCTION

Accurate and immediate streamflow forecasting in rivers is essential for disaster (e.g., flood and drought) prediction and water resource management (Suliman *et al.* 2015). Various conceptual and data-based forecasting models have been developed in recent decades and applied around the world (Noori & Kalin 2016). The conceptual models focus on the study of rainfall-runoff processes and underlying physical laws, which involve multiple factors, such as climatic conditions, land use, and evapotranspiration, and have been subjected to the requirements of excessive data (Karran *et al.* 2014; Samadi 2017). Artificial intelligence (AI) techniques have been popular data-based methods in recent years and have been successfully developed to model nonlinear and non-stationary hydrological time series (Anvari *et al.* 2014; Singh *et al.* 2015; Bui *et al.* 2016).

These methods have the ability to extract the relationship between inputs and outputs without the need for detailed physical process understanding. Among AI methods, artificial neural networks (ANN), support vector machine (SVM), and adaptive neuro-fuzzy inference systems (ANFIS) have been proven to be effective tools for modeling complex hydrologic systems (Babovic *et al.* 2001; Vojinovic *et al.* 2003; Taormina & Chau 2015).

The abilities of these machine learning methods to predict non-stationary time series are always limited due to the direct utilization of real data, which cannot avoid the effects of noise (Lotric & Dobnikar 2005). To overcome this limitation, noise reduction has become an important issue and is treated as a pre-processing method (Lang *et al.* 1996). For non-stationary hydrological time series

data, conventional de-noising schemes, such as Wiener filter and Kalman filter, are excessively dependent on the establishment of appropriate state space functions (Chou 2014). Wavelet-based noise reduction techniques are able to decompose a signal into several scales that represent both temporal and frequency domains and have had broad applications in various data analyses and signal processing in recent years (Patel *et al.* 2014). Discrete wavelet transform (DWT) as a data-based mathematical technique is particularly suited to time series de-noising and data compression (Evrendilek & Karakaya 2014). However, there are multiple factors influencing the efficiency of DWT, such as the choice of mother wavelet, decomposition level, threshold estimation method, and so on (Sang *et al.* 2014). Extraction of useful information from raw hydrological time series data based on most optimal factors was beneficial for the improvement of the daily streamflow forecasting accuracy of the machine learning methods.

The objective of this study was to compare the performances of ANN, SVM, and ANFIS models coupled with the DWT de-noising method for daily streamflow forecasting. A small forested basin located in Eastern China with large interannual variability in streamflow was chosen as a case study. The sensitivity of de-noising processing to the wavelet type, decomposition level, and threshold functions were examined. The key parameters of three data-based models were optimized by different optimization algorithms to obtain the best model structures. Raw and de-noised daily averaged flow time series data were applied to compare the effectiveness of regular and hybrid models.

STUDY AREA AND DATA

In this study, historical daily flow data of the Yuetan Basin were collected from the Bureau of Hydrology, Ministry of Water Resources of PR China to evaluate the performance of different methods. The Yuetan Basin (117°38'–118°10' E, 29°33'–29°50'N) is the source of the Xin'anjiang River and is located in Huangshan City, Eastern China, with a drainage area of 950 km² (Figure 1). It is a forest-dominated catchment, with sparse farming and residential land. The percent of forest cover is 95.44%, while the farm land and residential land accounts for 3.26% and 0.01%, respectively.

Mountainous regions and hills are the major landforms in the basin. Located in the subtropical monsoon climate zone, the mean annual temperature is approximately 16 °C. Precipitation is predominated by storms, which are significantly influenced by the terrain. The average precipitation is 1,760 mm, with the majority occurring during May to August. Without the effects of human activity, the streamflow varies widely in different seasons. Most maximum values occur in June and July, while the minimal streamflow is centralized from November to January. The location of the basin along with the hydrologic station is shown in Figure 1. Daily averaged flow data for a period of 14 years from 2000 to 2013 were collected from Yuetan Station. The total number of available data was 5,114 days, which included various hydrological conditions during different seasons of the year. For the purposes of model induction (Babovic & Keijzer 2000), the whole dataset was split into two sub-sets, of which the first 11 years (2000–2010) data were used for training purposes, and the remaining 3 years (2011–2013) of data were used for validation. The main statistics of the training data and validating data are given in Table 1. The skewness values of both the training and validating data series are quite high, which may decrease the estimation accuracy of the models.

Streamflows in natural canals were always thought to have nonlinear and dynamic structure (Babovic & Keijzer 2000; Kişi 2007). In this study, the Lyapunov exponent was calculated first to determine whether the daily averaged flow data of Yuetan station exhibited chaotic behavior (Rosenstein *et al.* 1993). The Lyapunov exponent of this time series was positive (0.43), which indicated that the time series exhibited chaos features (Wolf *et al.* 1985). During the process of chaos theory, the first step was reconstructing the dynamics in phase space (Ghorbani *et al.* 2010). The delay-time method was the most frequently used reconstruction method for a univariate time series, namely, the input variables of the models presented the previously observed daily flow values, while the output variable corresponded to the one-day-ahead flow value ($t + 1$). Numerous studies have demonstrated that autocorrelation function (ACF) and partial autocorrelation function (PACF) are useful to determine the number of daily flow values at different time lags that can best represent the time series (Lohani *et al.* 2012; He *et al.* 2014). Therefore, ACF and PACF were



Figure 1 | Location of the hydrologic station in the Yuetan Basin.

Table 1 | Descriptive statistics for the training and testing dataset

Descriptive statistics	Unit	Training data	Testing data
Min.	m ³ /s	0.11	0.19
Mean	m ³ /s	31.39	38.07
Median	m ³ /s	12.5	14.65
Max.	m ³ /s	1,500	1,740
Standard deviation	m ³ /s	71.63	103.63
Skewness	Dimensionless	8.75	9.02

applied first to determine the lag time of inputs. According to the results of ACF and PACF, as shown in Figure 2, the maximum four-lag time was determined ($t, t-1, t-2, t-3$). Several

combinations were considered as inputs of the models: (I) $Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}$, (II) Q_t, Q_{t-1}, Q_{t-2} , (III) Q_t, Q_{t-1} , (IV) Q_t .

However, the ACF and PACF mainly showed the dependence from the perspective of linearity, while they may not be helpful in determining the lags of data generated nonlinearly (Granger & Lin 1994). According to the time-delay embedding theorem presented by Takens (1981), a time series x_i generated from a nonlinear dynamic system was able to reconstruct a phase space, as follows (Sun et al. 2010):

$$x_i = (x_i, x_{i-\tau}, \dots, x_{i-(m-1)\tau})$$

where τ is an appropriate delay time and m is the embedding dimension. The choice of the delay time τ is a very important step in phase space reconstruction. The average mutual

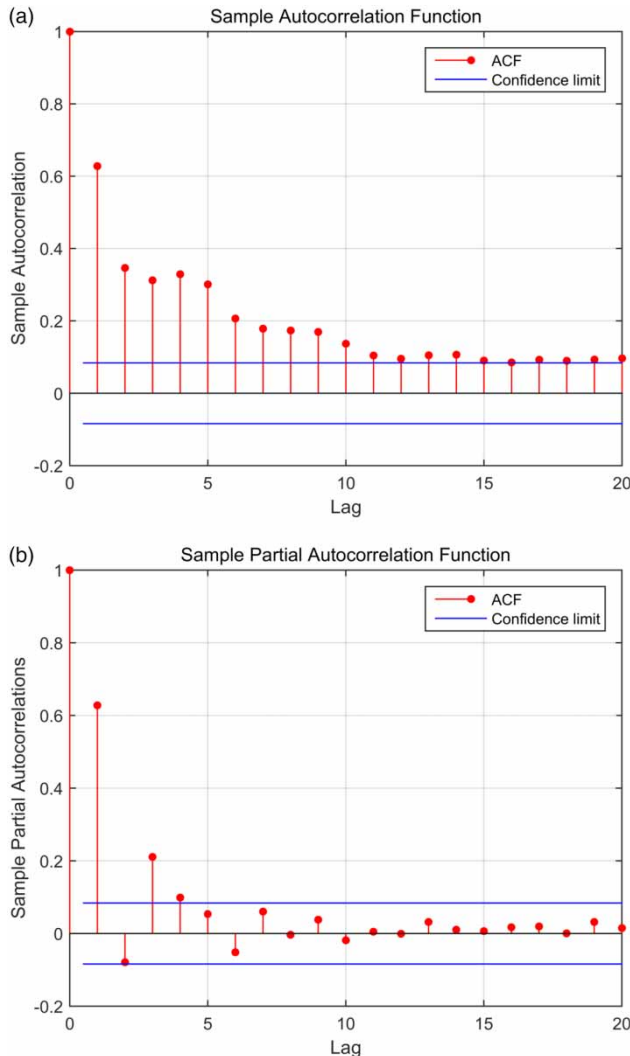


Figure 2 | ACF and PACF partial autocorrelation function of daily flow data.

information (AMI) analysis suggested by Fraser and others (Fraser & Swinney 1986; Sun et al. 2010) is commonly used to determine a proper delay time and is also applied in this study. We calculated AMI using time lags of 1–20 days and the first minimum AMI value occurred at $\tau = 3$, shown in Figure 3. Based on the optimal delay time 3, the embedding dimension m was calculated using the false nearest neighbors method (Kennel et al. 1992) and the result indicated that the optimal value of m was 8. Therefore, a multi-day-ahead input was tried in this study to predict Q_{t+1} : (V) Q_t , Q_{t-3} , Q_{t-6} , Q_{t-9} , Q_{t-12} , Q_{t-15} , Q_{t-18} , Q_{t-21} . The original input and output data were conveniently normalized to values between 0.1 and 0.9 before the training process to improve the efficiency of the models.

OPTIMIZED SUPPORT VECTOR MACHINE FOR REGRESSION

SVM is a recently developed supervised machine learning method for classification, which is then modified and used in regression problems (Kavousi-Fard et al. 2014). The general principles of support vector machine for regression (SVR) have been described in many studies (Kavousi-Fard et al. 2014; Li et al. 2016), so it is not necessary to list them in detail. The performance of SVR mainly depends on a set of parameters: the penalty parameter C , kernel function type and corresponding kernel parameters. In this study, a radius basis function (RBF) with high effectiveness and great speed was selected as the kernel function during the SVR training process. The penalty parameter C and kernel function's parameter c were determined by the particle swarm optimization algorithm (PSO), which was motivated by the social behavior of bird flocking (Kennedy 2010). Each particle in the PSO was given a random initial velocity and position, and then, during the process of particle movement, the best position of each particle was evaluated by a fitness function (Harish et al. 2015), which was expressed as the output error of the i th particle in this study:

$$f(x_i) = \sum_{k=1}^S t_k - p_k(x_i)$$

where f is the fitness value, S is the number of training samples, t_k is the observed value, p_k is the predicted value based on x_i , and x_i denotes the penalty parameter C and kernel function's parameter c . The 10-fold cross-validation method was applied to the optimized procedure, and the velocity and position of each particle were updated until the termination condition was satisfied. The optimum parameters obtained by PSO were tested with test data to check the performance of the SVR model.

OPTIMIZED ARTIFICIAL NEURAL NETWORKS

The ANN has recently been popular in many fields as a modeling technique to determine the nonlinear relationships between input variables and output variables (Giam

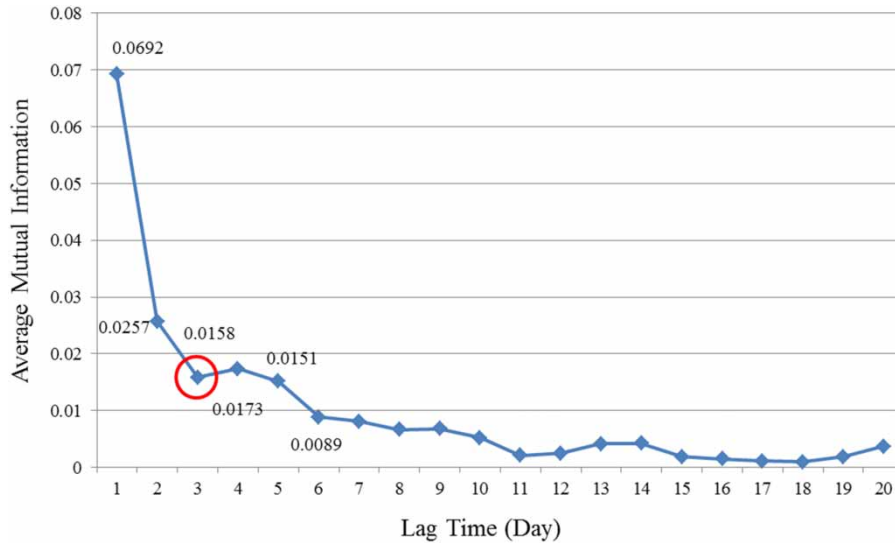


Figure 3 | AMI function of the daily flow time series.

& Olden 2015), and the back-propagation algorithm (BP) is the most commonly used supervised learning method for multilayer feed-forward networks (Li *et al.* 2017). BPNN is typically composed of an input layer, hidden layer, and output layer. BPNN has many advantages, such as its simple architecture, ease of construction, and rapid calculation speed (Liu *et al.* 2013). However, it has some issues in application, such as the choice of the neuron number in the hidden (NNH) layer and initialization of the network. A too small or too large NNH may lead to bad generalization capabilities and an over-fitting problem (Li *et al.* 2016). In this study, we calculated the range of NNH according to the following empirical formula:

$$N = \sqrt{M + J} + a$$

where N is NNH, M is the number of neurons of the input layer (1–4 and 8), J is the number of neurons of the output layer (1), and a is a constant from 1 to 10. The value ranges of NNH in this study were defined as 3–10 for inputs I–IV, and 4–13 for input V. The initial values of the network connection weights between the input layer and hidden layer as well as the hidden layer and output layer were important factors influencing the convergent behavior of BPNN. In this study, PSO was combined into BPNN and used to choose the appropriate connection weight matrices.

Furthermore, the 10-fold cross-validation method was applied to limit the over-fitting problem and improve the generalization ability of BPNN.

ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM

ANFIS is a fuzzy Sugeno model that maps inputs through input membership functions (MFs) and associated parameters to outputs through output MFs (Jang *et al.* 1997). More details of the mathematical formulations of ANFIS have been described in previous studies (Subasi 2007). The Gaussian and bell-shaped MFs have been increasingly popular for specifying fuzzy sets because of the smoothness and concise notation (Chang & Chang 2006). In this study, we chose the bell-shaped MFs as they had one additional parameter, which indicated that a non-fuzzy set can be more easily approached than Gaussian MFs. The hybrid learning algorithm, combining gradient descent and the least square techniques proposed by Jang *et al.* (Jang 1993), was applied to optimize the network parameters. During the model structure identification process, determination of the optimum number and the form of fuzzy rules is a crucial step. A fuzzy c-means clustering algorithm and subtractive clustering algorithm are applied to the dataset. After developing various models and repeatedly comparing

their results, the fuzzy c-means clustering algorithm was chosen to determine the appropriate number of clusters for each input.

WAVELET DE-NOISING METHOD

Hydrological time series often have complicated nonlinear and non-stationary characteristics, so that de-noising is a necessary task to reduce the noise interference (Babovic 2005; Sang 2013). Among the de-noising methods, wavelet threshold de-noising method was proposed by Donoho & Johnstone (1994) and proven as a simple and effective method by many studies (Wang et al. 2014). A large extent of noise could be suppressed, while singular information about the original signal was preserved (Lu et al. 2016). The wavelet threshold de-noising method was performed on the hydrological time series in this study, which consisted of three steps: (1) choosing the most appropriate wavelet function and number of resolution levels, and then, decomposing the original data with DWT; (2) applying several types of threshold functions to address the detailed coefficients of each level and obtaining the adjusted decomposed coefficients; (3) reconstructing the time series with adjusted coefficients and choosing the proper de-noise time series.

The description of DWT has occurred in many previous studies (Nalley et al. 2012). Daubechies wavelets were the most commonly used mother wavelets and had been proven to be effective in hydrologic time series modeling (Wei et al. 2012). The Daubechies wavelet family is often denoted by 'dbN', where db is the 'surname' and N is the wavelet order. Haar (db1), db2–db10 are common members of the Daubechies wavelet family. Each of them was applied to the whole hydrological dataset, and the one with the best result was chosen as the most appropriate mother wavelet. The number of decomposition levels (L) was another important factor that needed to be carefully selected during the decomposing process. Training with a small L is simple but may lead to inaccurate predictions without sufficient past information, while a high L causes hard training and slowly decreases the approximation error (Soltani 2002). Therefore, an L of 3–10 was processed for each mother wavelet, with the aim of determining the optimal L for the appropriate mother wavelet.

At each decomposition level, a suitable threshold (λ) needs to be selected, and the detailed coefficients are processed in the form of the threshold function (Zhang et al. 2015). We used two strategies to estimate the threshold in this study: (1) a universal threshold proposed by Donoho and Johnstone (Donoho 1995), $\lambda = \sigma\sqrt{2\log(N)}$, where N is the time series length and σ is the standard deviation of the noise; (2) a level-dependent threshold proposed by Birge & Massart (Misiti et al. 1996). This strategy kept all of the approximation coefficients for the level $J + 1$, and for level i from 1 to j , the n_i largest coefficients were kept according to the formula $n_i = \frac{M}{(J + 2 - i)^\alpha}$, where α and M are real numbers greater than 1. Generally, $\alpha = 3$ for de-noising purposes, and M is between $L(1)$ and $2 * L(1)$.

The choice of threshold function was critical as different functions showed various treatment strategies and mathematical estimation methods over and below the threshold λ . The two most widely used functions are the hard threshold function:

$$W'(j, k) = \begin{cases} W(j, k) & |W(j, k)| \geq \lambda \\ 0 & |W(j, k)| < \lambda \end{cases}$$

and soft threshold function:

$$W'(j, k) = \begin{cases} W(j, k) - \lambda & W(j, k) > \lambda \\ 0 & |W(j, k)| \leq \lambda \\ W(j, k) + \lambda & W(j, k) < -\lambda \end{cases}$$

However, the hard threshold function may lead to a larger variance in the reconstruction of the original series, while the soft threshold function may create an unnecessary bias when the true coefficients are large (Antoniadis 2007). To overcome the drawbacks of the hard and soft threshold functions, three modified threshold functions were applied in this study to process the coefficients.

1. Semi-soft threshold function:

$$W'(j, k) = \begin{cases} 0 & |W(j, k)| < \lambda_1 \\ \text{sgn}(W(j, k)) \left(\frac{\lambda(|W(j, k)| - \lambda_1)}{\lambda - \lambda_1} \right) & \lambda_1 \leq |W(j, k)| < \lambda \\ W(j, k) & |W(j, k)| \geq \lambda \end{cases}$$

2. Garrote threshold function:

$$W'(j, k) = \begin{cases} \operatorname{sgn}(W(j, k))\sqrt{(W(j, k))^2 - \lambda^2} & |W(j, k)| \geq \lambda \\ 0 & |W(j, k)| < \lambda \end{cases}$$

3. Generalized threshold function:

$$W'(j, k) = \begin{cases} \operatorname{sgn}(W(j, k))\left(|W(j, k)| - \lambda^m / |W(j, k)|^{(m-1)}\right) & |W(j, k)| \geq \lambda \\ 0 & |W(j, k)| < \lambda \end{cases}$$

where λ_1 is a secondary threshold defined as $\lambda_1 = mu * \lambda$ in the semi-soft threshold function. In the generalized threshold function, $m = 1, 2, \dots, \infty$; when $m = 1$, it is a soft threshold function and when $m = \infty$, it is a hard threshold function. We tried different threshold functions on the time series, and then, we used the modified coefficients to compute the inverse transform and reconstruct the de-noising time series. The accuracy of de-noising series were assessed by two indicators: the signal to noise ratio $SNR = -10 \log_{10} \frac{\sum_{i=1}^n (I_i - J_i)^2}{\sum_{i=1}^n (J_i)^2}$ and mean square error $MSE = \frac{1}{n} \sum_{i=1}^n (I_i - J_i)^2$, in which I_i and J_i are the real series and de-noised series, respectively.

The SNR was a measure of the desired signal strength relative to the background noise, and the decrease in the SNR meant more noise was included in the reconstructed time series. The MSE measured the difference between the raw series and de-noised series, and a higher value indicated a higher error. The reconstructed series with minimum MSE and maximum SNR were analyzed by PSO-BPNN, PSO-SVR, and ANFIS in the same way as the original series. All of the algorithms were performed by programming codes in MATLAB R2013b, and a Library for Support Vector Machines (LIBSVM) developed by Lin et al. (Chang & Lin 2011) was used to design the PSO-SVM model.

PERFORMANCE EVALUATION

The performances of the aforementioned models were evaluated by three statistical performance evaluation criteria, including the coefficient of efficiency (CE), root mean square error (RMSE), and mean absolute error (MAE). The CE is also known as the Nash-Sutcliffe coefficient and was defined as follows:

$$CE = 1 - \frac{\sum_{i=1}^n (Q_{io} - Q_{ip})^2}{\sum_{i=1}^n (Q_{io} - \bar{Q}_o)^2}$$

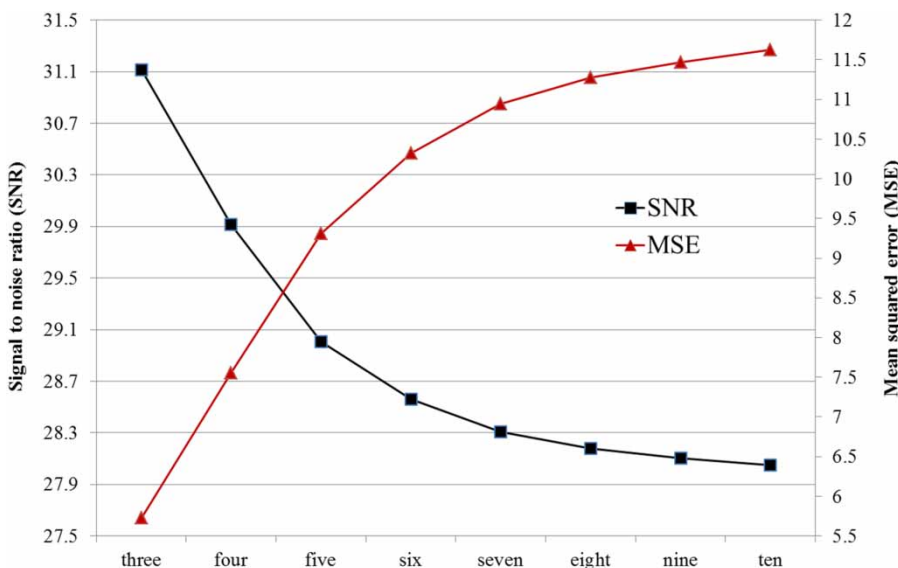


Figure 4 | Values of SNR and MSE from resolution levels 3–10.

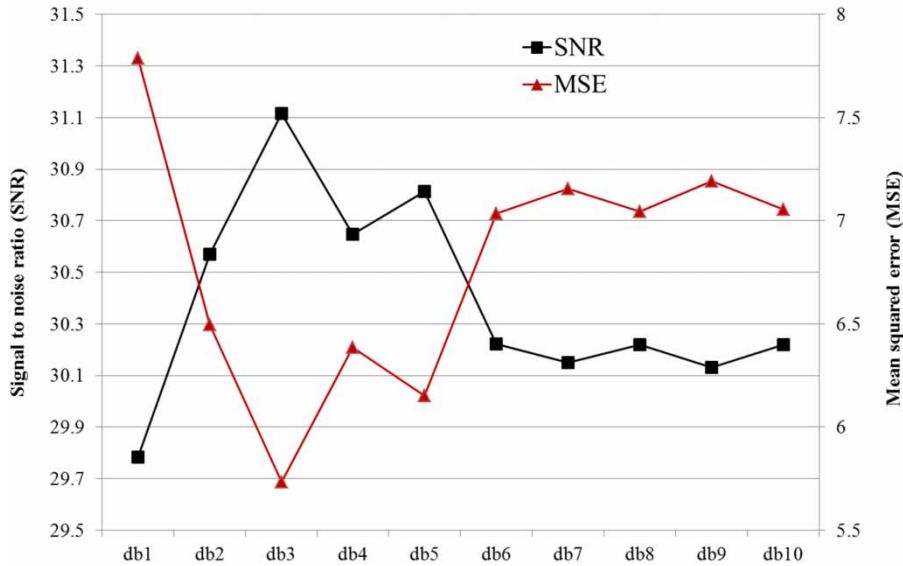


Figure 5 | SNR and MSE values from the mother wavelet db3–db10.

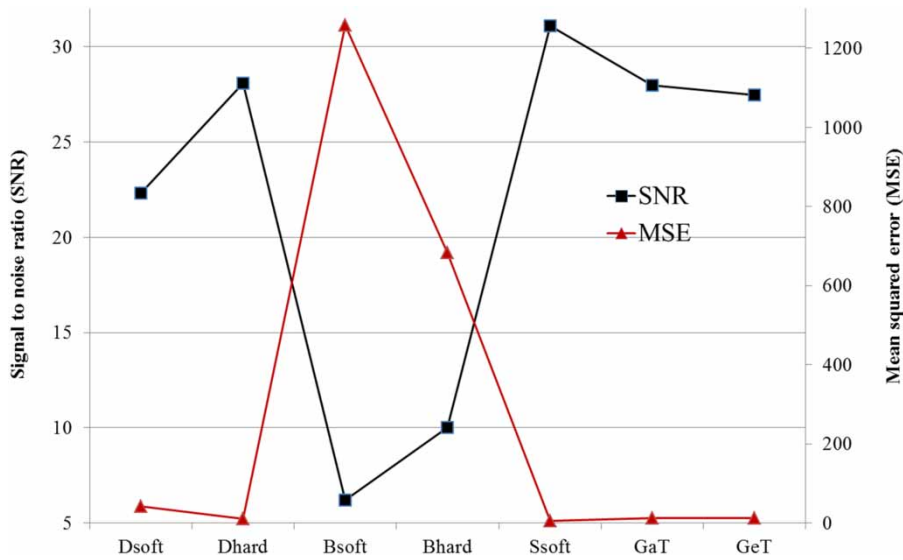


Figure 6 | SNR and MSE values from seven threshold functions.

The RMSE was calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{io} - Q_{ip})^2}{n}}$$

The MAE was defined as the average of absolute errors:

$$MAE = \frac{\sum_{i=1}^n |Q_{ip} - Q_{io}|}{n}$$

Table 2 | The best neurons number in hidden layer (NNH) for RS and WS

	Input	RS	WS
I	Q(t), Q(t-1), Q(t-2), Q(t-3)	6	4
II	Q(t), Q(t-1), Q(t-2)	7	3
III	Q(t), Q(t-1)	4	6
IV	Q(t)	3	7
V	Q(t), Q(t-3), Q(t-6), ..., Q(t-21)	5	10

where n is the number of input samples, Q_{io} and Q_{ip} are observed and predicted value of sample i , respectively. $\overline{Q_o}$ is the mean value of observed data. The values of $CE = 1$, $RMSE = 0$, and $MAE = 0$ would be obtained if the predicted values were exactly equal to the observed values.

RESULTS AND DISCUSSION

The de-noising effectiveness of the whole hydrological time series (2000–2013) at the Yuetan site was evaluated by the values of SNR and MSE mentioned above. The table including SNR and MSE values is supplied as Supplementary material (available with the online version of this paper).

To confirm the most appropriate resolution levels for decomposition, the results were compared among 3–10 resolution levels. Generally, the smoothing of the detailed signals improved with the increase of the resolution level number, while a greater resolution level led to error propagation (Chou 2014). In this study, with the increase of the resolution levels, the $SNRs$ were decreasing and $MSEs$ were increasing for all cases. Therefore, the number of resolution levels used in the wavelet decomposition was determined to be three. The trends of SNR and MSE towards resolution levels are shown in Figure 4, taking the mother wavelet $db3$ and semi-soft threshold function as examples.

Under resolution level three, the mother wavelet $db3$ offered better performance than the other wavelet functions,

Table 3 | The performance statistics of PSO-BPNN, PSO-SVM, and ANFIS during training and validating periods for RS and WS

Model	Input	CE		RMSE		MAE	
		Training	Validating	Training	Validating	Training	Validating
PSO-SVR_WS	I	0.85	0.82	31.87	46.95	9.94	11.21
	II	0.80	0.81	34.88	47.96	11.82	14.79
	III	0.70	0.62	43.13	68.17	11.10	13.97
	IV	0.62	0.45	47.91	82.30	12.99	17.97
	V	0.79	0.76	35.89	54.73	11.97	14.20
PSO-SVR_RS	I	0.80	0.81	35.23	48.00	14.36	14.96
	II	0.76	0.66	38.04	64.75	13.33	15.39
	III	0.69	0.65	43.49	65.52	11.78	14.09
	IV	0.61	0.46	48.70	81.51	15.26	20.00
	V	0.76	0.76	38.19	55.06	13.22	15.57
PSO-BPNN_WS	I	0.71	0.64	42.28	66.39	15.49	19.56
	II	0.69	0.58	43.34	71.97	13.98	17.72
	III	0.70	0.52	43.03	76.42	13.48	18.09
	IV	0.61	0.49	48.63	78.92	15.48	20.08
	V	0.68	0.71	44.08	60.46	14.84	17.72
PSO-BPNN_RS	I	0.67	0.66	45.06	64.15	16.88	21.04
	II	0.67	0.48	44.84	80.03	13.63	19.22
	III	0.69	0.54	43.36	74.75	14.11	18.46
	IV	0.62	0.49	48.17	78.94	14.69	19.56
	V	0.68	0.63	44.29	67.89	15.14	19.10
ANFIS_WS	I	0.78	0.81	36.94	47.98	10.69	12.88
	II	0.77	0.80	37.10	49.49	11.27	13.38
	III	0.68	0.69	44.17	61.31	11.99	14.24
	IV	0.64	0.60	47.03	69.92	12.97	16.26
	V	0.68	0.71	43.90	60.22	13.35	16.24
ANFIS_RS	I	0.74	0.74	40.08	56.83	12.21	14.57
	II	0.71	0.74	42.26	56.18	11.97	13.56
	III	0.69	0.67	43.30	63.20	12.47	15.35
	IV	0.61	0.61	48.72	69.32	13.52	16.25
	V	0.63	0.68	47.49	62.95	15.00	18.08

and it was selected as the objective wavelet. The values of *SNR* and *MSE* changed along with the wavelet functions, and the trends were not stable, as shown in Figure 5. Based on the determined resolution level and mother wavelet, a comparison can be performed to rank each threshold function's performance. Among seven threshold functions, the semi-soft threshold function (Ssoft) showed the best *SNR* and *MSE*, while the soft threshold function with the Birge and Massart strategy (Bsoft) gave the worst result (Figure 6). The ranks of the other functions were: a hard threshold function with the Donoho and Johnstone strategy (Dhard) > (better than) Garrote threshold function (GaT) > generalized threshold function (GeT) > soft threshold function with Donoho and Johnstone strategy (Dsoft) > hard threshold function with the Birge and Massart strategy (Bhard). By means of *SNR* and *MSE*, we compared the stability of the de-noising

performance of various wavelets and finally chose the wavelet with a scale of three and *db3* mother wavelet to de-noise the original hydrological time series based on Ssoft.

The wavelet de-noised dataset was separated into a training sub-set (2000–2010) and a validating sub-set (2011–2013) similar to the original one. A comparison between the results obtained from the raw time series (RS) and wavelet de-noised series (WS) during both training and validating periods was presented to evaluate the performances of PSO-SVR, PSO-BPNN, and ANFIS. To obtain the most effective performance of PSO-BPNN, the NNHs 3–10 were evaluated for inputs I–IV and 4–13 for input V during the training periods for both RS and WS. The NNH with the lowest *RMSE* and *MAE* of the same input may not be the same for RS and WS, which is shown in Table 2.

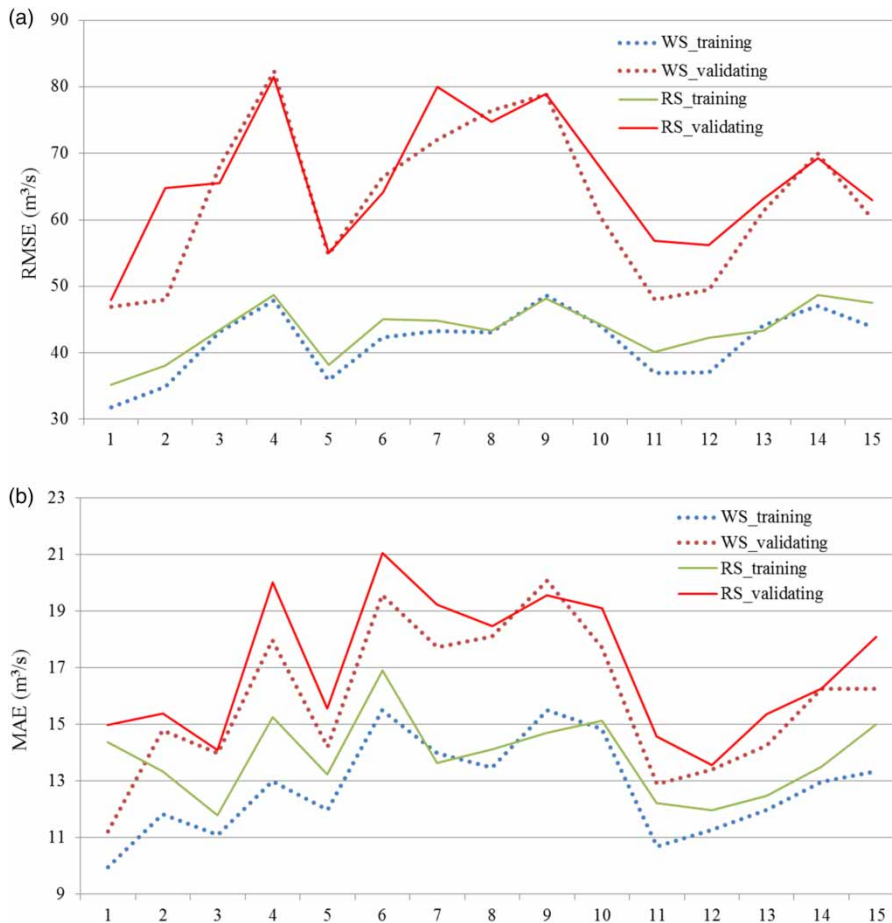


Figure 7 | *RMSE* and *MAE* values for all the cases of RS and WS during training and validating periods: (a) *RMSE* and (b) *MAE*.

The results of PSO-BPNN, PSO-SVR, and ANFIS with the best fit structures for RS and WS during the training and validating periods are summarized in Table 3. The performances during the two periods were similar for all models, which indicated good generalization capacities of these nonlinear systems. In order to find the differences of predictive capacities of RS and WS, the *RMSE* and *MAE* values were compared between the two series. The results of most cases showed that the *RMSE* and *MAE* values of WS were lower than RS during both periods (Figure 7), indicating a better predictability of WS and the necessity of data de-noising.

The comparison of results obtained from inputs I–V for the three WS models showed that the input consisting of four antecedent daily flow data (I) had a better fit than the others (Figure 8). With the lead time decreasing (II, III and IV), the results tended to deteriorate. The multi-day-ahead input (V) showed a moderate predictive capacity compared with the one-day-ahead inputs. It is worth

mentioning that the multi-day-ahead input had much better generalization capacities than inputs II–IV for the PSO-BPNN, which was not very satisfactory during the validating periods of both RS and WS.

Among the three models, PSO-SVR obtained the smallest values of *RMSE* and *MAE* as well as the highest value of *CE* for input I of WS, and so it was selected as the best-fit model for daily streamflow prediction in this study. ANFIS was slightly worse than PSO-SVR, but was apparently better than PSO-BPNN. The performances of PSO-SVR, PSO-BPNN, and ANFIS suggested that the three models showed good prediction accuracies for intermediate and moderate streamflow. However, it was an important measure to evaluate the predictive ability of a model on the estimations of the high and low streamflow. The *MAEs* of the three models for the 50 lowest values of the training data and validating data were compared and are shown in Figure 9. The differences between the simulated results and observed data of PSO-SVR and ANFIS

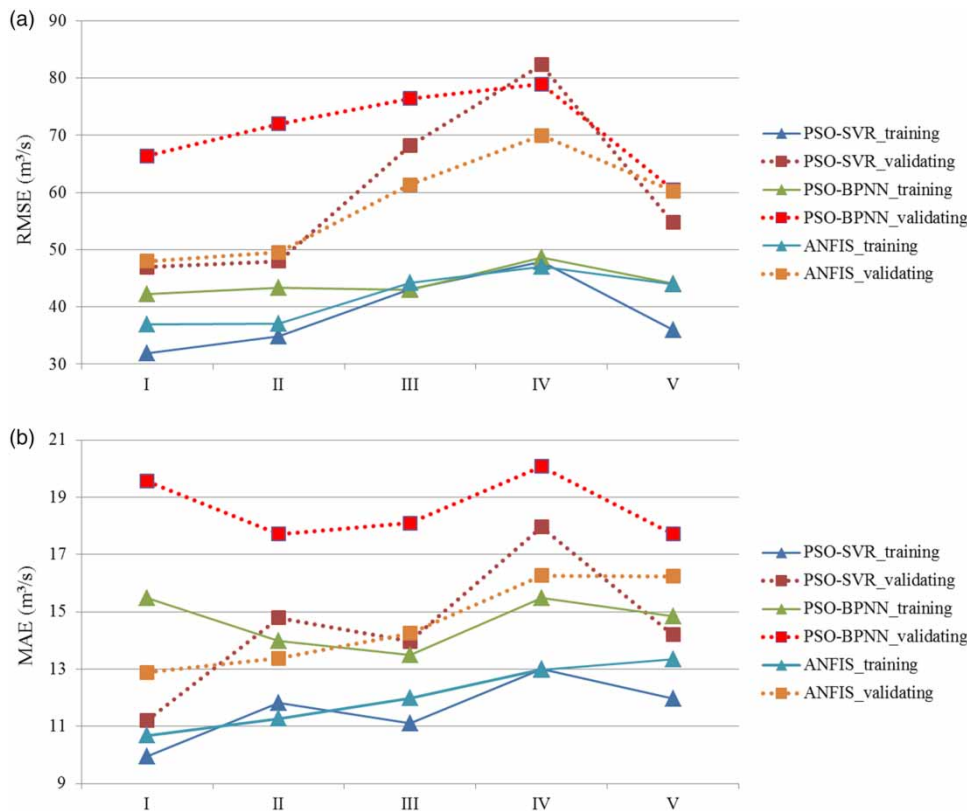


Figure 8 | *RMSE* and *MAE* values of inputs I–V for three WS models during training and validating periods: (a) *RMSE* and (b) *MAE*.

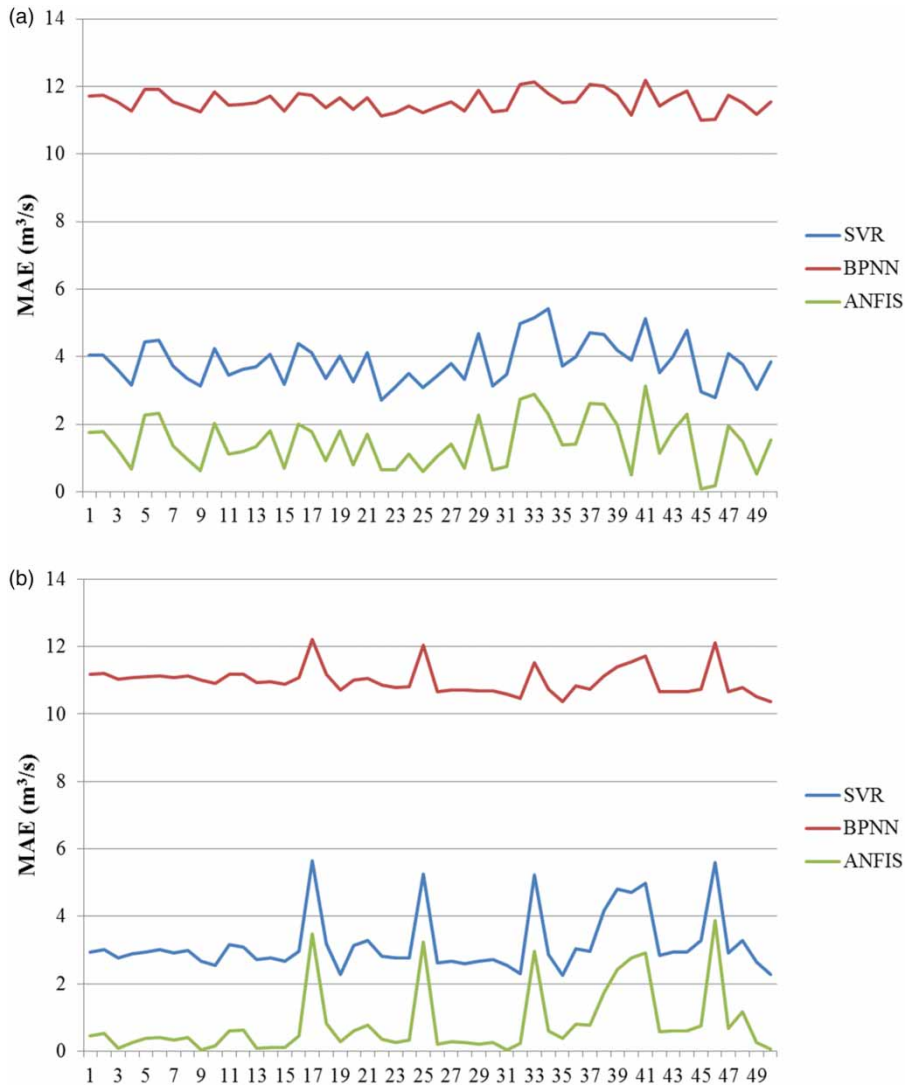


Figure 9 | The performances of three models on the 50 lowest values of the training data and testing data.

were apparently smaller than for PSO-BPNN during the two periods. The change trends of *MAE* were similar between PSO-SVR and ANFIS, but the prediction ability of ANFIS was superior to that of PSO-SVR with minimum data. The performances of the three models on the 50 highest values of the training data and validating data were also compared (Figure 10), and none appeared to be superior to the others. The average relative errors of PSO-SVR, PSO-BPNN, and ANFIS were 34.01%, 41.01%, and 31.86% for the highest training data and 28.83%, 56.70%, and 44.08% for the highest validating data, respectively. PSO-SVR showed a smoother trend of

errors and better generalization ability than the other two models for maximal data. Overall, PSO-SVR provided accurate and reliable daily flow predictions for moderate and high streamflow, and ANFIS was good at handling low values.

CONCLUSION

In this study, the performances of PSO-BPNN, PSO-SVR, and ANFIS for daily streamflow prediction in a forested basin were compared using one-day-ahead and

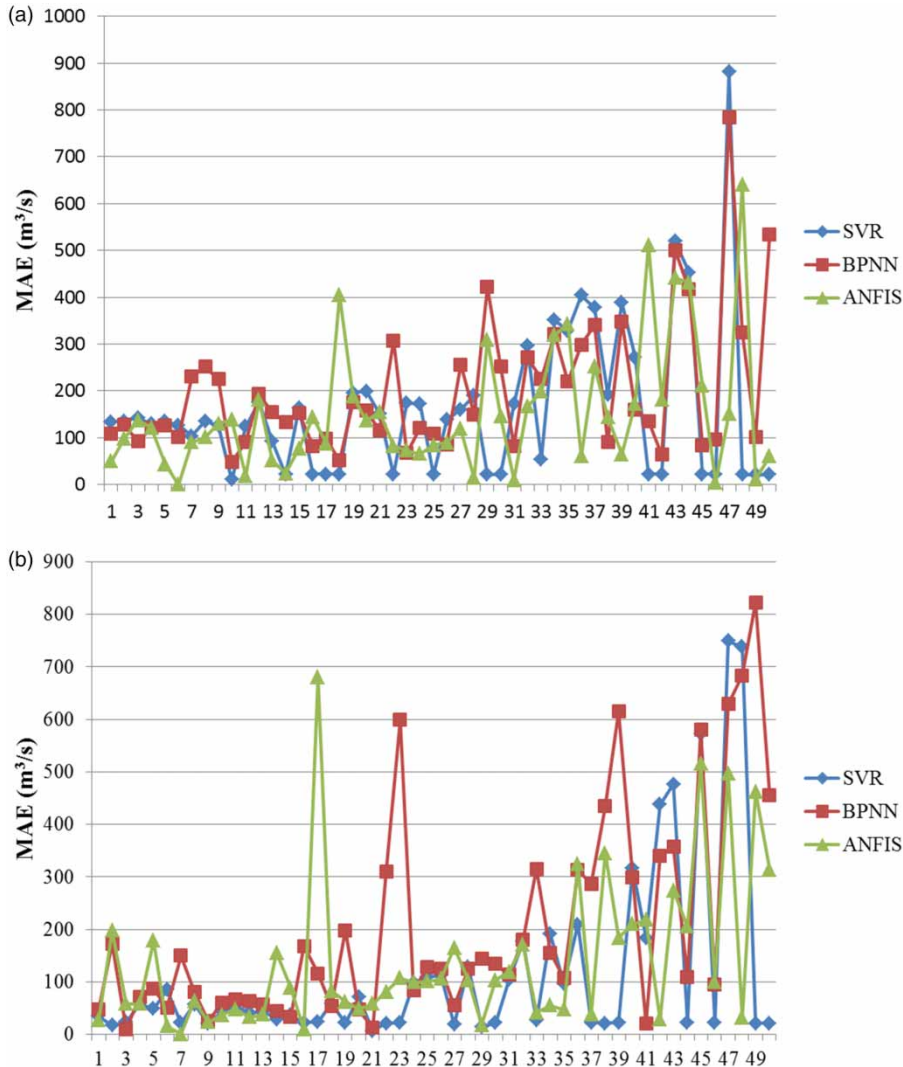


Figure 10 | The performances of three models on the 50 highest values of the training data and testing data.

multi-day-ahead inputs, respectively. The hydrological time series was first decomposed and de-noised by the wavelet modeling framework to enhance the forecasting performance. By discussing several key issues regarding the wavelet de-noising method, seven threshold functions were applied to identify and separate the deterministic components and noise of a hydrologic series. The wavelet with three levels and the db3 mother wavelet were chosen to de-noise the raw series using the Ssoft method. Both the RS and WS were used for daily flow prediction, and the predictive capacities of PSO-BPNN, PSO-SVR, and ANFIS were evaluated. The WS models

presented better results than the RS models for all situations. The further comparison of WS models indicated that the PSO-SVR model was superior to the PSO-BPNN and ANFIS models, and the 4-day ahead SVR_WS yielded the best performance among all of the models in terms of various model efficiencies. Due to the effects of climate and topography, extreme values were very common in the daily flow data of the Yuetan Basin. A comparison of the estimation of low flow and peak flow showed that the PSO-SVR model was slightly superior to the others for peak flow prediction, while the ANFIS model performed best for low flow prediction.

As the de-noising stage enhanced the forecasting performance and the proposed model could be applied for short-term forecasting of a hydrologic time series, it was difficult to identify a method that could maintain its prediction accuracy for extreme values. The results obtained here were limited to a single application, and further studies are needed to assess more cases and improve the predictive capacity by introducing a subsection function model or other new methods.

ACKNOWLEDGEMENTS

This work is supported by the Technology Research and Development Program of the Higher Education Institutions of Tianjin (043135202JW1716), the innovation team training plan of the Tianjin Education Committee (TD12-5037) and National Natural Science Foundation of China (No. 41372373). Comments and suggestions from two anonymous reviewers and the editor are greatly appreciated.

REFERENCES

- Antoniadis, A. 2007 Wavelet methods in statistics: some recent developments and their applications. *Statistics Surveys* **1**, 16–55.
- Anvari, S., Mousavi, S. J. & Morid, S. 2014 Sampling/stochastic dynamic programming for optimal operation of multi-purpose reservoirs using artificial neural network-based ensemble streamflow predictions. *Journal of Hydroinformatics* **16**, 907–921.
- Babovic, V. 2005 Data mining in hydrology. *Hydrological Processes* **19**, 1511–1515.
- Babovic, V. & Keijzer, M. 2000 Forecasting of river discharges in the presence of chaos and noise. *NATO Science Series 2 Environmental Security* **71**, 405–420.
- Babovic, V., Cañizares, R., Jensen, H. R. & Klinting, A. 2001 Neural networks as routine for error updating of numerical models. *Journal of Hydraulic Engineering* **127**, 181–193.
- Bui, D. T., Pradhan, B., Nampak, H., Bui, Q.-T., Tran, Q.-A. & Nguyen, Q.-P. 2016 Hybrid artificial intelligence approach based on neural fuzzy inference model and metaheuristic optimization for flood susceptibility modeling in a high-frequency tropical cyclone area using GIS. *Journal of Hydrology* **540**, 317–330.
- Chang, C.-C. & Lin, C.-J. 2011 LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2**, 27.
- Chang, F.-J. & Chang, Y.-T. 2006 Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Advances in Water Resources* **29**, 1–10.
- Chou, C.-m. 2014 Random modeling of daily rainfall and runoff using a seasonal model and wavelet denoising. *Mathematical Problems in Engineering* **11**, doi:10.1155/2014/917365.
- Donoho, D. L. & Johnstone, I. M. 1994 Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.
- Donoho, D. L. 1995 De-noising by soft-thresholding. *IEEE Transactions on Information Theory* **41**, 613–627.
- Evrendilek, F. & Karakaya, N. 2014 Regression model-based predictions of diel, diurnal and nocturnal dissolved oxygen dynamics after wavelet denoising of noisy time series. *Physica A: Statistical Mechanics and its Applications* **404**, 8–15.
- Fraser, A. M. & Swinney, H. L. 1986 Independent coordinates for strange attractors from mutual information. *Physical Review A, General Physics* **33**, 1134–1140.
- Ghorbani, M. A., Kisi, O. & Aalinezhad, M. 2010 A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest lyapunov methods. *Applied Mathematical Modelling* **34**, 4050–4057.
- Giam, X. & Olden, J. D. 2015 A new R 2-based metric to shed greater insight on variable importance in artificial neural networks. *Ecological Modelling* **313**, 307–315.
- Granger, C. & Lin, J. L. 1994 Using the mutual information coefficient to identify lags in nonlinear models. *Journal of Time Series Analysis* **15**, 371–384.
- Harish, N., Mandal, S., Rao, S. & Patil, S. 2015 Particle swarm optimization based support vector machine for damage level prediction of non-reshaped berm breakwater. *Applied Soft Computing* **27**, 313–321.
- He, Z., Wen, X., Liu, H. & Du, J. 2014 A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *Journal of Hydrology* **509**, 379–386.
- Jang, J.-S. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* **23**, 665–685.
- Jang, J., Sun, C.-T. & Mizutani, E. 1997 *Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, Englewood Cliffs, NJ.
- Karran, D. J., Morin, E. & Adamowski, J. 2014 Multi-step streamflow forecasting using data-driven non-linear methods in contrasting climate regimes. *Journal of Hydroinformatics* **16**, 671–689.
- Kavousi-Fard, A., Samet, H. & Marzbani, F. 2014 A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. *Expert Systems with Applications* **41**, 6047–6056.
- Kennedy, J. 2010 Particle swarm optimization. In: *Encyclopedia of Machine Learning* (C. Sammut & G. I. Webb, eds). Springer, Boston, MA, pp. 760–766.

- Kennel, M. B., Brown, R. & Abarbanel, H. D. 1992 Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A* **45**, 3403–3411.
- Kişî, Ö. 2007 Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering* **12**, 532–539.
- Lang, M., Guo, H., Odegard, J. E., Burrus, C. S. & Wells, R. O. 1996 Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters* **3**, 10–12.
- Li, X., Sha, J. & Wang, Z.-l. 2016 A comparative study of multiple linear regression, artificial neural network and support vector machine for the prediction of dissolved oxygen. *Hydrology Research* **48** (5), 1214–1225.
- Li, X., Sha, J. & Wang, Z.-L. 2017 Chlorophyll-a prediction of lakes with different water quality patterns in China based on hybrid neural networks. *Water* **9**, 524.
- Liu, M., Wang, M., Wang, J. & Li, D. 2013 Comparison of random forest, support vector machine and back propagation neural network for electronic tongue data classification: application to the recognition of orange beverage and Chinese vinegar. *Sensors and Actuators B: Chemical* **177**, 970–980.
- Lohani, A., Kumar, R. & Singh, R. 2012 Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology* **442**, 23–35.
- Lotric, U. & Dobnikar, A. 2005 Predicting time series using neural networks with wavelet-based denoising layers. *Neural Computing & Applications* **14**, 11–17.
- Lu, J.-y., Hong, L., Dong, Y. & Zhang, Y.-s. 2016 A new wavelet threshold function and denoising application. *Mathematical Problems in Engineering* **8**, doi:10.1155/2016/3195492.
- Misiti, M., Misiti, Y., Oppenheim, G. & Poggi, J.-M. 1996 *Wavelet Toolbox*. The MathWorks Inc., Natick, MA, 15, 21.
- Nalley, D., Adamowski, J. & Khalil, B. 2012 Using discrete wavelet transforms to analyze trends in streamflow and precipitation in quebec and Ontario (1954–2008). *Journal of Hydrology* **475**, 204–228.
- Noori, N. & Kalin, L. 2016 Coupling SWAT and ANN models for enhanced daily streamflow prediction. *Journal of Hydrology* **533**, 141–151.
- Patel, A. X., Kundu, P., Rubinov, M., Jones, P. S., Vértés, P. E., Ersche, K. D., Suckling, J. & Bullmore, E. T. 2014 A wavelet method for modeling and despiking motion artifacts from resting-state fMRI time series. *Neuroimage* **95**, 287–304.
- Rosenstein, M. T., Collins, J. J. & De Luca, C. J. 1993 A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena* **65**, 117–134.
- Samadi, S. Z. 2017 Assessing the sensitivity of SWAT physical parameters to potential evapotranspiration estimation methods over a coastal plain watershed in the southeastern United States. *Hydrology Research* **48**, 395–415.
- Sang, Y.-F. 2013 A review on the applications of wavelet transform in hydrology time series analysis. *Atmospheric Research* **122**, 8–15.
- Sang, Y.-F., Liu, C., Wang, Z., Wen, J. & Shang, L. 2014 Energy-based wavelet de-noising of hydrologic time series. *PLoS One* **9**, e110733.
- Singh, G., Panda, R. K. & Lamers, M. 2015 Modeling of daily runoff from a small agricultural watershed using artificial neural network with resampling techniques. *Journal of Hydroinformatics* **17**, 56–74.
- Soltani, S. 2002 On the use of the wavelet decomposition for time series prediction. *Neurocomputing* **48**, 267–277.
- Subasi, A. 2007 Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction. *Computers in Biology and Medicine* **37**, 227–244.
- Suliman, A. H. A., Jajarmizadeh, M., Harun, S. & Darus, I. Z. M. 2015 Comparison of semi-distributed, GIS-based hydrological models for the prediction of streamflow in a large catchment. *Water Resources Management* **29**, 3095–3110.
- Sun, Y., Babovic, V. & Chan, E. S. 2010 Multi-step-ahead model error prediction using time-delay neural networks combined with chaos theory. *Journal of Hydrology* **395**, 109–116.
- Takens, F. 1981 Detecting strange attractors in turbulence. *Lecture Notes in Mathematics* **898**, 366–381.
- Taormina, R. & Chau, K.-W. 2015 Neural network river forecasting with multi-objective fully informed particle swarm optimization. *Journal of Hydroinformatics* **17**, 99–113.
- Vojinovic, Z., Kecman, V. & Babovic, V. 2003 Hybrid approach for modeling wet weather response in wastewater systems. *Journal of Water Resources Planning and Management* **129**, 511–521.
- Wang, D., Singh, V. P., Shang, X., Ding, H., Wu, J., Wang, L., Zou, X., Chen, Y., Chen, X. & Wang, S. 2014 Sample entropy-based adaptive wavelet de-noising approach for meteorologic and hydrologic time series. *Journal of Geophysical Research: Atmospheres* **119**, 8726–8740.
- Wei, S., Song, J. & Khan, N. I. 2012 Simulating and predicting river discharge time series using a wavelet-neural network hybrid modelling approach. *Hydrological Processes* **26**, 281–296.
- Wolf, A., Swift, J. B., Swinney, H. L. & Vastano, J. A. 1985 Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena* **16**, 285–317.
- Zhang, B., Sun, L., Yu, H., Xin, Y. & Cong, Z. 2015 A method for improving wavelet threshold denoising in laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **107**, 32–44.

First received 6 July 2017; accepted in revised form 9 November 2017. Available online 29 November 2017