Improving prediction accuracy of river discharge time series using a Wavelet-NAR artificial neural network

Shouke Wei, Depeng Zuo and Jinxi Song

ABSTRACT

This study developed a wavelet transformation and nonlinear autoregressive (NAR) artificial neural network (ANN) hybrid modeling approach to improve the prediction accuracy of river discharge time series. Daubechies 5 discrete wavelet was employed to decompose the time series data into subseries with low and high frequency, and these subseries were then used instead of the original data series as the input vectors for the designed NAR network (NARN) with the Bayesian regularization (BR) optimization algorithm. The proposed hybrid approach was applied to make multi-step-ahead predictions of monthly river discharge series in the Weihe River in China. The prediction results of this hybrid model were compared with those of signal NARNs and the traditional Wavelet-Artificial Neural Network hybrid approach (WNN). The comparison results revealed that the proposed hybrid model could significantly increase the prediction accuracy and prediction period of the river discharge time series in the current case study.

Key words | Bayesian regularization, NAR network, river discharge, wavelet transformation, Weihe River

Shouke Wei (corresponding author)
Department System Analysis,
Integrated Assessment and Modelling,
The Swiss Federal Institute of Aquatic Science and Technology (EAWAG), 8600 Dübendorf,
Switzerland
and
Apmosian SciTech International Inc.,
BC V5P 3R1, Vancouver,
Canada
E-mail: *shouke.wei@gmail.com*

Depeng Zuo
College of Water Sciences,
Beijing Normal University,
100875 Beijing,
China

Jinxi Song
College of Urban and Environmental Sciences,
Northwest University,
710069 Xi'an,
China

INTRODUCTION

River discharge prediction for any time interval (yearly, monthly, daily or hourly) is essential for planning various activities in the river catchment management (Wei *et al.* 2012). A variety of methods have been developed to predict river discharge, including time series methods (Shamseldin & O'Connor 1996; Porporato & Ridolfi 2001; Mohammadi *et al.* 2006; Wei 2008), conceptual models (Nash & Sutcliffe 1970; Toth *et al.* 2000; Collischonn *et al.* 2005) and physical models (Garrote & Bras 1995; Fukushima 1988; Spruill *et al.* 2000).

In the last few decades, the artificial neural network (ANN) has been successfully employed to various different fields including hydrological flow estimation and prediction (Atiya *et al.* 1999; Campólo *et al.* 1999; Birikundavyi *et al.* 2002; Huang *et al.* 2004; Kişi 2007). ANN is a type of function approximator, which is good at forecasting nonlinear hydrologic time series (Kisi & Cigizoglu 2007; Wei *et al.* 2012). The advantages of ANN models are that they do not require information on whether the data are from a specific statistical distribution or stationary, nor what the exact relationships amongst the various input variables are (Maier & Dandy 1996). Even though the ANN modeling procedure does not require a detailed knowledge of hydrological characteristics, a well-trained ANN model can easily be applied to water resources management issues (Srinivasulu & Jain 2006). ANN also has the advantages of self-learning, self-organizing, and self-adapting (Ertay & Çekyay 2005; Feng & Hong 2008). Many studies compared the results of ANN models with those of traditional model approach, and the comparison results displayed that ANN models have good forecasting performance (Tsai & Lee 1999; Abrahart & See 2000; Kişi 2005; Sahoo *et al.* 2009; Bahrami *et al.* 2011).

However, when river discharge time series are characterized by high nonlinearity and non-stationarity with excessive noise, ANN will be unable to capture accurately a river discharge behavior (Wei *et al.* 2012). Later on, many researchers integrated ANN with other modeling

doi: 10.2166/hydro.2012.143

approaches to increase simulation accuracy, such as ANN and ARIMA (Auto Regressive Integrated Moving Average) approach (Cigizoglu 2003; Kişi 2005; Jain & Kumar 2007), ANN and physical model (Demirel *et al.* 2009), phase-space reconstruction and ANN approaches (Sivakumar *et al.* 2002; Wu *et al.* 2009b), conceptual and ANN integrated model (Tokar & Markus 2000; Srinivasulu & Jain 2009), ANN and non-parametric methods (Brath *et al.* 2002), and ANN and fuzzy hybrid techniques (Chang & Chen 2001; Aqil *et al.* 2007).

More recently, Wavelet and ANN (WNN) hybrid modeling approach (Wei *et al.* 2012) and even more complicated combination method such as Wavelet-Bootstrap-ANN hybrid model (Tiwari & Chatterjee 2010, 2011) have attracted increasing interest due to advantages in terms of good fitting and prediction accuracy compared to other traditional models. Wavelet transformation analysis is therefore becoming a popular analysis technique due to its ability to reveal simultaneously both spectral and temporal information within one signal (Nourani *et al.* 2009). This method advances Fourier analysis, where the basic shortcoming was that the Fourier spectrum contained only globally average information (Wei *et al.* 2012). Wavelet transformation is able to decompose a time series into its subseries and capture useful information at different resolution levels, and thus it has been to analyze data and make forecasting. Two essential contributions to this topic are Murtagh *et al.* (2004) and Adamowski (2008). The former employed wavelet transformation as a mean of handling time series data when future data are unknown. The latter developed forecasting models based on wavelet and cross-wavelet constituent components, which displayed that the proposed models can be used with great accuracy as a stand-alone forecasting method for short-term river flood. For the WNN hybrid model, it was first proposed by Aussem *et al.* (1998) to predict a financial time series. For river discharge estimation, Kim & Valdés (2003) applied dyadic wavelet transforms and neural networks to forecast droughts in the Conchos River Basin in Mexico, and the results displayed that the conjunction model significantly improved the forecasting ability of neural networks. Wang & Ding (2003) applied wavelet network model to predict daily discharge of the Yangtze River in China, which showed that the hybrid model could increase the forecast accuracy and extend the prediction time. Cannas *et al.* (2006) employed continuous and discrete wavelet transforms and data partitioning to investigate the effects of data processing for river flow forecasting using a neural network, with results that showed that networks trained with pre-processed data performed better than networks trained with an unrecompensed signal. Kişi (2008) applied a neuro-wavelet model to forecast (1-month-ahead) stream flows of the Canakdere and Goksudere Rivers in Turkey and compared the results with those of single multi-layer perceptron (MLP), multilinear regression (MLR) and auto-regression (AR) models, revealing that the combined WNN model could increase both the forecast accuracy and performance. Partal (2009) combined different artificial neural network algorithms and wavelet transform for monthly river flow forecasting in Turkey with comparison results revealing that wavelet and feed-forward back-propagation model was superior to other models. Kişi (2009) proposed the application of a neuro-wavelet conjunct model for forecasting daily intermittent stream-flow and the comparison results revealed that the suggested model could significantly increase the forecast accuracy of single ANN. Some other studies on the application of wavelet-neural network for forecasting river discharge in recent years (e.g. Adamowski & Sun 2010; Pramanik *et al.* 2010; Adamowski & Chan 2011; Krishna *et al.* 2011; Tiwari & Chatterjee 2011; Wei *et al.* 2012) have proved that wavelet and artificial neural network hybrid models have better performances in river flow forecasting than single artificial neural network models. However, the model prediction results in those previous studies still lack accuracy. Besides, network training with smaller data sets have not been discussed in those literature except that Wei *et al.* (2012) discussed the topic of small data sets for net training using one river discharge time series from the Weihe River, China.

The objectives of this study include: (1) to develop a Wavelet nonlinear autoregressive network (WNAR) hybrid modeling method to improve the prediction accuracy; (2) to estimate river discharge time series of Weihe River in China using the proposed method; (3) to investigate network generalization improvement under small data set by means of Bayesian regularization (BR); and (4) to compare the prediction results of the hybrid models with those of the single NAR network (NARN) model and traditional WNN.

STUDY AREA AND MATERIALS

The Weihe River originates from the Wushu Mountain in Weiyuan County of the Gansu province in west-central China, and runs across the Shaanxi province and joins the Yellow River at Tonguan in Shaanxi. The river is 818 kilometers long with a watershed area of 135,000 km^2, and is the largest tributary of the Yellow River (Figure 1). The river has played a large role in the development of West China and the health of the ecosystem of the Yellow River. However, since the 1980s, the river has suffered many problems, including greatly reduced annual river runoff, geologic hazards caused by water projects, high concentrations of sediment and consequential heavy flooding, as well as heavy water pollution (Song *et al.* 2007; Wei *et al.* 2012). Since the late 1990s, many parts of the river have lost ecosystem functionality, which have restricted the sustainable development of the region (Zhao 2003; Wang *et al.* 2004). In this context, it is of significant interest to have an advanced technique to model the river's discharge with improved prediction accuracy.

The data sets used in this study are monthly river discharge time series collected from three gages of the Weihe River, namely Linton, Huanxian, and Xianyang. Huaxian series and Xianyang series include 504 month (42 years) flow data from January 1960 to December 2001, and Lintong series include 492 month (41 years) data from January 1961 to December 2001. The data sets in the last 48 months (around 10%) of the three data series are used for network prediction test and the rest of the data for network training. Data set divisions and their statistical analysis are presented in Table 1. The analysis results displayed that values of mean and standard deviation in training sets are much larger than those in testing set. From the Skewness and Kurtosis of the available data, it revealed that distribution of the training data sets are more positive skewed and more peaked than that of the testing data sets. Those indicate that it might be difficult for a single ANN model to predict the testing data.

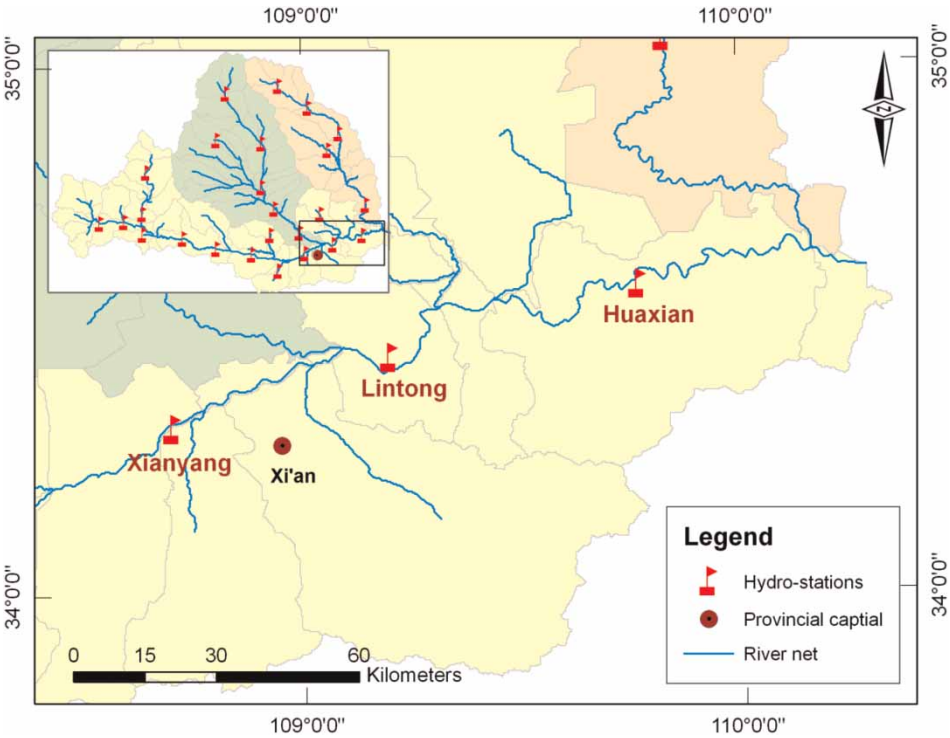

Figure 1 | A sketch of the water system of the Weihe River Basin in China, where Huaxian gage, Lintong gage and Xianyang gage are the study area.

Table 1 | Statistics of river discharge data set for WNAR and NAR networks

Gage	Data set	Obs.	Mean	Std.	Skew.	Kurt.	Max.	Min.
Huaxian	Training	456	236.5	271.6	2.62	12.0	2,000	2.67
	Test	48	111.0	114.7	1.76	5.63	504.0	8.59
	Entire	504	224.5	263.2	2.72	12.8	2,000	2.67
Lintong	Training	444	223.7	235.3	2.30	9.34	1,490	15.4
	Test	48	110.7	104.9	1.66	5.15	455	14.6
	Entire	492	212.6	228.3	2.39	9.97	1,490	14.6
Xianyang	Training	456	135.6	153.8	2.47	10.5	974	3.65
	Test	48	47.45	57.84	1.89	5.71	227	2.32
	Entire	504	127.2	149.6	2.56	11.2	974	2.32

METHODS

In this study two modeling techniques, nonlinear autoregressive (NAR) artificial neural network and discrete wavelet analysis (DWA) were used. We used MATLAB Wavelet Toolbox™ 4 and Neural Network Toolbox™ 6 for the analyses.

Nonlinear autoregressive (NAR) network

An artificial neural network (ANN) is defined as a structure comprised of a number of simple interconnected operating elements called neurons (or units, cells, or nodes), as inspired by the biological nervous system (Beale *et al.* 2010; Wei *et al.* 2012). A neural network can perform a particular function mapping between inputs and outputs, by adjusting the values of the connections (weights) between neurons. The important decisions necessary to build a neural network model include neural network type, network structure, methods of pre- and post-processing of input/output data, choice of training algorithms as well as training stop criteria (Wang *et al.* 2006; Wei *et al.* 2012). The most common network structure is Multi-layer Feed-forward Neural Network (MLFN), which is a static network. Dynamic neural network is the extension of static neural network by considering time. In this study, NAR network was used. NAR network (NARN) is one type of dynamic neural network that is good at time series prediction, where there is only one series involved. The future values of a time series $Y(t)$ are predicted only from d past values of that series. This form of prediction can be written as follows:

$$Y(t) = f(Y(t-1), Y(t-2), \ldots, Y(t-d)) \quad (1)$$

Determining the numbers of hidden layers and neurons (i.e. network structure) is the most important task in building the network. The standard NARN is feed-forward network, including three layers in its structure, i.e. input, hidden and output layers, with a sigmoid transfer function in the hidden layer and a linear transfer function in the output layer (Figure 2(a)). While a trial and error procedure to determine the number of neurons is still the most commonly used methods some alternative algorithms have also been proposed (Wei *et al.* 2012). The training of the network was performed using the error back-propagation (BP) algorithm, the most popular one in the water resources literature (Cigizoglu 2004). The main working process of the BP network is the input units, hidden units and output units which are completely connected in a feed-forward way, and each error signal can be back-propagated from the output to input layer to adjust the weight. These processes can continue iteratively until the network output matches the target within a specified level (Wei *et al.* 2012). Data normalization is preprocessing of transforming data set into small intervals usually [−1, 1] or [0, 1] before training a network. This process is important to insure that all variables receive

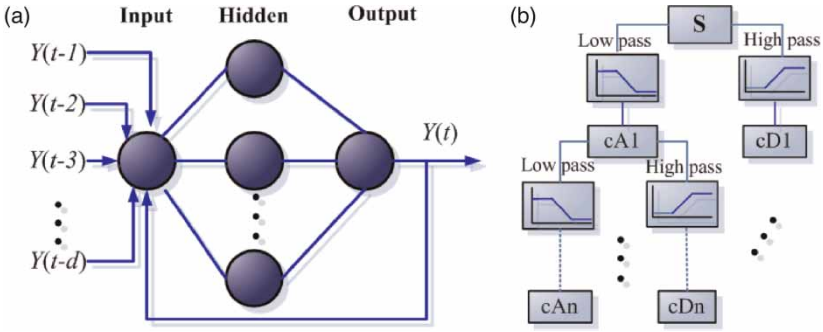

Figure 2 | A sketch of (a) a standard nonlinear autoregressive network (NARN), (b) wavelet decomposition analysis.

equal attention and that the efficiency of the training network is improved (Dawson & Wilby 2001; Wei *et al.* 2012). Various methods can be used for data normalization (Wang *et al.* 2006).

Overfitting is one of the common problems that usually occurs during neural network training, in which the error on the training set is driven to a very small value, but when new data are presented to the network the error is large (Wei *et al.* 2012). Two methods are usually used for improving the network generalization: early-stopping and regularization. In Matlab, Bayesian regularization (BR) (MacKay 1991) was used to improve the network generalization, because it enables the determination of the optimal regularization parameters in an automated fashion (Beale *et al.* 2010). In the hydrological literature, Wei *et al.* (2012) firstly employed this method to compare the modeling results of a WNN model without generalization and with BR. This regularization method is usually expressed by:

$$\text{msereg} = \gamma \text{mse} + (1-\gamma)\text{msw} \tag{2}$$

$$\text{msw} = \frac{1}{n}\sum_{i=1}^{n} w_j^2 \tag{3}$$

where msereg means the regularization mse, γ is the performance ratio, w, msw are the weights of network mse, and their average, respectively.

Discrete wavelet analysis

Wavelet analysis (WA) is a promising time-frequency technique for signal analysis, and has several advantages over the traditional Fourier analysis (FA) and the adapted 'short-time Fourier transform'. WA is an improved version of the 'short-time Fourier transform' which can elucidate time characteristics in data (Daubechies 1990). WA is a windowing technique with variable-sized regions, which allows the use of long time intervals for low-frequency information, and shorter regions for high-frequency information. In this sense, it is usually characterized as a *mathematical microscope*. There are two main types of wavelet transformation methods: continuous wavelet transformation (CWT) and discrete wavelet transformation (DWT). CWT is usually defined by the following equation in the literature (e.g. Kişi 2008; Partal 2009):

$$\begin{aligned} W_f(a,\tau) &= \int_{-\infty}^{+\infty} f(t)\psi_{a,\tau}*(t)\,\mathrm{d}t \text{ with } \psi_{a,\tau}*(t) \\ &= |a|^{-1/2}\psi*\left(\frac{t-\tau}{a}\right), \quad a \in R, \tau \in R, a \neq 0 \end{aligned} \tag{4}$$

where $W_f(a,\tau)$ are the wavelet coefficients, t is time interval, $f(t)$ presents the input signal, $\psi(t)$ is a base wavelet function also called the mother wavelet, * corresponds to the complex conjugate of $\psi(t)$, a is a scaling factor stretching or compressing the mother wavelet to the frequency of the signal, and τ is a translating factor shifting the mother wavelet to the time domain of the signal. Equation (4) presents that the wavelet transformation is the decomposition of $f(t)$ under different resolution levels (scales).

However, continuous wavelet transformation calculates wavelet coefficients at every possible scale, which is a time consuming process producing excessive (often redundant) amounts of data. Thus DWT is usually used in practice. It

can be expressed by:

$$\text{Let } a = a_0^j,\ \tau = ka_0^j\tau_0,\ a_0 > 0,\ \tau_0 \in R,\ \forall j, k = 0, 1, 2, 3, \ldots, m \in Z \tag{5}$$

where $(\psi_{a,\tau}(t))$ can be written as:

$$\psi_{j,k}(t) = a_0^{-j/2}\psi[a_0^{-j}(t - ka_0^j\tau_0)] = a_0^{-j/2}\psi(a_0^{-j}t - k\tau_0) \tag{6}$$

Therefore, the discrete wavelet transformation can correspondingly be expressed by:

$$W_f(j, k) = a_0^{-j/2}\int_{-\infty}^{+\infty} f(t)\psi*(a_0^{-j}t - k\tau_0)\,\mathrm{d}t \tag{7}$$

The simplest and most efficient practice is that scales and position are selected based on the powers of two logarithms, called *dyadic* scales and positions (Mallat 1989). That is, let $a_0 = 2$, $\tau_0 = 1$, then the discrete wavelet transformation becomes a binary one:

$$W_f(j, k) = 2^{-j/2}\int_{-\infty}^{+\infty} f(t)\psi*(2^{-j}t - k)\,\mathrm{d}t \tag{8}$$

$W_f\ (a,\ \tau)$ and $W_f\ (j,\ k)$ can reflect the characters of the original time series in frequency (a or j) and time domains (τ or k). When a or j is small, the frequency resolution is very low, but the time domain is very high. When a or j become large, the frequency resolution is high, but the time domain is low.

As a discrete time series $f\ (t)$, in which $f\ (t)$ occurs at discrete integer time steps t, the dyadic discrete wavelet transformation can be written as:

$$W_f(j, k) = \sum_{j,k\in z} f(t)2^{-j/2}\psi(2^{-j}t - k) \tag{9}$$

The input signal can be reconstructed using the equation:

$$f(t) = \sum_{j,k\in z} W_f(j, k)\psi_{i,k}(t) \tag{10}$$

In this equation, wavelet coefficients $W_f\ (j,k)$ are divided into an approximation (or low frequency) coefficient (cA_n) at level n through a low pass filter $l(\psi_{i,k}(t))$, and detail (or high frequency) coefficients (cD1, cD2, cD3,…, cD_n) at different levels 1, 2,…, n through a high pass filter $h(\psi_{i,k}(t))$ (Figure 2(b)). cA_n provides background information on the original signal, while cD1, cD2, cD3,…,cD_n contains the detail information on the original signal such as period and break, jump. Then the original signal can be expressed as:

$$f(t) = cA_n l(\psi_{i,k}(t)) + \sum_{n=1} cD_n h(\psi_{i,k}(t)) \tag{11}$$

or simplified to the form:

$$f(t) = A_n(t) + \sum D_n(t) \tag{12}$$

where $A_n(t)$ is the approximation of the original signal at level n, and $D_n(t)$ is the details of the original signal at levels $n = 1, 2, 3,\ldots, m$.

RESULTS AND DISCUSSIONS

The two modeling techniques, discrete wavelet analysis (DWA) and NAR artificial neural network, were integrated to produce a Wavelet-NAR network (WNAR) hybrid model. This hybrid modeling approach integrates the advantages of wavelet transformation and artificial neural network models, and this process included two steps: (1) The river discharge time series, i.e. a signal, was decomposed with DWT into an approximation (An) with low frequency, and details (D1, D2,…, Dn) with high frequency; and (2) these subseries were collectively used as the input for the NAR network, with the original series as the target. We used these subseries collectively so that all the information of the original is maintained.

We compared Wavelet-NAR hybrid modeling results with those from the single NAR network in two steps. Firstly, we compared them under the same network structure. Secondly, we compared the results of Wavelet-NAR with those of the improved NAR network. The correlation coefficient (R), root mean squared error ($RMSE$), mean absolute error (MAE), and mean absolute relative error ($MARE$) were used to evaluate the model accuracy in general. In addition,

extreme error indices (Wei *et al.* 2012), maximum absolute relative error (ARE_{mx}) and minimum absolute relative error (ARE_{mn}) were used to measure the model performances to predict river discharge at certain time points. Absolute relative errors of maximum discharge (ARE_{mxd}) and minimum discharge (ARE_{mnd}) are employed to assess the model accuracy to predict the highest and lowest discharges. These evaluation indices are defined by Equations (13)–(18).

$$R = \frac{\sum_{t=1}^{n}(Y_t - \bar{Y}_t)(\hat{Y}_t - \bar{\hat{Y}}_t)}{\sqrt{\sum_{t=1}^{n}(Y_t - \bar{Y}_t)^2 \sum_{t=1}^{n}(\hat{Y}_t - \bar{\hat{Y}}_t)^2}} \quad (13)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{T}(\hat{Y}_t - Y_t)^2} \quad (14)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}\left|\hat{Y}_t - Y_t\right| \quad (15)$$

$$\mathrm{MARE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{Y}_t - Y_t}{Y_t}\right| \quad (16)$$

$$\mathrm{ARE_{mx}} = \max\left|\frac{\hat{Y}_t - Y_t}{Y_t}\right| \text{ and } \mathrm{ARE_{mn}} = \min\left|\frac{\hat{Y}_t - Y_t}{Y_t}\right| \quad (17)$$

$$\mathrm{ARE_{mx\,d}} = \left|\frac{\hat{Y}_{\max} - Y_{\max}}{Y_{\max}}\right| \text{ and } \mathrm{ARE_{mn\,d}} = \left|\frac{\hat{Y}_{\min} - Y_{\min}}{Y_{\min}}\right| \quad (18)$$

where t is time unit, $\hat{Y}_t$ is simulated river discharge, Y_t is observed discharge, $\bar{Y}_t$ and $\bar{\hat{Y}}_t$ are the mean ones, $Y_{\max}$ and $Y_{\min}$ are the maximum and minimum observed discharges, $\hat{Y}_{\max}$ and $\hat{Y}_{\min}$ are the predicted values of the maximum and minimum observed discharges.

Wavelet decomposition

Two important issues in wavelet analysis are to choose the wavelet type and the appropriate scale numbers in wavelet multi-resolution decomposition. There are hundreds of mother wavelets available. There is no common principle for the selection of mother wavelet. As Goring (2006) stated, which mother wavelet to use is one of the questions that perplexes beginners, but with experience you find it makes little difference. Goring (2006) preferred to use Daubechies Wavelet 5 (db5) for almost everything. We employed db5 as the mother wavelet to decompose the series because: (1) Daubechies wavelet is one of a widely used wavelet family, which was formulated by the Belgian mathematician Ingrid Daubechies in 1988 (Mohamed & Atta 2010; Wei *et al.* 2012); (2) our data set is monthly river discharge data sets, which are comparably smooth; and (3) db5 is one of the commonly used wavelets in db wavelet family (Wei *et al.* 2012), which is suitable to decompose comparably smooth data set. The rule for the appropriate scales of wavelet numbers is thought that the largest scales should be shorter than the size of testing data (predictions) (Wu *et al.* 2009a). Wu *et al.* (2009a) have not given any explanation about this rule, but this concept is clear. For example, suppose that we have a new data set containing only 40 numbers used to test a WNN model prediction ability. According to discrete wavelet decomposition method – the powers of two logarithms, or *dyadic*, 40 data number can only be decomposed using the largest scale 5 (2^1-2^2-2^3-2^4-2^5). If scale 6 is used, 2^6 is 64, which means that the data set has not enough data for this decomposition. In our case, the testing sizes of three flow series are 48 months (4 years), and thus the largest scales are chosen as 5 for the three data series. Therefore, the flow data sets are decomposed into various details (Ds) and an approximation (A5) at five resolution levels (2^1-2^2-2^3-2^4-2^5) using db5 DWT. The new decomposed subseries present variations of the original times series on different periods. As an example, Figure 3 illustrates the decomposition of the original river discharge series in the Xianyang gage. The six wavelet components (five details and one approximation) of the wavelet decomposition displayed 2-month mode (D1), 4-month mode (D2), 8-month mode (D3), 16-month mode (D4), 32-month mode (D5), and the approximation mode (A5) of the flow time series.

The basic algorithm for the DWT is based on a simple scheme: convolution and downsampling, which is not limited to dyadic length. Border distortions usually arise when a convolution is performed on finite-length signals (Beale *et al.* 2010). Cui and Song (2007) further stated that orthonormal dbN wavelet functions are all asymmetric except Haar wavelet

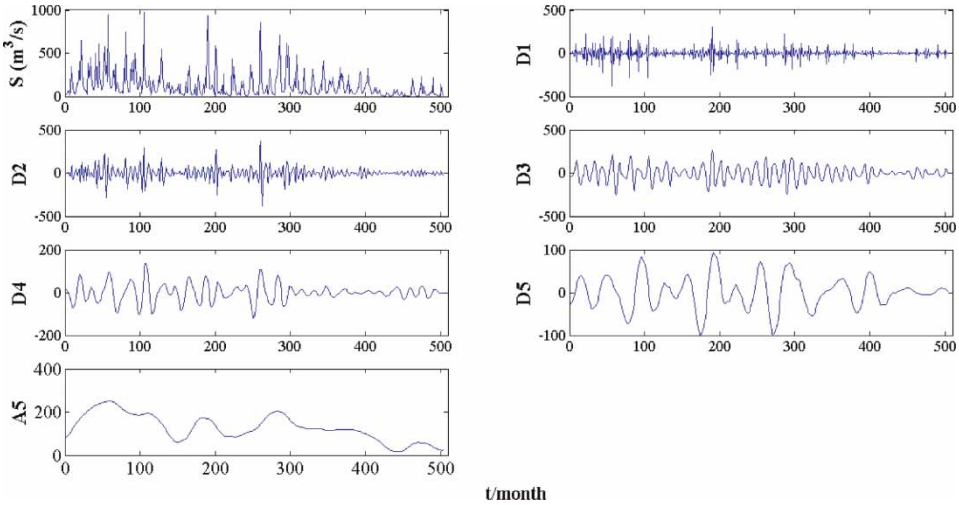

Figure 3 | Wavelet decomposition of the original river discharge, taking Xianyang discharge series for example, with level5 using Daubechies Wavelet No. 5 (db5).

basis (i.e. db1), and a comparatively great distortion and error, especially in the border will occur if such asymmetrical wavelet basis is selected. Many methods are used to solve such boundary effects, including zero-padding, smooth padding, periodic extension, and boundary value replication (symmetrization). Symmetrization is the default mode of the wavelet transform in the Matlab Wavelet Toolbox. This method works well in general for images though it has the disadvantage of artificially creating discontinuities of the first derivative at the border (Beale *et al.* 2010). We also synthesized the signal by reconstructing it from the wavelet coefficients to test if the decomposition is valid without boundary effects. The maximum absolute errors of the synthesis are 6.8494×10^{-9}, 4.7148×10^{-9} and 2.9681×10^{-9} for Huaxian series, Lintong series and Xiangyang series, and such small errors revealed that the decomposition is very valid.

Network training processes

We used a three layer network structure (p, d, m, n) for training both WNAR and NAR network models, in which p is the input vectors, d is the feedback delay, m is the neurons in the hidden layers and n is the output vectors. While two methods, early stopping and regularization, could be used to solve the overfitting problem as both these methods can ensure network generalization when applied properly; for a small data set, however, Bayesian regularization provides a much better generalization performance than early stopping (Wei *et al.* 2012). This is because Bayesian regularization does not need a validation data set separated from the training data set, and consequently can utilize all available data. BR was implemented in the training function *trainbr*, via the Matlab Neural Network Toolbox User's Guide (Beale *et al.* 2010).

Before network training, the data series of river discharge (Q_t) was normalized into the range [−1, 1] using Matlab function mapminmax because that BR algorithm generally works best when the network inputs and targets are scaled approximately in the range [−1, 1] (Beale *et al.* 2010; Wei *et al.* 2012). NAR network is a dynamic network, where the output is fed back to the input of the feedforward network (closeloop) (Figure 2), while a static backpropagation network (open loop) for training is more accurate. This is because the true output is available during the training, and thus the true output can be used instead of feeding back the estimated output (Beale *et al.* 2010). The training process started with 1 neuron ($m = 1$) and feedback delay from 1 ($d = 1$), and the neuron numbers and feedback delay ranges were increased progressively if the network performance could not be improved after 50 time training. The process was continued until the best performance was met in terms of significance of time delay, R, RMSE, MAE and MARE.

Significance of time delay

The adequacy of the model was evaluated by checking if time delay in the model is significant in terms of error

autocorrelation function (EACF) and input-error cross-correlation function (IECF). EACF describes how the simulation errors are related in time, and IECF displays how the errors are correlated with the input sequences. For a perfect prediction model, there should be only one nonzero value of the autocorrelation function, and all of the input-error cross-correlations should be zero.

As an example, Figure 4 illustrates the plots of EACF and IECF of the river discharge series of the Xianyang gage. After feedback delay 3 ($d = 3$) and the 10 neurons ($m = 10$) were used in the hidden layer, error autocorrelations were approximately in the 95% confidence limits around zero besides for the one at zero lag (Figure 4(a)) and all of the input-error cross-correlations fell within the confidence bounds around zero (Figure 4(b)), thus WNAR model is adequate. For the single NARN1 model under the same structure ($d = 3$, $m = 10$, $n = 1$), on the contrary, the time delay is not significant since there are error autocorrelations at lags ± 12 (Figure 4(c)) and significant input-error correlations (Figure 4(d)) at lags ± 10, ± 11 and ± 12. In this sense, the NARN model should be improved, and improvement is usually done by increasing the number of delay or/

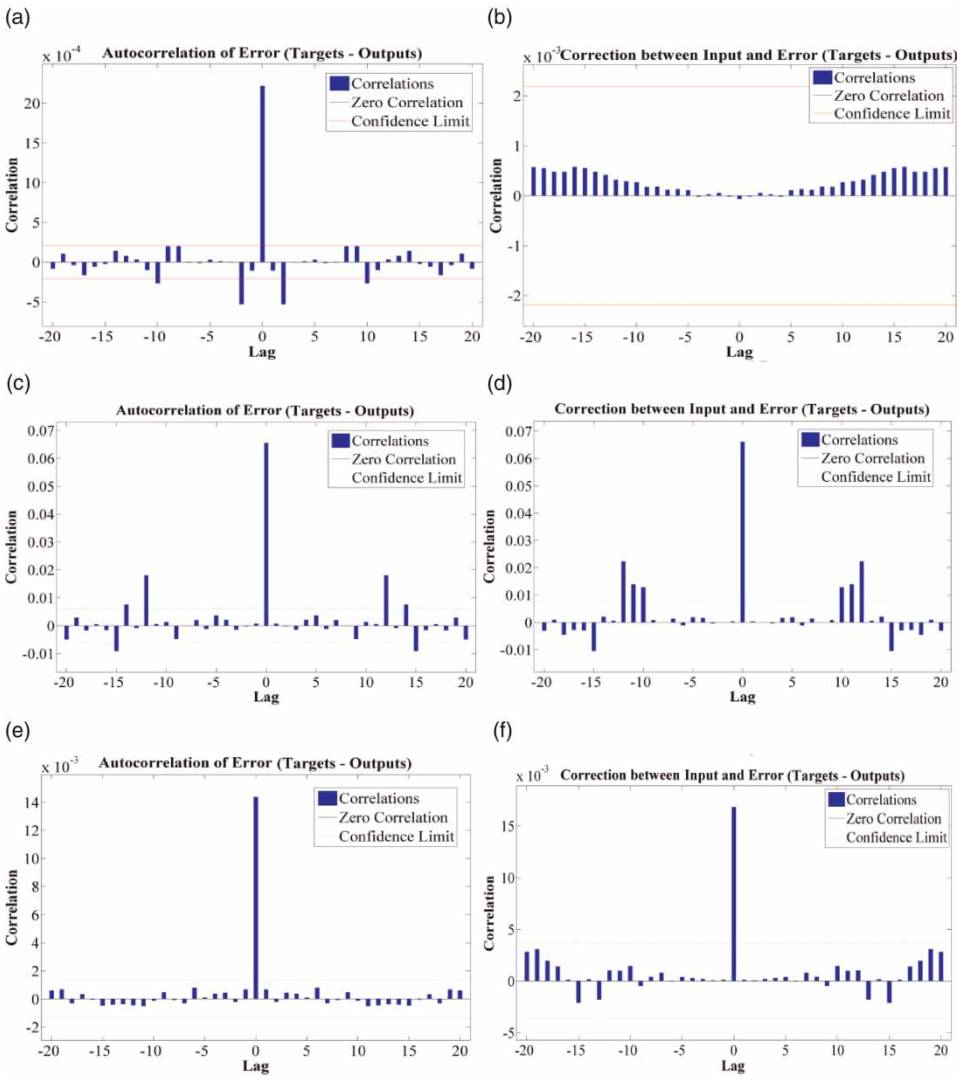

Figure 4 | Plots of the significance of feedback time delay of the models, taking Xianyang discharge series for an instance: (a) error autocorrelation function (EACF) of WNAR model, (b) input-error cross-correlation function (IECF) of WNAR model, (c) EACF of a single NARN model (NARN1), (d) IECF of NARN1, (e) EACF of the improved NARN model (NARN2), and (f) IECF of NARN2.

and numbers of neurons in hidden layer. All the error autocorrelations and the input-error cross-correlations of the improved NARN model (NARN2) fall within the confidence bounds around zero (Figure 4(e), 4(f)) when 13 neurons were used in the hidden layer and feedback delay reached the range from 1 to 12. However, 13 neurons are high, which might result in an unstable ANN.

Training results evaluation

The results of model performances of training are presented in Table 2. In terms of largest correlation coefficients, the WNAR hybrid models displayed much stronger correlation between observed values and fitting value than those of NARN1 and NARN2 models. For instance, the stronger correlation of the WNAR hybrid model and its comparison to two single NARN models for Lintong gage are illustrated using regression plots (Figure 5), in which the network outputs are plotted with respect to targets. For a perfect fit, the data should fall along a 45 degree line, where the network outputs are equal to the targets. R value of 0.99 in training indicated that the goodness-of-fit of WNAR was best (Figure 5(a)), though the fit of NARN2 was reasonably good due to R value of 0.87 (Figure 5(e)).

In addition, WNAR model reached the best fitting performances in terms of evaluation indices, for example, RMSE (0.045), MAE (0.034) and MARE (0.079) under the network structure ($d = 3$, $m = 10$, $n = 1$) in the Lintong gage (Table 2). In contrast, the single NAR1 has poor performance to fit river discharge under the same structure due to bigger error indices (RMSE = 0.257, MAE = 0.167 and MARE = 2.453). While the improved single NARN model (NARN2) could significantly increase the simulation performances in terms of the evaluation indices (RMSE = 0.180, MAE = 0.118 and MARE = 2.115), NARN was still less precise to fit historical data comparing with hybrid model WNAR.

Model testing evaluation

Open loop network architecture is useful for network training, but closed loop architecture is more useful for multi-step-ahead predictions. Function *closeloop* was used to convert open loop to close loop in Matlab. The general testing evaluation results for model prediction ability of WNAR, NARN1 and NARN2 using the observed discharge data from the last 48 months are summarized in Table 3. The testing results were very similar with those of training. In Lintong gage, for instance, regression plots comparing the degrees of correlations between observed values and predicted values of the WNAR hybrid model and the two single NARN

Table 2 | Evaluation results of the WNAR and NAR network training performances

Gage	Mode	Model structure[a]	R[b]	RMSE[c]	MAE[d]	MARE[e]
Huaxian	WNAR[f]	(6,1:3,10,1)	0.990	0.038	0.029	0.054
	NARN1[g]	(1,1:3,10,1)	0.610	0.216	0.134	0.435
	NARN2[h]	(1,1:12,13,1)	0.844	0.147	0.092	0.299
Lintong	WNAR[f]	(6,1:3,10,1)	0.990	0.045	0.034	0.079
	NARN1[g]	(1,1:3,10,1)	0.596	0.257	0.167	2.453
	NARN2[h]	(1,1:12,13,1)	0.866	0.180	0.118	2.115
Xianyang	WNAR[f]	(6,1:3,10,1)	0.989	0.047	0.035	0.079
	NARN1[g]	(1,1:3,10,1)	0.590	0.256	0.160	0.745
	NARN2[h]	(1,1:12,13,1)	0.927	0.120	0.081	0.241

[a]Structure (*m, d, n, p*): the first number (*m*), the second one (*d*), the third one (*n*) and the fourth one (*p*) signify input vectors, feedback delay, neurons hidden layer, and output vectors, respectively.
[b]Correlation coefficient.
[c]Root mean squared error.
[d]Mean absolute error.
[e]Mean absolute relative error.
[f]Wavelet-NAR network hybrid model.
[g]A single NAR network model.
[h]The improved NAR network model.

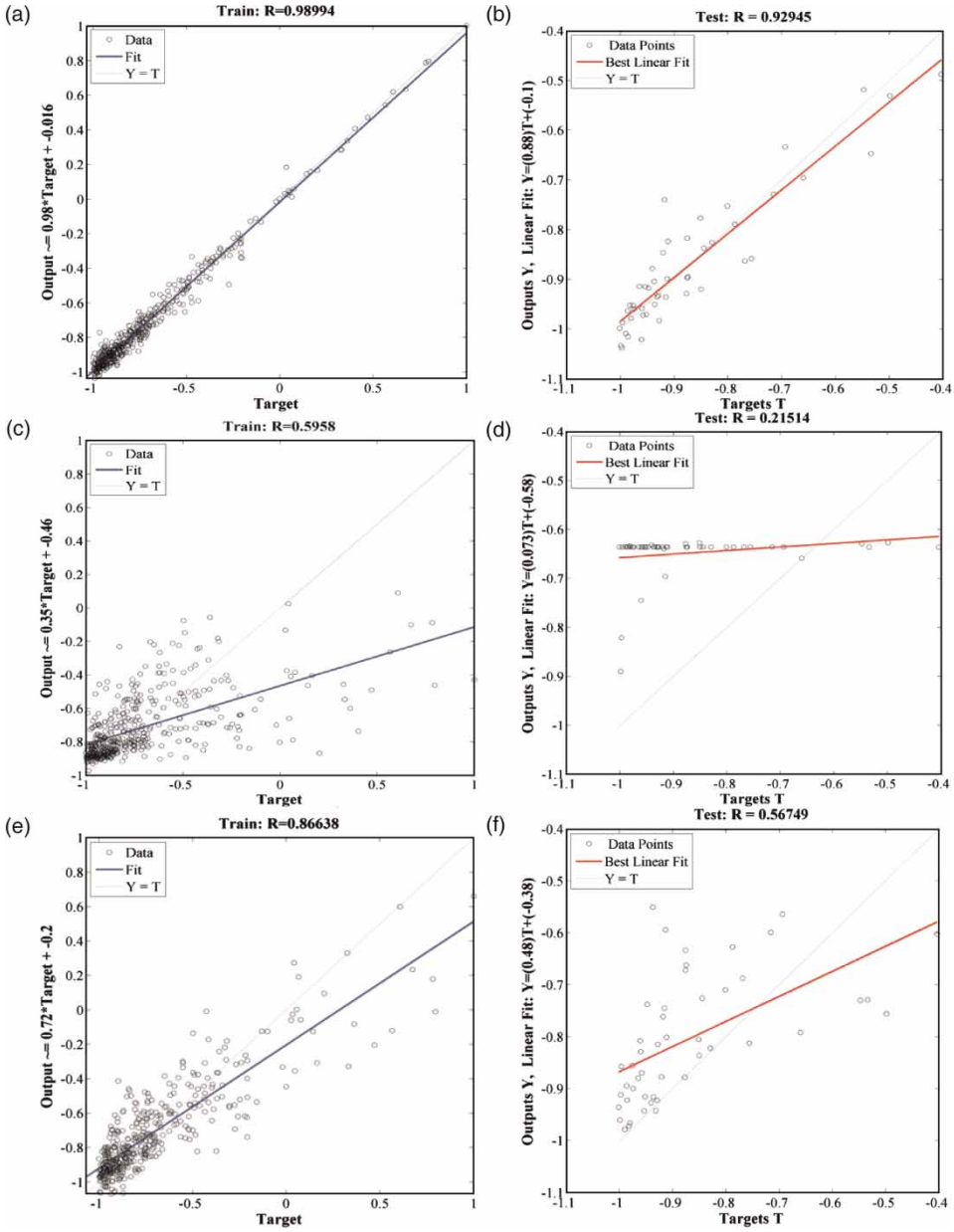

Figure 5 | Regression plots displaying the network outputs with respect to targets, taking Xianyang discharge series for an instance: (a) training performance of model WNAR, (b) testing result of model WNAR, (c) training performance of model NARN1, (d) testing results of model NARN1, (e) training performance of model NARN2, and (f) testing results of model NARN2.

models are illustrated in (Figure 5). For WNAR, R value of 0.929 in testing indicated that the goodness-of-fit was very good (Figure 5(b)). In contrast, NARN1 and NARN2 exhibited a less precise performance for river discharge prediction in terms of lower R values (0.215 and 0.567) in testing (Figure 5(d), 5(f)) even though the higher R value of 0.866 in training indicated that NARN2 was also able to fit historical data well (Figure 5(e)). In addition, the WNAR model significantly improved the general prediction performance of intermediate and moderate discharge in terms of lowest evaluation indices (RMSE = 0.053, MAE = 0.040 and MARE = 0.050) (Table 3). For further comparison of extreme error indices, it indicated that the lowest ARE_{mx} (0.211), ARE_{mn} (0.001), and ARE_{mxd} (0.122) and ARE_{mnd}

Table 3 | Evaluation results of the WNAR and NAR network prediction performance testing

Gage	Mode	Model structure[a]	R^2[b]	RMSE[c]	MAE[d]	MARE[e]
Huaxian	WNAR[f]	(6,1:3,10,1)	0.912	0.047	0.034	0.042
	NARN1[g]	(1,1:3,10,1)	0.117	0.233	0.220	0.241
	NARN2[h]	(1,1:12,13,1)	0.417	0.200	0.156	0.176
Lintong	WNAR[f]	(6,1:3,10,1)	0.929	0.053	0.040	0.050
	NARN1[g]	(1,1:3,10,1)	0.215	0.262	0.245	0.277
	NARN2[h]	(1,1:12,13,1)	0.567	0.313	0.245	0.279
Xianyang	WNAR[f]	(6,1:3,10,1)	0.919	0.050	0.038	0.045
	NARN1[g]	(1,1:3,10,1)	0.128	0.331	0.311	0.329
	NARN2[h]	(1,1:12,13,1)	0.321	3.067	2.375	2.619

[a]Structure (*m, d, n, p*): the first number (*m*), the second one (*d*), the third one (*n*) and the fourth one (*p*) signify input vectors, feedback delay, neurons hidden layer, and output vectors, respectively.
[b]Correlation coefficient.
[c]Root mean squared error.
[d]Mean absolute error.
[e]Mean absolute relative error.
[f]Wavelet-NAR network hybrid model.
[g]A single NAR network model.
[h]The improved NAR network model.

(0.061) also proved that the WNAR hybrid model had much stronger prediction ability than single NARN models, especially for the highest and lowest discharge events, even though the NARN2 model also increased the prediction ability of the highest and lowest discharge (Table 4).

Prediction of river discharge

The visualized comparison results of the 48 months (i.e. 4-year period) ahead predictions of the WNAR hybrid model and observations are illustrated in Figure 6. The detailed comparison of the predicted results of the

Table 4 | Extreme error indices for WNAR and NAR network prediction performance testing

Gage	Model	Model structure[a]	ARE^b_{mx}	ARE^c_{mn}	ARE^d_{mxd}	ARE^e_{mnd}
Huaxian	WNAR[f]	(6,1:3,10,1)	0.208	0.001	0.185	0.098
	NARN1[g]	(1,1:3,10,1)	0.376	0.052	0.132	0.086
	NARN2[h]	(1,1:12,13,1)	0.477	0.010	0.154	0.022
Lintong	WNAR[f]	(6,1:3,10,1)	0.211	0.001	0.122	0.061
	NARN1[f]	(1,1:3,10,1)	0.574	0.003	0.173	0.207
	NARN2[h]	(1,1:12,13,1)	0.825	0.001	0.133	0.239
Xianyang	WNANR[f]	(6,1:3,10,1)	0.001	0.126	0.110	0.052
	NARN1[g]	(1,1:3,10,1)	0.414	0.024	0.240	0.234
	NARN2[h]	(1,1:12,13,1)	10.06	0.063	3.712	1.134

[a]Structure (*m, d, n, p*): the first number (*m*), the second one (*d*), the third one (*n*) and the fourth one (*p*) signify input vectors, feedback delay, neurons hidden layer, and output vectors, respectively.
[b]Maximum absolute relative error.
[c]Minimun absolute relative error.
[d]Absolute relative error of maximum river discharge.
[e]Absolute relative error of minimum river discharge.
[f]WNAR = Wavelet-NAR network hybrid model.
[g]A single NAR network model.
[h]The improved NAR network model.

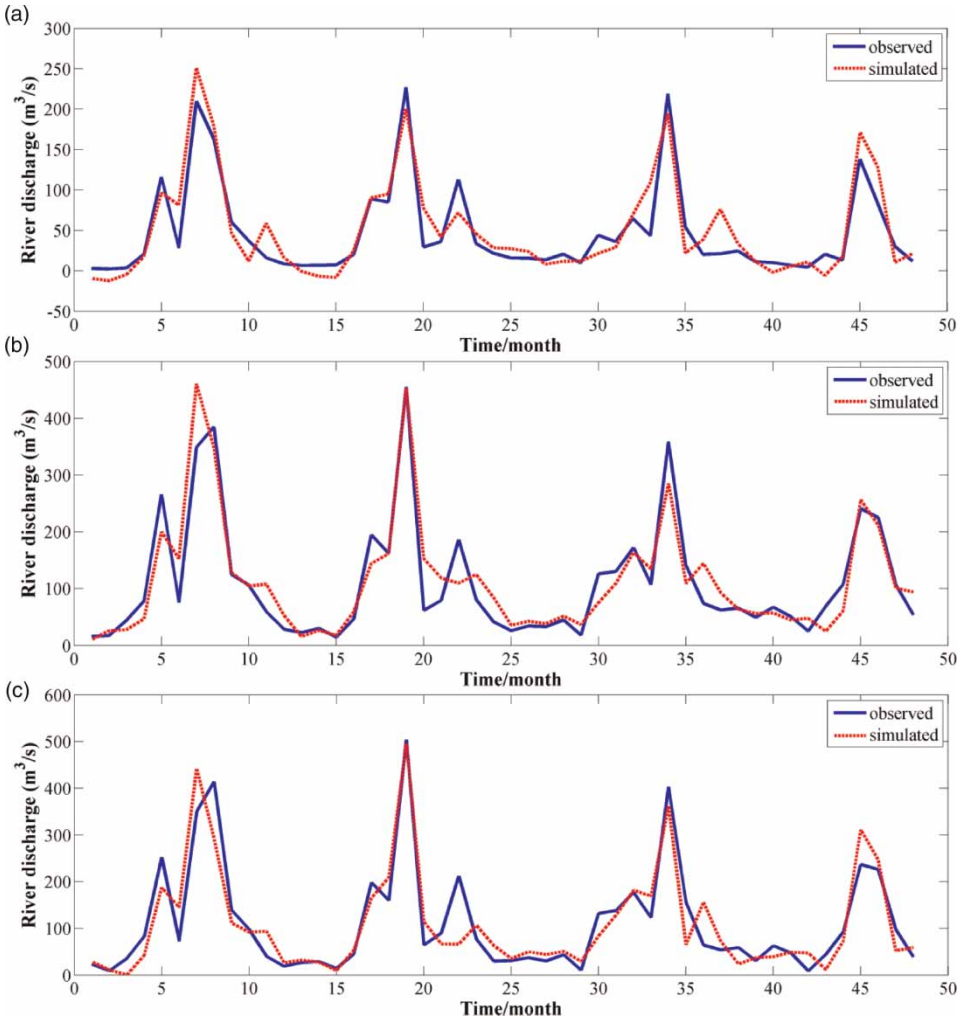

Figure 6 | Plots of the comparison results between the predictions of WNAR model and observations: (a) Xianyang gage, (b) Lintong gage, and (c) Huaxian gage.

WNAR and two NARN models and their evaluations are displayed in Table 5. In general, these results further revealed that the 48 month ahead predictions of WNAR were much more accurate than those of the single NARN models based on both maximum, minimum and mean absolute relative errors (ARE_{mx}, MARE and ARE_{mn}), and the AREs of highest and lowest discharge (ARE_{mxd} and ARE_{mnd}). In addition, comparing the prediction results of WNAR hybrid model for the three gages, it displayed that WNAR has the best performance to accurately predict Lintong discharge due to the lowest evaluation indices, $ARE_{mx} = 1.48$, $ARE_{mn} = 0.00$, $MARE = 0.40$, $ARE_{mxd} = 0.00$ and $ARE_{mnd} = 0.20$. In contrast, WNAR has a comparatively weaker performance to accurately predict Xianyang discharge due to the negative prediction values in some years.

A comparison between WNAR and WNN

Multi-layer Feed-forward Neural Network (MLFN) is used in the traditional WNN model. MLFN is a static network, but NARN is one type of dynamic neural network, the extension of static neural network by considering time. Therefore, the WNAR model should have better performance in theory. In the previous studies of WNN hydrid modeling approaches, Wei *et al.* (2012) proposed a WNN method combining db5 wavelet and a feed-forward multi-layer perceptron ANN with back-propagation algorithm and

Table 5 | Comparison results of WNAR model and single NAR network (NARN) models ($\times m^3 s^{-1}$)

	Huaxian gage							Lintong gage							Xianyang gage						
		WNAR[b]		NARN1[c]		NARN2[d]			WNAR[b]		NARN1[c]		NARN2[d]			WNAR[b]		NARN1[c]		NARN2[d]	
Time/ month	O[a]	S[e]	ARE[f]	S[e]	ARE[f]	S[e]	ARE[f]	O[a]	S[e]	ARE[f]	S[e]	ARE[f]	S[e]	ARE[f]	O[a]	S[e]	ARE[f]	S[e]	ARE[f]	S[e]	ARE[f]
1	23.6	28.6	0.2	136.0	4.76	131.0	4.55	16.2	9.9	0.39	96.5	4.95	44.4	1.74	2.7	(9.3)	4.38	74.9	26.3	(15)	6.48
2	9.3	11.1	0.20	154.0	15.65	240.0	24.95	17.3	25.8	0.49	147.3	7.52	80.6	3.66	2.3	(12.4)	6.32	82.4	34.52	(75)	33.29
3	35.5	0.2	1.00	223.2	5.29	51.1	0.44	44.3	27.7	0.37	203.3	3.59	157.0	2.54	3.5	(4.6)	2.34	122.9	34.51	(87)	26.03
4	82.5	41.9	0.49	263.0	2.19	27.7	0.66	77.9	46.6	0.40	239.6	2.08	203.5	1.61	20.6	17.3	0.16	140.8	5.83	(136)	7.60
5	252.0	187.2	0.26	298.3	0.18	37.4	0.85	266.0	200.2	0.25	267.2	0.00	168.8	0.37	116.0	96.8	0.17	165.3	0.43	(211)	2.82
6	72.6	144.5	0.99	315.4	3.34	91.4	0.26	75.7	152.7	1.02	281.4	2.72	191.3	1.53	28.3	81.1	1.87	178.4	5.30	(227)	9.02
7	350.0	442.4	0.26	323.6	0.08	126.0	0.64	349.0	460.8	0.32	288.9	0.17	214.3	0.39	210.0	251.2	0.20	190.4	0.09	(377)	2.79
8	414.0	291.7	0.30	324.7	0.22	87.0	0.79	385.0	347.6	0.10	290.7	0.25	195.5	0.49	163.0	179.3	0.10	196.6	0.21	(565)	4.47
9	139.0	111.8	0.20	323.1	1.32	12.3	0.91	125.0	128.6	0.03	290.2	1.32	159.1	0.27	60.4	47.0	0.22	200.9	2.33	(819)	14.57
10	97.9	91.9	0.06	320.7	2.28	45.0	0.54	106.0	104.3	0.02	288.5	1.72	105.1	0.01	37.1	11.8	0.68	202.8	4.47	(241)	7.49
11	39.9	93.8	1.35	318.6	6.98	87.9	1.20	59.3	108.0	0.82	286.9	3.84	68.9	0.16	16.1	58.8	2.65	203.8	11.66	1080	66.08
12	19.0	26.3	0.39	317.1	15.69	55.0	1.89	28.6	53.1	0.86	285.5	8.98	36.0	0.26	8.7	16.8	0.94	203.9	22.57	1702	195.78
13	26.9	32.0	0.19	316.3	10.76	40.2	0.49	22.7	15.8	0.31	284.6	11.54	31.6	0.39	6.8	(0.5)	1.07	203.9	29.16	1738	256.09
14	28.8	27.6	0.04	315.9	9.97	9.9	0.66	29.9	26.5	0.11	284.1	8.50	40.1	0.34	6.9	(6.7)	1.97	203.7	28.44	(1622)	235.43
15	14.0	9.7	0.31	315.9	21.56	18.4	0.31	14.6	17.6	0.20	283.9	18.45	62.9	3.30	7.1	(8.2)	2.15	203.6	27.55	(1506)	212.27
16	45.5	53.5	0.18	316.0	5.94	23.6	0.48	46.4	59.4	0.28	283.9	5.12	111.4	1.40	19.9	24.8	0.24	203.4	9.22	(1103)	56.41
17	198.0	164.8	0.17	316.1	0.60	9.3	0.95	195.0	144.4	0.26	283.9	0.46	153.7	0.21	88.9	90.3	0.02	203.3	1.29	3865	42.48
18	160.0	209.4	0.31	316.2	0.98	35.5	0.78	162.0	161.7	0.00	284.0	0.75	229.4	0.42	84.9	95.1	0.12	203.3	1.39	2485	28.26
19	504.0	495.6	0.02	316.2	0.37	82.5	0.84	455.0	452.9	0.00	284.0	0.38	308.6	0.32	227.0	200.1	0.12	203.2	0.10	9	0.96
20	63.9	114.5	0.79	316.3	3.95	252.0	2.94	61.6	152.9	1.48	284.1	3.61	346.8	4.63	29.3	77.8	1.66	203.2	5.94	316	9.78
21	90.2	66.3	0.26	316.3	2.51	72.6	0.20	79.2	118.5	0.50	284.1	2.59	314.6	2.97	36.0	41.9	0.16	203.2	4.65	(79)	3.19
22	212.0	66.1	0.69	316.3	0.49	350.0	0.65	186.0	109.5	0.41	284.1	0.53	245.9	0.32	113.0	71.8	0.36	203.2	0.80	1201	9.62
23	77.3	106.6	0.38	316.3	3.09	414.0	4.36	80.6	124.6	0.55	284.1	2.53	162.1	1.01	33.9	46.2	0.36	203.2	5.00	(758)	23.35
24	30.0	63.2	1.11	316.3	9.54	139.0	3.63	41.4	84.1	1.03	284.1	5.86	103.9	1.51	21.8	28.4	0.30	203.2	8.32	(1811)	84.09
25	30.9	35.3	0.14	316.3	9.24	97.9	2.17	25.8	35.7	0.38	284.1	10.01	72.7	1.82	15.9	27.4	0.73	203.2	11.78	1304	81.04
26	37.2	49.8	0.34	316.3	7.50	39.9	0.07	34.3	42.4	0.24	284.1	7.28	89.5	1.61	15.7	23.9	0.52	203.2	11.95	(539)	35.36
27	30.2	43.8	0.45	316.3	9.47	19.0	0.37	33.1	38.4	0.16	284.1	7.58	122.0	2.69	13.6	8.2	0.40	203.2	13.94	(2357)	174.31
28	43.5	50.7	0.16	316.3	6.27	26.9	0.38	44.9	51.8	0.15	284.1	5.33	141.7	2.16	20.6	11.8	0.43	203.2	8.87	645	30.29
29	10.7	29.5	1.75	316.3	28.56	28.8	1.69	18.2	36.5	1.00	284.1	14.61	120.6	5.63	9.8	11.9	0.21	203.2	19.72	(334)	35.06
30	132.0	83.9	0.36	316.3	1.40	14.0	0.89	126.0	75.3	0.40	284.1	1.25	136.1	0.08	44.1	21.5	0.51	203.2	3.61	1182	25.79
31	138.0	128.4	0.07	316.3	1.29	45.5	0.67	130.0	109.5	0.16	284.1	1.19	217.5	0.67	36.1	28.8	0.20	203.2	4.63	2110	57.44
32	177.0	181.7	0.03	316.3	0.79	198.0	0.12	172.0	163.9	0.05	284.1	0.65	290.2	0.69	64.5	69.6	0.08	203.2	2.15	(207)	4.21

(*continued*)

Table 5 | continued

	Huaxian gage							Lintong gage							Xianyang gage						
		WNAR[b]		NARN1[c]		NARN2[d]			WNAR[b]		NARN1[c]		NARN2[d]			WNAR[b]		NARN1[c]		NARN2[d]	
Time/ month	O[a]	S[e]	ARE[f]	S[e]	ARE[f]	S[e]	ARE[f]	O[a]	S[e]	ARE[f]	S[e]	ARE[f]	S[e]	ARE[f]	O[a]	S[e]	ARE[f]	S[e]	ARE[f]	S[e]	ARE[f]
33	123.0	169.0	0.37	316.3	1.57	160.0	0.30	107.0	134.5	0.26	284.1	1.66	285.4	1.67	43.3	109.0	1.52	203.2	3.69	1966	44.40
34	403.0	362.1	0.10	316.3	0.22	504.0	0.25	359.0	285.1	0.21	284.1	0.21	215.1	0.40	219.0	194.6	0.11	203.2	0.07	1895	7.65
35	156.0	65.1	0.58	316.3	1.03	63.9	0.59	141.0	108.8	0.23	284.1	1.01	146.6	0.04	54.2	21.8	0.60	203.2	2.75	(1233)	23.76
36	63.8	156.2	1.45	316.3	3.96	90.2	0.41	73.9	144.2	0.95	284.1	2.84	105.7	0.43	20.3	38.1	0.88	203.2	9.01	(2802)	139.01
37	53.8	72.0	0.34	316.3	4.88	212.0	2.94	62.3	92.3	0.48	284.1	3.56	77.3	0.24	21.2	76.1	2.59	203.2	8.59	928	42.75
38	58.8	23.5	0.60	316.3	4.38	77.3	0.31	65.7	63.5	0.03	284.1	3.32	57.9	0.12	24.5	33.6	0.37	203.2	7.30	345	13.10
39	30.5	37.3	0.22	316.3	9.37	30.0	0.02	49.3	56.3	0.14	284.1	4.76	57.7	0.17	11.5	11.9	0.03	203.2	16.67	(3448)	300.81
40	62.7	39.2	0.37	316.3	4.04	30.9	0.51	67.4	57.3	0.15	284.1	3.22	72.5	0.08	10.2	(2.0)	1.20	203.2	18.93	1532	149.16
41	47.5	49.0	0.03	316.3	5.66	37.2	0.22	50.8	44.7	0.12	284.1	4.59	77.8	0.53	7.0	5.4	0.23	203.2	28.03	(1286)	184.66
42	8.6	47.3	4.50	316.3	35.82	30.2	2.52	25.3	47.4	0.87	284.1	10.23	93.1	2.68	4.4	10.9	1.46	203.2	44.98	(733)	166.76
43	44.1	11.5	0.74	316.3	6.17	43.5	0.01	68.7	24.7	0.64	284.1	3.14	152.2	1.22	20.5	(5.9)	1.29	203.2	8.91	(1644)	81.22
44	92.1	72.6	0.21	316.3	2.43	10.7	0.88	108.0	61.2	0.43	284.1	1.63	264.5	1.45	13.1	17.5	0.34	203.2	14.51	3319	252.33
45	237.0	311.5	0.31	316.3	0.33	132.0	0.44	241.0	256.4	0.06	284.1	0.18	336.8	0.40	138.0	171.4	0.24	203.2	0.47	1022	6.40
46	226.0	246.2	0.09	316.3	0.40	138.0	0.39	225.0	211.8	0.06	284.1	0.26	310.9	0.38	84.0	128.9	0.54	203.2	1.42	245	1.92
47	98.2	52.1	0.47	316.3	2.22	177.0	0.80	107.0	100.7	0.06	284.1	1.66	257.3	1.40	30.1	10.2	0.66	203.2	5.75	(415)	14.78
48	39.0	58.6	0.50	316.3	7.11	123.0	2.15	54.1	94.3	0.74	284.1	4.25	208.7	2.86	12.1	21.0	0.73	203.2	15.80	(2135)	177.46
Max[g]	504.0	495.6	4.50	324.7	35.82	504.0	24.95	455.0	460.8	1.48	290.7	18.45	346.8	5.63	227.0	251.2	6.32	203.9	44.98	3865	300.81
Min_{11}	8.6	0.2	0.02	154.0	0.08	9.3	0.01	14.6	15.8	0.00	147.3	0.00	31.6	0.01	2.3	(12.4)	0.02	82.4	0.07	(3448)	0.96
Mean	123.4	121.6	0.61	313.5	6.32	104.3	1.46	122.0	125.4	0.40	281.7	4.21	167.4	1.27	53.2	57.7	0.85	198.5	11.02	129.3	77.64
Mxd[i]	504.0	495.6	0.02	316.2	0.37	82.5	0.84	455.0	452.9	0.00	284.0	0.38	308.6	0.32	227.0	200.1	0.12	203.2	0.10	9.1	0.96
Mnd[j]	8.6	47.3	4.50	316.3	35.78	30.2	2.51	14.6	17.6	0.20	283.9	18.45	62.9	3.30	2.3	(12.4)	6.32	82.4	34.52	(75)	33.29

[0]signifies that the predicted values are negative.
[a]Observed river discharge.
[b]Wavelet-NAR network hybrid model.
[c]Single NAR network model.
[d]Improved NAR network model.
[e]Predicted river discharge.
[f]Absolute relative error (ARE).
[g]Maximum value.
[h]Minimum value.
[i]Maximum river discharge and related ARE indices.
[j]Minimum river discharge and related ARE indices.

Bayesian regularization optimization algorithm to model a river discharge time series of Weijiabao gage, located on the upper stream of Xianyang gage in Weihe River, China. This previous study revealed that the proposed WNN hybrid approach has a very good performance in training and prediction. In order to compare the results of WNAR with traditional WNN, we employed the WNN approach proposed in Wei *et al.* (2012) to make the modeling and analysis of the three gages in this study. A comparison of the training and prediction performance results of WNAR and WNN are displayed in Table 6. The results suggest that WNAR has much better prediction performance than that of WNN in terms of larger R, smaller RMSE, MAE and MARE though the training results of WNN are also very good.

CONCLUSIONS

A variety of linear and nonlinear methods have been employed to model and predict river discharge time series, such as ARIMA, NAR, ANN, NAR, conceptual model (CM), physical model (PM) and hybrid models. More recently, many studies have proved that wavelet-artificial neural network hybrid modeling approach (WNN) is able to well improve the prediction performance of the models. This study developed a wavelet–nonlinear autoregressive artificial neural network (WNAR) hybrid model approach to improve the model prediction performance of river discharge time series. The WNAR models were successfully applied to three hydrological gages in the Weihe River, China to predict 48-month ahead river discharge time series.

The results from WNAR hybrid model were compared with those from the single NAR network (NARN) models and a traditional WNN hybrid model. Comparison of the results between WNAR and NARN displayed that while NARN model could fit the historical data well provided sufficient neurons in the hidden layers and significant feedback delays in the training process. In general, single NARN models failed to accurately predict the highest and lowest river discharges. In contrast, WNAR hybrid model displayed much better performance both in the training and testing periods than the single NARN models. Comparison of the WNAR hybrid models with the single NARN models, via model evaluation indices, suggested that the WNAR models also significantly improved the general estimation of intermediate and moderate discharges, while comparison of extreme error indices indicated that the WNAR model was also able to accurately forecast the highest and lowest river discharges. A comparison of the training and prediction performance results of WNAR and WNN revealed that WNAR has much better prediction performance than WNN though the WNN has also very good training results. However, since three time series data set in only one river site were used for the simulation and prediction analysis of WNAR artificial neural network, studies on more rivers are required to conclusively prove the advantages of the proposed Wavelet-NAR network modeling approach.

Table 6 | Comparison results of WNAR and WNN performance evaluation

Gage	Mode	Training R[a]	RMSE[b]	MAE[c]	MARE[d]	Forecasting R[a]	RMSE[b]	MAE[c]	MARE[d]
Huaxian	WNAR[e]	0.990	0.038	0.029	0.054	0.912	0.047	0.034	0.042
	WNN[f]	0.969	0.034	0.024	1.005	0.696	0.066	0.045	2.181
Lintong	WNAR[e]	0.990	0.045	0.034	0.079	0.929	0.053	0.040	0.050
	WNN[f]	0.984	0.029	0.021	0.692	0.640	0.088	0.052	1.939
Xianyang	WNAR[e]	0.989	0.047	0.035	0.079	0.919	0.050	0.038	0.045
	WNN[f]	0.987	0.026	0.019	0.670	0.662	0.045	0.032	3.388

[a]Correlation coefficient.
[b]Root mean squared error.
[c]Mean absolute error.
[d]Mean absolute relative error.
[e]Wavelet-Nonlinear Autoregressive Network (WNAR) hybrid model.
[f]Wavelet-BP Artificial Neural Network (WNN) hybrid model.
[g]The improved NAR network model.

ACKNOWLEDGEMENTS

This work is part of the project 'Determination of Environmental Flow Requirement and Its Safeguard Measures in the Wei River in China (2009DFA22980)' and the project 'Effect of Physical and Chemical Actions on the Variation of Streambed Hydraulic Conductivity (51079123)'. The former was supported by the Sino-Swiss Science and Technology Cooperation Program, Switzerland and the Ministry of Science and Technology, China. The latter was supported by National Natural Science Foundation of China.

REFERENCES

Abrahart, R. J. & See, L. 2000 Comparing neural network and autoregressive moving average techniques for the provision of continuous river flow forecasts in two contrasting catchments. *Hydrol. Process.* **14**, 2157–2172.

Adamowski, J. 2008 River flow forecasting using wavelet and cross-wavelet transform models. *J. Hydrol. Process.* **22**, 4877–4891.

Adamowski, J. & Chan, H. F. 2011 A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* **407**, 28–40.

Adamowski, J. & Sun, K. 2010 Development of a wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J. Hydrol.* **390**, 85–91.

Aqil, M., Kita, I., Yano, A. & Nishiyama, S. 2007 Analysis and prediction of flow from local source in a river basin using a Neuro-fuzzy modeling tool. *J. Environ. Manage.* **85**, 215–223.

Atiya, A. F., El-Shoura, S. M., Shaheen, S. I. & El-Sherif, M. S. 1999 A comparison between neural-network forecasting techniques–case study: river flow forecasting. *IEEE T. Neural Network.* **10**, 402–409.

Aussem, A., Campbell, J. & Murtagh, F. 1998 Wavelet-based feature extraction and decomposition strategies for financial forecasting. *J. Comput Intell. Finance* **6**, 5–12.

Bahrami, J., Kavianpour, M., Abdi, M., Telvari, A., Abbaspour, K. & Rouzkhash, B. 2011 A comparison between artificial neural network method and nonlinear regression method to estimate the missing hydrometric data. *J. Hydroinf.* **13**, 245–254.

Beale, M. H., Hagan, M. T. & Demuth, M. H. 2010 *Neural Network Toolbox 7 User's Guide.* MathWorks, Inc., Natick, MA.

Birikundavyi, S., Labib, R., Trung, H. T. & Rousselle, J. 2002 Performance of neural networks in daily streamflow forecasting. *J. Hydrol. Eng.* **7**, 392–398.

Brath, A., Montanari, A. & Toth, E. 2002 Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrol. Earth Syst. Sci.* **6**, 627–639.

Campólo, M., Andreussi, P. & Soldati, A. 1999 River flood forecasting with a neural network model. *Water Resour. Res.* **35**, 1191–1197.

Cannas, B., Fanni, A., See, L. & Sias, G. 2006 Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Phys. Chem. Earth* **31**, 1164–1171.

Chang, F. J. & Chen, Y. C. 2001 A counterpropagation fuzzy-neural network modeling approach to real time streamflow prediction. *J. Hydrol.* **245**, 153–164.

Cigizoglu, H. K. 2003 Incorporation of ARMA models into flow forecasting by artificial neural networks. *Environmetrics* **14**, 417–427.

Cigizoglu, H. K. 2004 Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons. *Adv. Water Resour.* **27** (2), 185–195.

Collischonn, W., Haas, R., Andreolli, I. & Tucci, C. E. M. 2005 Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *J. Hydrol.* **305**, 87–98.

Cui, H. & Song, G. 2007 Study of the wavelet basis selections. In: *Computational Intelligence and Security* (Y. Wang, Y. M. Cheng & H. Liu, eds). Springer, Berlin, Heidelberg, pp. 1009–1017.

Daubechies, I. 1990 The wavelet transform, time-frequency localization and signal analysis. *IEEE T. Inform. Theory* **36**, 961–1005.

Dawson, C. W. & Wilby, R. L. 2001 Hydrological modelling using artificial neural networks. *Prog. Phys. Geog.* **25**, 80–108.

Demirel, M. C., Venancio, A. & Kahya, E. 2009 Flow forecast by SWAT model and ANN in Pracana basin, Portugal. *Adv. Eng. Softw.* **40**, 467–473.

Ertay, T. & Çekyay, B. 2005 Integrated Clustering Modeling with Backpropagation Neural Network for Effcient Customer Relationship Management. In: *Intelligent Data Mining: Techniques and Applications* (D. Ruan, G. Chen, E. E. Kerre & G. Wets, eds). Springer, Berlin, Heidelberg, pp. 355–373.

Feng, L. & Hong, W. 2008 On hydrologic calculation using artificial neural networks. *Appl. Math. Lett.* **21**, 453–458.

Fukushima, Y. 1988 A model of river flow forecasting for a small forested mountain catchment. *Hydrol. Process.* **2**, 167–185.

Garrote, L. & Bras, R. L. 1995 A distributed model for real-time flood forecasting using digital elevation models. *J. Hydrol.* **167**, 279–306.

Goring, D. 2006 *Orthogonal Wavelet Decomposition*. Available via Mul-gor Consulting Ltd, http://www.tideman.co.nz/Salalah/OrthWaveDecomp.html (accessed on 20 March 2011).

Huang, W., Xu, B. & Chan Hilton, A. 2004 Forecasting flows in Apalachicola River using neural networks. *Hydrol. Process.* **18**, 2545–2564.

Jain, A. & Kumar, A. M. 2007 Hybrid neural network models for hydrologic time series forecasting. *Appl. Soft Comput.* **7**, 585–592.

Kim, T. & Valdés, J. B. 2003 Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. *J. Hydrol. Eng.* **8**, 319.

Kişi, Ö. 2005 Daily river flow forecasting using artificial neural networks and auto-regressive models. *Turk. J. Eng. Environ. Sci.* **29**, 9–20.

Kişi, Ö. 2007 Streamflow forecasting using different artificial neural network algorithms. *J. Hydrol. Eng.* **12**, 532–539.

Kişi, Ö. 2008 Stream flow forecasting using neuro – wavelet technique. *Hydrol. Process.* **22** (20), 4142–4152.

Kişi, Ö. 2009 Neural networks and wavelet conjunction model for intermittent streamflow forecasting. *J. Hydrol. Eng.* **8**, 773–782.

Kisi, O. & Cigizoglu, H. K. 2007 Comparison of different ANN techniques in river flow prediction. *Civ. Eng. Environ. Syst.* **24**, 211–231.

Krishna, B., Rao, Y. R. S. & Nayak, P. C. 2011 Time series modeling of river flow using wavelet neural networks. *J. Water Resour. Prot.* **3**, 50–59.

MacKay, D. J. C. 1991 Bayesian interpolation. *Neural Comput.* **4**, 415–447.

Maier, H. R. & Dandy, G. C. 1996 The use of artificial neural networks for the prediction of water quality parameters. *Water Resour. Res.* **32**, 1013–1022.

Mallat, S. G. 1989 Multiresolution approximations and wavelet orthonormal bases of L 2 (R). *T. Am. Math. Soc.* **315**, 69–88.

Mohamed, M. A. & Atta, M. M. 2010 Automated classification of galaxies using transformed domain features. *Int. J. Comput. Sci. Network Security* **10**, 86–91.

Mohammadi, K., Eslami, H. R. & Kahawita, R. 2006 Parameter estimation of an ARMA model for river flow forecasting using goal programming. *J. Hydrol.* **331**, 293–299.

Murtagh, F., Starck, J. L. & Renaud, O. 2004 On neuro-wavelet modeling, *J. Decis. Support Syst.* **37**, 475–484.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I-A discussion of principles. *J. Hydrol.* **10**, 282–290.

Nourani, V., Alami, M. T. & Aminfar, M. H. 2009 A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Eng. Appl. Artif. Intel.* **22**, 466–472.

Partal, T. 2009 River flow forecasting using different artificial neural network algorithms and wavelet transform. *Can. J. Civil Eng.* **36**, 26–38.

Porporato, A. & Ridolfi, L. 2001 Multivariate nonlinear prediction of river flows. *J. Hydrol.* **248**, 109–122.

Pramanik, N., Panda, R. K. & Singh, A. 2010 Daily river flow forecasting using wavelet ANN hybrid models. *J. Hydroinf.* **12**, 1–15.

Sahoo, G. B., Schladow, S. G. & Reuter, J. E. 2009 Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. *J. Hydrol.* **378**, 325–342.

Shamseldin, A. Y. & O'Connor, K. M. 1996 A nearest neighbour linear perturbation model for river flow forecasting. *J. Hydrol.* **179**, 353–375.

Sivakumar, B., Jayawardena, A. W. & Fernando, T. 2002 River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *J. Hydrol.* **265**, 225–245.

Song, J. X., Xu, Z. X., Liu, C. M. & Li, H. E. 2007 Ecological and environmental instream flow requirements for the Wei River – the largest tributary of the Yellow River. *Hydrol. Process.* **21**, 1066–1073.

Spruill, C. A., Workman, S. R. & Taraba, J. L. 2000 Simulation of daily and monthly stream discharge from small watersheds using the SWAT model. *T. Asae* **43**, 1431–1439.

Srinivasulu, S. & Jain, A. 2006 A comparative analysis of training methods for artificial neural network rainfall-runoff models. *Appl. Soft Comput.* **6**, 295–306.

Srinivasulu, S. & Jain, A. 2009 River flow prediction using an integrated approach. *J. Hydrol. Eng.* **14**, 75–75.

Tiwari, M. K. & Chatterjee, C. 2010 Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J. Hydrol.* **394**, 45–470.

Tiwari, M. K. & Chatterjee, C. 2011 A new Wavelet-Bootstrap-ANN hybrid model for daily discharge forecasting, *J. Hydroinf.* **13**, 500–519.

Tokar, A. S. & Markus, M. 2000 Precipitation-runoff modeling using artificial neural networks and conceptual models. *J. Hydrol. Eng.* **5**, 156–161.

Toth, E., Brath, A. & Montanari, A. 2000 Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.* **239**, 132–147.

Tsai, C. P. & Lee, T. L. 1999 Back-propagation neural network in tidal-level forecasting. *J. Waterw. Port. Coastal. Ocean Eng.* **125**, 195–202.

Wang, W. & Ding, J. 2003 Wavelet network model and its application to the prediction of hydrology. *Nature and Science* **1**, 67–71.

Wang, W., Gelder, P. H., Vrijling, J. K. & Ma, J. 2006 Forecasting daily streamflow using hybrid ANN models. *J. Hydrol.* **324**, 383–399.

Wang, W., Kong, J., Duan, L., Wang, Y. & Ma, X. 2004 Research on the conversion relationships between the river and groundwater in the Yellow River drainage area. *Sci. China Series E. Tech. Sci.* (in Chinese) **47**, 25–41.

Wei, S. 2008 *On the use of game theoretic models for water resources management*. Ph.D. Thesis, Brandenburg University of Technology (BTU) Cottbus, Germany.

Wei, S., Song, J. & Khan, N. I. 2012 Simulating and predicting river discharge time series using a wavelet-neural network hybrid modeling approach. *Hydrol. Process.* **26**, 281–296.

Wu, C. L., Chau, K. W. & Li, Y. S. 2009a Methods to improve neural network performance in daily flows prediction. *J. Hydrol.* **372** (1–4), 80–93.

Wu, C. L., Chau, K. & Li, Y. S. 2009b Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resour. Res.* **45**, W08432.

Zhao, H. 2003 Evaluation of water quality and countermeasures of pollution prevention for Weihe river. *Northwest Water Resour. Water Eng.* (in Chinese) **14** (1), 28–31.

First received 20 October 2011; accepted in revised form 5 March 2012. Available online 13 June 2012