Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

Methods to improve neural network performance in daily flows prediction

C. L. Wu, K. W. Chau* and Y.S. Li
Dept. of Civil and Structural Engineering, Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong, People's Republic of China
(*Email: cekwchau@polyu.edu.hk)

ABSTRACT

In this paper, three data-preprocessing techniques, moving average (MA), singular spectrum analysis (SSA), and wavelet multi-resolution analysis (WMRA), were coupled with artificial neural network (ANN) to improve the estimate of daily flows. Six models, including the original ANN model without data preprocessing, were set up and evaluated. Five new models were ANN-MA, ANN-SSA1, ANN-SSA2, ANN-WMRA1, and ANN-WMRA2. The ANN-MA was derived from the raw ANN model combined with the MA. The ANN-SSA1, ANN-SSA2, ANN-WMRA1 and ANN-WMRA2 were generated by using the original ANN model coupled with SSA and WMRA in terms of two different means. Two daily flow series from different watersheds in China (Lushui and Daning) were used in six models for three prediction horizons (i.e. one-, two-, and three-day-ahead forecast). The poor performance on ANN forecast models was mainly due to the existence of the lagged prediction. The ANN-MA, among six models, performed best and eradicated the lag effect. The performances from the ANN-SSA1 and ANN-SSA2 were similar, and the performances from the ANN-WMRA1 and ANN-WMRA2 were also similar. However, the models based on the SSA presented better performance than the models based on the WMRA at all forecast horizons, which meant that the SSA is more effective than the WMRA in improving the ANN performance in the current study. Based on an overall consideration including the model performance and the complexity of modeling, the ANN-MA model was optimal, then the ANN model coupled with SSA, and finally the ANN model coupled with WMRA.

KEYWORDS

Daily flows prediction, artificial neural network, lagged prediction, moving average, singular spectral analysis, wavelet multi-resolution analysis

1. Introduction

Artificial Neural Networks (ANNs) have gained significant attention in past two decades and been widely used for hydrological forecasting. The ASCE Task Committee on Application of Artificial Neural Networks in Hydrology (*2000*) and Dawson and Wilby (*2001*) give good state-of-the-art reviews on ANN modeling in hydrology. Many studies focused on streamflow predictions have proven that ANN is superior to traditional regression techniques and time-series models including Autoregressive (AR) and Autoregressive Moving Average (ARMA) (*Raman and Sunilkumar, 1995; Jain et al., 1999; Thirumalaiah and Deo, 2000; Abrahart and See, 2002; Castellano-Me´ndeza et al., 2004; Kişi, 2003,2005*). Besides, ANN is also compared with nonlinear prediction (NLP) method which is derived from the chaotic time series (*Farmer and Sidorowich, 1987*). Laio et al. (*2003*) carried out a comparison of ANN and NLP for flood predictions and found that ANN performed slightly better at long forecast time while the situation was reversed for shorter time. Sivakumar et al. (*2002*) found that ANN was worse than NLP in short-term river flow prediction.

The ANN is able to capture the dynamics of the flow series by using previously observed flow values as inputs during the forecasting of daily flows from the flow data alone. As a consequence, the high autocorrelation of the flow data often introduce the lagged predictions for the ANN model. The issue of lagged predictions in the ANN model has been mentioned by some researchers (*Dawson and Wilby, 1999; Jian and Srinivasulu, 2004; de Vos and Rientjes, 2005; Muttil and Chau, 2006*). De Vos and Rientjes (*2005*) suggested that an effective solution to the forecasting lag effect is to obtain new model inputs by moving average (MA) over the original discharge data.

As known, a natural flow series can be viewed as a quasi-periodic signal, which is contaminated by various noises at different flow levels. Cleaner signals used as model inputs will improve the model performance. Therefore, signal decomposition techniques for the purpose of data-preprocessing may be favorable. Two such techniques are known as singular spectral analysis (SSA) and wavelet multi-resolution analysis (WMRA). Briefly, the SSA decomposes a time series into a number of components with simpler structures, such as a slowly varying trend, oscillations and noise. The SSA uses the basis functions characterized by data-adaptive nature, which makes the approach suitable for the analysis of some nonlinear dynamics (*Elsner and Tsonis, 1997*). A time series in the WMRA breaks down into a series of linearly independent detail signals and one approximation signal by using discrete wavelet transform with a specific wavelet function such as the Haar wavelet. Mallat (*1989*) presented a complete theory for wavelet multi-resolution signal decomposition (also mentioned as pyramid decomposition algorithm). Moreover, the continuous wavelet transform can conduct a local signal analysis at which point the traditional Fourier and SSA are, however, less effective (*Howell and Mahrt, 1994*), which can be referred to Torrence and Compo (*1998*) for a practical guide. Nevertheless, the signal analysis in the time-frequency space is not the point of concern in this study.

The techniques of SSA and WMRA have been successfully introduced to the field of hydrology (*Lisi et al., 1995*; *Sivapragasam et al., 2001; Marques et al., 2006; Partal and Kişi, 2007*). Sivapragasam et al. (*2001*) established a hybrid model of support vector machine (SVM) in conjunction with the SSA for the forecasting of rainfall and runoff. A considerable improvement in the model performance was obtained in comparison with the original SVM model. However, the paper did not explicitly mention how to combine SVM with the SSA. The applications of WMRA to precipitation and discharge were presented in the work of Partal and Kişi (*2007*) and Partal and Cigizoglu (*2008*) respectively, where the WMRA was applied to each model input variable. Results from their studies indicated that the WMRA is highly promising for improvement of the model performance.

The objective of this study is to evaluate the effectiveness of the three data-preprocessing techniques of MA, SSA, and WMRA in the improvement of the ANN model performance. To explore the SSA or WMRA, the ANN model is coupled with the components of SSA or WMRA in terms of two different methods. One is that the raw flow data is first decomposed, then a new flow series is obtained by components filter, and finally the new flow series is used to generate the model inputs. The type of model is named as ANN-SSA1 or ANN-WMRA1. The other method is the same as Partal and Cigizoglu (*2008*), based on which the model is named as ANN-SSA2 or ANN-WMRA2. With the original ANN model and the ANN-MA, there are six models for the flow data forecasting in all. This paper is organized in the following manner. Section 2 presents the two sets of streamflow data. Section 3 describes the modeling methods including a brief introduction of ANN, MA, SSA, and WMRA, and how to construct the hybrid ANN models. The application of the forecast models to the flow data is presented in Section 4 where relevant points include decomposition of flows, the identification of the ANN's architecture, implementation of

ANN models, and forecasting results and discussion. Section 5 sheds light on main conclusions in this study.

2. Streamflow Data

Daily mean flow data from two rivers of Lushui and Daning are used in this study. The two rivers belong to direct tributaries of Yangtze River, and are both located in Hubei province, People Republic of China. The flow data from Lushui River were acquired at Tongcheng hydrology station which is at the upper stream of Lushui watershed (hereafter, the flow data is referred to as Lushui series). The watershed has an area of 224 km^2. The data period covers a 5 years long duration (Jan. 1, 1984- Dec. 31, 1988). The flow data from Daning River were collected at Wuxi hydrology station which is at the upper and middle streams of Daning watershed. The drainage area controlled by Wuxi station is 2 001km^2. The flow data spanned 20 years (from Jan. 1, 1988 to Dec. 31, 2007).

In the process of modeling of ANNs, the raw flow data is often partitioned into three parts as training set, cross-validation set and testing set. The training set serves the model training and the testing set is used to evaluate the performances of models. The cross-validation set help to implement an early stopping approach in order to avoid overfitting of the training data. The same data partition was adopted in two daily flow series: the first half of the entire flow data as training set and the first half of the remaining data as cross-validation set and the other half as testing set.

Table 1 presents related information about two rivers and some descriptive statistics of the original data and three data subsets, including mean (μ), standard deviation (S_x), coefficient of variation (C_v), skewness coefficient (C_s), minimum (X_{min}), and maximum (X_{max}). As shown in Table 1, the training set cannot fully include the cross-validation or testing set. Due to the weak extrapolation ability of ANN, it is suggested that all data should be scaled to the interval [-0.9, 0.9] rather than [-1, 1] when ANN employs the hyperbolic tangent functions as transfer functions in the hidden layer and output layer.

Fig. 1 estimates the autocorrelation functions (ACF), average mutual information (AMI), and partial autocorrelation functions (PACF) from lag 0 to lag 30 days for the two flow series. The AMI measures the general dependence of two variables (*Fraser and Swinney, 1986*) whereas the ACF and PACF show the dependence from the perspective of linearity. The first order autocorrelation of each flow data is large (0.59 for Lushui, and 0.7 for Wuxi). The rapid decaying pattern of the PACF confirms the dominance of autoregressive process, relative to the moving-averaging process revealed by the ACF.

3. Methods

3.1. Artificial neural networks

An ANN is a massively parallel-distributed information processing system with highly flexible configuration and so has an excellent nonlinearity capturing ability. The feed-forward multilayer perceptron (MLP) among many ANN paradigms is by far the most popular, which usually uses the technique of error back propagation to train the network configuration. The architecture of the ANN consists of the number of hidden layers and the number of neurons in input layer, hidden layers and output layer. ANNs with one hidden layer are commonly used in hydrologic modeling (*Dawson and Wilby, 2001; de Vos and Rientjes, 2005*) since these networks are considered to provide enough complexity to accurately simulate the nonlinear-properties of the hydrologic process. A three-layer ANN is therefore chosen for the present study, which comprises the input layer with I nodes, the hidden layer with H nodes (neurons), and the output layer with one node. The hyperbolic tangent functions are used as transfer functions in the hidden layer and output layer. The purpose of network training is to optimize the weights w connecting neighboring layers and bias θ of each neuron in hidden layer and output layer. The Levernberg-Marquardt (LM) training algorithm is used here for adjusting the weights and bias.

3.2. Moving Average

The moving average method smoothes data by replacing each data point with the average of the K neighboring data points, where K may be called the length of memory window. The method is based on the idea that any large irregular component at any point in time will exert a smaller effect if we average the point with its immediate neighbors (*Newbold et al., 2003*). The most common moving average method is the unweighted moving average, in which each value of the data carries the same weight in the smoothing process. For time series $\{x_1, x_2, \cdots, x_N\}$, the K-term unweighted moving average is written as $x_t^* = (\sum_{i=0}^{K-1} x_{t-i})/K$ (where $t = K, \ldots N; x_t^*$ stands for the moving average value) when the backward moving mode is adopted (*Lee et al., 2000*). Choice of the window length K is by a trial and error procedure of minimizing the ANN prediction error.

3.3. Singular Spectrum Analysis

The implementation of SSA can be referred to Vautard *et al.* (*1992*) and Elsner and Tsonis (*1997*). Four steps are summarized for the implementation. The first step is to construct the 'trajectory matrix'. The 'trajectory matrix' results from the method of delays. In the method of delays, the coordinates of the phase space will approximate the dynamic of the system by using lagged copies of the time series. Therefore, the 'trajectory matrix' can reflect the evolution of the time series with a careful choice of (τ, L) window. For time series $\{x_1, x_2, \cdots, x_N\}$, the 'trajectory matrix' is denoted by

$$\mathbf{X} = \frac{1}{\sqrt{N}} \begin{pmatrix} x_1 & x_{1+\tau} & x_{1+2\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & x_{2+2\tau} & \cdots & x_{2+(L-1)\tau} \\ x_3 & x_{3+\tau} & x_{3+2\tau} & \cdots & x_{3+(L-1)\tau} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-(L-1)\tau} & x_{N-(L-2)\tau} & x_{N-(L-3)\tau} & \cdots & x_N \end{pmatrix} \quad (1)$$

where L is the embedding dimension (also called singular number in the context of SSA), τ is the lagged (or delay) time. The matrix dimension is $R \times L$ where $R = N-(L-1)\tau$. The next step is the singular value decomposition (SVD) of the trajectory matrix $\mathbf{X}$. Let $\mathbf{S} = \mathbf{X}^{\mathrm{T}}\mathbf{X}$ (called the lagged-covariance matrix). With SVD, $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{DLE}^{\mathrm{T}}$ where $\mathbf{D}$ and $\mathbf{E}$ are left and right singular vectors of $\mathbf{X}$, and $\mathbf{L}$ is a diagonal matrix of singular values. $\mathbf{E}$ consists of orthonormal columns, and is also called the 'empirical orthonormal functions' (EOFs). Substituting $\mathbf{X}$ into the definition of $\mathbf{S}$ yields the formula of $\mathbf{S} = \mathbf{EL}^2\mathbf{E}^{\mathrm{T}}$. Further $\mathbf{S}$=$\mathbf{E} \wedge \mathbf{E}^{\mathrm{T}}$ since $\mathbf{L}^2 = \wedge$ where $\wedge$ is a diagonal matrix consisting of ordered values $0 \le \lambda_1 \le \lambda_2 \le \cdots \lambda_L$. Therefore, the right singular vectors of $\mathbf{X}$ are the eigenvectors of $\mathbf{S}$ (*Elsner and Tsonis, 1997*). In other words, the singular vectors $\mathbf{E}$ and singular values of $\mathbf{X}$ can be respectively attained by calculating the eigenvectors and the square roots of the eigenvalues of $\mathbf{S}$.

The first two steps involve the decomposition stage of SSA, and the next two steps belong to the recovering stage. The third step is to calculate the principal components (a_i^k's) by projecting the original time record onto the eigenvectors as follows:

$$a_i^k = \sum_{j=1}^{L} x_{i+(j-1)\tau} e_j^k \text{ , for } \quad i = 1, 2 \cdots, N-(L-1)\tau \tag{2}$$

where e_j^k represents the jth component of the kth eigenvector. As known, each principal component is a filtered process of the original series with length $N-(L-1)\tau$, not length N as desired, which poses a problem in real-time prediction.

The last step is to generate reconstruction components (RCs) whose lengths are the same as the original series. The generation of each RC depends on a convolution of one principal component with the corresponding singular vector, given by Vautard et al. (*1992*). Therefore, The m RCs can be achieved if all m principal components and their associated eigenvectors are employed in the process of signal reconstruction. Certainly, the original record can be filtered by choosing p ($< L$) RCs from all L RCs.

3.4. Wavelet Multiresolution Analysis (WMRA)

The WMRA utilizes discrete wavelet transform (DWT) to decompose a raw signal into a series of component signals. Referred to the work of Daubechie (*1992*) and Kücük and AǵiralÍoǵlu (*2006*), the DWT is briefly presented below.

(1) *Wavelet transform*

Let $f(t)$ be a continuous time series with $t \in [-\infty, \infty]$, the continuous wavelet transform of $f(t)$ with respect to a wavelet function $\psi(t)$ is defined by the linear integral operator

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}^{*}(t)dt \text{ with } \psi_{a,b}^{*}(t) = \frac{1}{\sqrt{|a|}}\psi^{*}\left(\frac{t-b}{a}\right) \tag{3}$$

where $W(a,b)$ is the wavelet coefficients and a and b are real numbers; the ($*$) indicates complex conjugation. Thus, the wavelet transform is a function of two variables, a and b. The parameter "a" can be interpreted as a dilation (a > 1) or contraction (a < 1) factor of the wavelet function $\psi(t)$ corresponding to different scales of observation. The parameter "b" can be interpreted as a temporal translation or shift of the function $\psi(t)$, which allows the study of the signal $f(t)$ locally around the time b. The wavelet transform therefore expresses a time series in three-dimensional space: time (b), scale/frequency (a), and wavelt spectrum $|W(a,b)|^2$.

The DWT is to calculate the wavelet coefficients on discrete dyadic scales and positions in time. Discrete wavelet functions have the form by choosing $a = a_0^m$ and $b = nb_0 a_0^m$ in Eq. (3) as:

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

$$\psi_{m,n}(t) = a_0^{-m/2}\psi\left(\frac{t - nb_0 a_0^m}{a_0^m}\right) = a_0^{-m/2}\psi(a_0^{-m}x - nb_0) \tag{4}$$

where m and n are integers that control the wavelet dilation and shift respectively, and $a_0 > 1$, $b_0 > 0$ are fixed. The appropriate choices for a_0 and b_0 depend on the wavelet function. A common choice for them is $a_0 = 2, b_0 = 1$. Now Assuming a discrete time series x_t, where x_i occurs at the discrete time i, the DWT becomes

$$W_{m,n} = 2^{-m/2}\sum_{i=0}^{N-1} x_i \psi(2^{-m} i - n) \tag{5}$$

where $W_{m,n}$ is the wavelet coefficient for the discrete wavelet function with scale $a = 2^m$ and location $b = 2^m n$. In this study, the wavelet function is derived from the family of Daubechies wavelets with the 3 order.

(2) *Multiresolution analysis (MRA)*

The Mallat's decomposition algorithm (*Mallat, 1989*) is employed in this study. According to the Mallat's theory, the original discrete time series x_t is decomposed into a series of linearly independent approximation and detail signals.

The process consists of a number of successive filtering steps as depicted in Fig. 2. Fig. 2(a) displays an entire MRA scheme, and Fig. 2(b) shows the filtering operation between two adjacent resolutions. The original signal x_t is first decomposed into an approximation and accompanying detail. The decomposition process is then iterated, with successive approximation being decomposed in turn so that the finest-resolution original signal is transformed into many coarser-resolution components (*Kücüķ and Aġalİoġlu, 2006*). As shown in Fig. 2(b), the approximation cA_{i+1} is achieved by letting cA_i pass through the low-pass filter H' and downsampling by two (denoted as $\downarrow 2$) whereas the detailed version cD_{i+1} is obtained by letting cA_i pass through the high-pass filter G' and downsampling by two. The details are therefore the low-scale, high frequency components whereas the approximations are the high-scale, low-frequency components. Finally, the original signal x_t is decomposed into many detailed components and one approximation component which denotes the coarsest resolution. Following the procedure, the raw flow data can be decomposed into $m+1$ components if the m in DWA is set.

3.5. ANNs integrated with data preprocessing techniques

To explore the capability of ANNs, five ANN models are generated with the aid of the above three data-processing techniques. These data-preprocessing techniques are aimed at improving mapping relationship between inputs and output of the ANN model by smoothing raw flow data. Six forecasting models are described as follows.

(1) ANN

The original ANN model (hereafter referred to as ANN) directly employs original flow data to generate model input/output pairs. It is used as the baseline model for the purpose of comparison with the other five proposed models.

(2) ANN-MA

The moving average method first smoothes the original flow data, and then the smoothed data are used to form the model inputs. The model is hereafter referred to as ANN-MA.

(3) ANN-SSA1 and ANN-SSA2

The raw flow data is first decomposed by SSA into L RCs, and then the raw flow data is filtered by selecting $p\,(\le L)$ from all L RCs. A new flow series is generated by summing the selected p RCs. Finally, the new flow series is used to generate the model inputs. The type of model is hereafter referred to as ANN-SSA1.

Different from ANN-SSA1, the model inputs are first derived from the original flow data, and then each input variable series of the model is filtered by selecting $p\,(< L)$ from all L RCs. A new series for each input variable is formed by summing the chosen p RCs. The type of model is hereafter called ANN-SSA2. Obviously, the p may be different for each input variable of ANN-SSA2 whereas the p is the same for each input variable in ANN-SSA1.

(4) ANN-WMRA1 and ANN-WMRA2

The ANN-WMRA1 and ANN-WMRA2 are established in combination with the WRMA instead of SSA. The idea behind the modelling is identical to the ANN-SSA1 and ANN-SSA2.

3.6. Evaluation of model performances

The Person's correlation coefficient (r) or the coefficient of determination ($R^2=r^2$), have been identified as inappropriate measures in hydrologic model evaluation by Legates and McCabe (*1999*). The coefficient of efficiency (*CE*) (*Nash and Sutcliffe, 1970*) is a good alternative to r or R^2 as a 'goodness-of-fit' or relative error measure in that it is sensitive to differences in the observed and forecasted means and variances. Legates and McCabe (*1999*) also suggested that a complete assessment of model performance should include at least one absolute error measure (e.g., RMSE) as necessary supplement to a relative error measure. Besides, the Persistence Index (PI) (*Kitanidis And Bras, 1980*) was adopted here for the purpose of checking the prediction lag effect. Three measures were therefore used in this study. They are formulated as: $\mathrm{CE} = 1 - \sum_{i=1}^{n}(\mathrm{T}_i - \hat{\mathrm{T}}_i)^2 / \sum_{i=1}^{n}(\mathrm{T}_i - \bar{\mathrm{T}})^2$, $\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\mathrm{T}_i - \hat{\mathrm{T}}_i)^2}$, and $\mathrm{PI} = 1 - \sum_{i=1}^{n}(\mathrm{T}_i - \hat{\mathrm{T}}_i)^2 / \sum_{i=1}^{n}(\mathrm{T}_i - \mathrm{T}_{i-l})^2$. In these equations, n is the number of observations, $\hat{\mathrm{T}}_i$ stands for forecasted flow, T_i represents observed flow, $\bar{\mathrm{T}}$ denotes average observed flow, and T_{i-l} is the flow estimate from a so-call persistence model (or called naïve model) that basically takes the last flow observation (at time i minus the lead time l) as a prediction. CE and PI values of 1 stands for perfect fits.

4. Application of models to the flow data

4.1 Decomposition of daily flow data

(1) *Decomposition by SSA*

Decomposition of the raw flow data by SSA requires identifying the parameter pair (τ, L). The choice of L represents a compromise between information content and statistical confidence. The value of L should be able to clearly resolve different oscillations hidden in the original signal. In other words, some leading eigenvalues should be identified. Fig. 3 displays the sensitivities of the eigenvalue decomposition to the singular number L for Lushui and Wuxi. Results show that about 5 leading eigenvalues stand out for different L, which implies the leading eigenvalues are insensitive to L. These leading eigenvalues are associated with lower frequency oscillations. For the convenience of filtering operation later, L is set a small value of 5 in the present study. Fig. 4 presents the sensitivities of the eigenvalue decomposition to the lag time τ when $L = 5$. Results suggest that the eigenvalues can be distinguished when $\tau = 1$, which means that original signal can be resolved distinctly. The final parameter pair (τ, L) in SSA were therefore set as (1, 5) for two studied flow data series.

Taking the flow data of Lushui as the example, Fig. 5 presents five RCs and the original flow series excluding the testing data. The RC1 represents an obvious low-frequency oscillation, which exhibits a similar mode to the original flow series. The other RCs reflect high-frequency oscillations, part of which can be deleted so as to improve the mapping between inputs and output of ANN models. Fig. 6 depicts AMI and cross-correlation function (CCF) between RCs and the original flow data. The last plot in Fig. 6 denotes the average of AMI and CCF, which was generated by averaging the results in the plots of five RCs. The average indicates an overall correlation either being positive or negative. The best positive correlation occurs at lag 1. The RC1 among all 5 RCs exhibits the best positive correlation with the original flow series. The correlation quickly shifts from positive value to negative value for other RCs with the increase of lag time. In essence, the positive or negative value of CCF may indicate that the RC makes a positive or negative contribution to the output of model when the RC is used as the input of model. Therefore, deleting RCs, which are negative correlations with the model output if the average AMI or CCF is positive, can improve the performance of the forecasting model. This is the underlying reason that ANN is coupled with SSA or WRMA in this study.

(2) *Decomposition by WMRA*

The WMRA decomposes an original signal into many components at different scales (or frequencies). Each of components plays a distinct role in the original flow series. The low-frequency component generally reflects the identity (periodicity and trend) of the signal whereas the high-frequency component uncovers details (*Kücük and AġalÌoġlu, 2006*). An important issue in the WMRA is to choose the appropriate number of scales. The largest scales should be shorter than the size of testing data. The sizes of testing data are 550 days (1.25 years) for Lushui and 1826 days (5 years) for Wuxi. The largest scale *m* is therefore chosen as 8 and 10 for Lushui and Wuxi respectively. Thus, the flow data of Lushui was decomposed at 8 wavelet resolution levels (2^1-2^2-2^3-2^4-2^5-2^6-2^7-2^8 day), and the flow data of Wuxi was decomposed at 10 wavelet resolution levels (2^1-2^2-2^3-2^4-2^5-2^6-2^7-2^8-2^9-2^{10}day). Fig. 7 shows the original flow data of Lushui (excluding testing data) and 9 wavelet components (8 details components and 1 approximation component). For the purpose of distinction with the components of SSA, one wavelet component at some scale is expressed by DWC with the power of 2. For instance, DWC1 stands for the component at the scale of 2^1 day and DWC2 represents the component at the scale of 2^2 day whereas DWC9 denotes the approximation for the Lushui flow series. The approximation component at the right end of Fig. 7 is the residual which reflects the trend of the flow data. As revealed in Fig. 7, detail components at scales of 2^8 (256 day) and 2^7 (128 day) are characterized by notable periodicity, which partially exhibits annual oscillation and semi-annual oscillation in the original flow series. Relative weak periodic signals occur at scales of 16 days, 32 days and 64 days. Other high-frequency components tend to capture the details (or noises) of the original flow series. Hence, the inputs of model can be filtered by deleting some high-frequency DWCs.

AMI and CCF between DWCs and the original flow are presented in Fig. 8. The last plot in Fig. 8 describes the average plots of AMI and CCF, which was generated by averaging the plots of 9 DWCs. The DWCs with lower frequencies including 2^8, 2^9 and 2^{10} days always keep positive correlations with the original flow data within a long lag time. The approximation component DWC9 also exhibits a positive correlation with the original flow data with a long lag time. It can be seen from other plots of DWCs that the correlation coefficient shifts between positive and negative values.

4.2 Identification of the ANN architecture

Six models' architectures need to be identified depending on the raw or filtered flow data before models can be applied to the flow prediction. The ANN model is used as a paradigm to shed light on the procedure.

The architecture identification of the ANN model includes determining model inputs and the number of nodes (or neurons) in the hidden layer when there is one model output. The selection of appropriate model inputs is crucial in model development. There is no any theoretic guide for the selection of model inputs although a large number of methods have been reported in literature which was reviewed by Bowden et al. (*2005*). These methods appear very subjective. Sudheer et al. (*2002*) suggested that the statistical approach depending on cross-, auto- and partial-auto-correlation of the observed data is a good alternative to the trial-and-error method in identifying model inputs. The statistical method was also successfully applied to daily suspended sediment data by Kişi (*2008*). The model input in this method is mainly determined by the plot of PACF. The essence of this method is to examine the dependence between the input and output data series. According to this method, the model inputs were originally considered to take previous 6 daily flows for Luishui and previous 13 daily flows for Wuxi because the PACF within the confidence band occurs at lag 6 for Luishui and lag 13 for Wuxi (Fig. 1).

The false nearest neighbors (FNNs) (*Kennel et al., 1992; Abarbanel et al., 1993*) is another commonly used method to identify model inputs, which is for the perspective of dynamics reconstruction of a system (*Wang et al., 2006*). The following discussion outlines the basic concepts of the FNN algorithm. Suppose the point $\mathbf{Y}_{\mathrm{i}}=\{x_i,x_{i+\tau},x_{i+2\tau},\cdots,x_{i+(L-1)\tau}\}$ has a neighbor $\mathbf{Y}_{\mathrm{j}}=\{x_j,x_{j+\tau},x_{j+2\tau},\cdots,x_{j+(L-1)\tau}\}$, the criterion that $\mathbf{Y}_{\mathrm{j}}$ is viewed as a false neighbor of $\mathbf{Y}_{\mathrm{i}}$ is:

$$\frac{\left|x_{i+L\tau}-x_{j+L\tau}\right|}{\left\|\mathbf{Y}_{\mathrm{i}}-\mathbf{Y}_{\mathrm{j}}\right\|}>\mathrm{R}_{\mathrm{tol}} \tag{6}$$

where $\|\ \|$ stands for the distance in a Euclidean sense, $\mathrm{R}_{\mathrm{tol}}$ is some threshold with the common range of 10 to 30. For all points i in the vector state space, Eq. (6) is performed and then the percentage of points which have FNNs is calculated. The algorithm is repeated for increasing L until the percentage of FNNs drops to zero (or some acceptable small number, denoted by R_{P}, such as R_{P}=1%), where L is the target L. Setting $\mathrm{R}_{\mathrm{tol}}=15$ and $\tau=1$, the

Percentage of FNNs (FNNP) as a function of L were calculated for the two flow series, shown in Fig. 9. The values of L are 6 and 8 respectively for Lushui and Wuxi when R_P=2%, and the values of L are 7 and 12 for Lushui and Wuxi when R_P=1%. The final selection of model inputs were 6 for Luishui (i.e. using $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}$ and Q_{t-6} as input to predict Q_t) and 8 (i.e. using $Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}, Q_{t-5}, Q_{t-6}, Q_{t-7}$ and Q_{t-8} as input to predict Q_t) for Wuxi by trial and error among three potential model inputs.

The ensuing task is to optimize the size of the hidden layer with the chosen three inputs and one output. The optimal size H of the hidden layer was found by systematically increasing the number of hidden neurons from 1 to 10 until the network performance on the cross-validation set was no longer improved significantly. The identified ANN architecture was: 6-8-1for Lushui and 8-9-1 for Wuxi. Note that the identified ANN model was used as the baseline model for the following hybrid operation with data-preprocessing techniques.

4.3 Implementation of models

(1) *ANN-MA*

The window length K (see Section 3.2) can be determined by varying K from 1 to 10 depending on the identified ANN model. The targeted value of K was associated with the optimal network performance in terms of RMSE. The final K was 3, 5, and 7 at one-, two-, and three-day-ahead forecast horizons for the two flow data.

(2) *ANN-SSA1 (or ANN-WMRA1)*

According to the methodological procedure for the ANN-SSA1 and ANN-WMRA1, the further tasks include sorting out contributing components from RCs or DWCs and determining the ANN-SSA1 architecture. RCs from Lushui were employed to describe the implementation of the ANN-SSA1.

The determination of the effective RCs depends on the correlation coefficients between RCs and the original flow data (Fig. 6). The procedure includes the following steps:

- Identify that the average of CCF (shown at the right below corner of Fig. 6) is positive or negative. For one-day-ahead prediction, the average of CCF is positive (0.18) at lag 1.
- Sort the value of CCFs at lag 1 for all RCs in a descending order (in an ascending order, if the average of CCF is negative). For the one-day-ahead prediction, the new order is RC1, RC2, RC3, RC4, and RC5, which is the same as the original order.
- Use the ANN model to conduct predictions in which $p\ (\le L)$ RCs generating the new model inputs systematically decreases from all 5 RCs at the beginning to only RC1 at the end. The target value of p is associated with the minimum RMSE amongst five runs of the ANN model.

Fig. 10 shows the results of the RCs and DWCs filter at all three prediction horizons. It can be seen from Fig. 10(1) that two components (RC1and RC2) were remained for one-day-ahead prediction, two components (RC1 and RC5, because the new order is RC1-RC5-RC2-RC4-RC3) for two-day-ahead prediction, and only component (RC1, due to the new order being RC1-RC4-RC5-RC3-RC2) for three-day-ahead prediction. Fig. 10(2) shows that most of DWCs are kept. For instance, only DWC1 (detail at the scale of 2^1 day) of all 9 DWCs was deleted for one-step-ahead prediction, two DWCs (DWC1 and DWC2) were deleted for two-step-ahead prediction, and three DWCs (DWC1, DWC2, and DWC3) were removed for three-step-ahead prediction. The values of vali-RMSE in Fig. 10 also show that the SSA is superior to the WMRA in the improvement of the ANN performance.

Based on the remained RCs or DWCs, the number of nodes in the hidden layer of the ANN model is optimized again. The identified architectures of ANN-SSA1 and ANN-WMRA1 were the same as the original ANN model (i.e., 6-8-1 for Lushui and 8-9-1 for Wuxi).

(3) *ANN-SSA2 (or ANN-WMRA2)*

Implementation of the ANN-SSA2 or ANN-WMRA2 can be referred to Partal and Kişi (*2007*) and Partal and Cigizoglu (*2008*) for details. A three-step procedure of the implementation is: to firstly use the SSA or WRMA to decompose each input variable series of the original ANN model; to then select effective components for each input variable; to finally generate a new input variable series by summing selected effective components. Obviously, the procedure is very time-consuming because it has to be repeated for each ANN input variable. There is no definite criterion for the selection of RCs or DWCs. A basic principle is to remain these components that make a positive contribution to the model output. A trial and error approach was therefore employed in the present study. The value of CCF in Figs. 6 and 8 indicate the contribution of each component in each input variable to the ANN output. For instance, the values of CCF at lag 1 denote the correlation coefficients between components of Q_{t-1} and the output variable Q_t. Table 2 lists the effective components of each input variable for ANN-SSA2 and ANN-WMRA2 based on the Lushui flows. It can be seen that the SSA is more effective than the WMRA because most of DWCs are remained for each input variable of the ANN-WMRA2 whereas only one or two RCs are kept for each input variable of the ANN-SSA2. With new model inputs, identified architectures of the ANN-SSA2 and ANN-WMRA2 were also the same as the original ANN model (i.e., 6-8-1 for Lushui and 8-9-1 for Wuxi).

4.4 forecasting results and discussion

Fig. 11 shows the scatter plots and hydrographs of the results of one-day-ahead prediction of the ANN model using the flow data of Lushui and Wuxi. The ANN model seriously underestimates a number of moderate and high magnitudes of the flows. The low values of CE and PI demonstrate that a time lag may exist between the forecasted and observed flows. A representative detail of the hydrographs is presented in Figs. 12(1) and 12(2), in which the prediction lag effect is fairly obvious. Figs. 12(3) and 12(4) illustrate the lag values at one-, two-, and three-day-ahead forecast horizons for Lushui and Wuxi on a basis of the CCF between forecasted and observed discharges. The value of CCF at zero lag corresponds to the actual performance (i.e. correlation coefficient) of the modes. The lag at which the value of CCF is maximized, is an expression for the mean lag in the model forecast. Therefore, there were 1, 2, and 4 days lag for Lushui, and 1, 2, and 3 days lag for Wuxi, which are respectively associated with one-, two-, and three-day-ahead forecasting.

The scatter plots of simulation results of one-day-ahead prediction based on the Lushui flows by using the ANN-SSA1, ANN-SSA2, ANN-WMRA1, and ANN-WMRA2, are presented in Fig. 13. Each of the four models exhibits a noticeable improvement in the performance compared with the ANN model. The remarkable improvement is, however, from the ANN-SSA1 and ANN-SSA2 in terms of RMSE, CE and PI. Fig. 14 describes the representative detail of the hydrographs of the four prediction models. The lagged predictions can be clearly found in the detail plots derived from the ANN-WMRA1 and

ANN-WMRA1, in particular in the latter. Figs. 15 and 16 present the scatter plots and detail parts of the hydrographs from the same four models based on the flows of Wuxi. A great improvement in the model performance can be seen when the four models are compared with the ANN model. In terms of RMSE, CE and PI, the ANN-SSA1 and ANN-SSA2 performed better than the ANN-WMRA1 and ANN-WMRA2. The detail plots in Fig. 16(1) and16 (3) also indicate that the ANN-SSA1 and ANN-SSA2 can reasonably approximate the flows of Wuxi. In contrast, the ANN-WRMA1 and ANN-WRMA1 underestimate quite a number of peak value flows. Furthermore, the lag effect is still visible in Fig. 16 (2) and 16(4).

In terms of the ANN-SSA1 and ANN-SSA2, the scatter plots with low spread, the low RMSE, and high CE and PI indicate excellent model performance. The matched-perfectly detail plots in Figs. 16(1) and 16(3) also show that the two models highly approximate the flows of Wuxi.

The ANN-MA simulation results of one-day-ahead prediction are presented in Fig. 17. Figs. 17(1), 17(3), and 17(5) depict the scatter plots, the hydrographs and the CCF curves at three forecasting horizons based on the flows of Lushui. Figs. 17(2), 17(4), and 17(6) demonstrate the scatter plots, the hydrographs and the CCF curves at three forecasting horizons depending on the flows of Wuxi. The results from Figs. 17(5) and 17(6) show that the issue of lagged prediction is completely eliminated by the MA because the maximum CCF occurs at zero lag. Compared with the other five models (ANN, ANN-SSA1, ANN-SSA2, ANN-WMRA1, and ANN-WMRA2), the ANN-MA model exhibits the best model performance including the scatter plots with low spread, the low RMSE, and high CE and PI. The matched-perfectly detail plots in Fig. 17 (2) and 17(4) indicate that the two models are fairly adequate in reproducing the observed flows of Lushui and Wuxi. In addition, the ANN-MA model also shows a great ability in capturing the peak value of flows (depicted in Figs. 17 (1), 17(2), 17(3), and 17(4)).

Table 3 and 4 summarize the forecasting performance of all six models in terms of RMSE, CE, and PI at three prediction horizons. The ANN model shows markedly inferior results compared with the other five models. The ANN-MA among all models holds the best performance at each prediction horizon. It can be also seen that the performances from the ANN-SSA1 and ANN-SSA2 are similar, and the same situation appears between the ANN-WMRA1 and ANN-WMRA2. However, the models based on SSA provide noticeably better performance than the models based on WMRA at each forecast horizon, which means that the SSA is more effective than the WMRA in improving the ANN performance in the current study.

5. Conclusions

In this study, the conventional ANN model was coupled with three different data-preprocessing techniques, i.e., MA, SSA, and WMRA. As a result, six ANN models, the original ANN model, ANN-MA, ANN-SSA1, ANN-SSA2, ANN-WMRA1, and ANN-WMRA2, were proposed to forecast two daily flow series of Lushui and Wuxi. To apply these models to the flow data, the memory length K of MA, the lag time and embedding dimension (τ, L) of SSA, and the largest scale m of WMRA needed to be decided in advance. The K for one-, two-, and three-day-ahead were 3, 5, and 7 days for each flow data by trial and error. The values of (τ, L) were set as the value of (1, 5) for each flow data by sensitivity analysis. The largest scale m of WMRA was 8 and 10 for Lushui and Wuxi respectively depending on the length size of the testing data.

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

The results from the original ANN model were disappointing due to the existence of the prediction lag effect. The analysis of CCF between predicted and observed flows revealed that the lags at three prediction horizons were 1, 2, and 4 days lag for Lushui and 1, 2, and 3 days lag for Wuxi. All three data-preprocessing techniques could improve the ANN performance. The ANN-MA, among all six models, performed best and eradicated the lag effect. It could be also seen that the performances from the ANN-SSA1 and ANN-SSA2 were similar, and the same situation appeared between the ANN-WMRA1 and ANN-WMRA2. However, the models based on SSA provided noticeably better performance than the models based on WMRA at each forecast horizon, which meant that the SSA was more effective than the WMRA in improving the ANN performance in the current study.

Under the overall consideration including the model performance and the complexity of modeling, the ANN-MA model was optimal, then the ANN model coupled with SSA, and finally the ANN model coupled with WMRA.

References

Abarbanel, H. D. I., Brown,R., Sidorowich, J. J., and Tsimring, L.S., (1993). The analysis of observed chaotic data in physical systems, ***Reviews of Modern Physics***, 65(4), 1331-1392.

Abrahart, R.J., and See, L., (2002). Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments, ***Hydrology and Earth System Sciences***, 6(4), 655-670.

ASCE Task Committee on Application of the Artificial Neural Networks in Hydrology, (2000). Artificial neural networks in hydrology II: hydrologic applications, ***J. Hydrol. Engng***, ASCE, 5 (2), 124–137.

Bowden, G.J., Dandy, G.C., Maier, H.R., (2005). Input determination for neural network models in water resources applications: Part 1—background and methodology, ***Journal of Hydrology,*** 301, 75–92.

Castellano-Méndeza, M., González-Manteigaa, W., Febrero-Bande, M., Manuel Prada-Sáncheza, J., and Lozano-Calderón, R., (2004). Modeling of the monthly and daily behavior of the runoff of the Xallas river using Box–Jenkins and neural networks methods, ***Journal of Hydrology***, 296, 38–58.

Daubechies, I., (1992). Ten Lectures on Wavelets CSBM-NSF Series, ***Application Mathematics***, 61. SIAM publication, Philadelphia PA.

Dawson, C. W. and Wilby, R. L.,(1999). A comparison of artificial neural networks used for river flow forecasting, ***Hydrology and Earth System Sciences***, 3, 529–540.

Dawson, C. W., and Wilby, R. L., (2001). Hydrological Modeling Using Artificial Neural Networks, ***Progress in Physical Geography***, 25(1), 80-108.

De Vos, N.J. and Rientjes, T.H.M., (2005). Constraints of artificial neural networks for rainfall -runoff modeling: trade-offs in hydrological state representation and model evaluation, ***Hydrology and Earth System Sciences***, 9, 111-126.

Elsner, J. B. and Tsonis, A. A.. (1997). ***Singular Spectrum Analysis: A New Tool in Time Series Analysis***, Plenum Press, New York.

Farmer, J. D., and Sidorowich, J. J., (1987). Predicting chaotic time series, ***Phys. Rev. Lett.***, 59(4), 845–848.

Fraser, A.M. and Swinney, H.L., (1986). Independent coordinates for strange attractors from mutual information, ***Physical Review A***, 33(2), 1134-1140.

Howell, J.F. and Mahrt, L., (1994). ***'An adaptive decomposition: application to turbulence' in Wavelet in Geophysics***, Academic Press, New York, 107–128.

Jain, S. K., Das, A., and Drivastava, D. K., (1999). Application of ANN for reservoir inflow prediction and operation, ***J. Water. Resour. Plann. Manage.***, 125(5), 263–271.

Jain, A., and Srinivasulu, S., (2004). Development of effective and efficient rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques, ***Water Resour. Res.***, 40, W04302.

Kennel, M. B., Brown, R., Abarbanel, H. D. I., (1992). Determining embedding dimension for phase space reconstruction using geometrical construction. ***Phy. Rev. A.***, 45(6), 3403-3411.

Kişi, O., (2003). River flow modeling using artificial neural networks, ***Journal of Hydrologic Engineering***, 9(1), 60-63.

Kişi, O., (2005). Daily river flow forecasting using artificial neural networks and auto-regressive models, ***Turkish J. Eng. Env. Sci,*** 29, 9-20.

Kişi, O., (2008). Constructing neural network sediment estimation models using a data-driven algorithm, ***Mathematics and Computers in Simulation***, 79, 94–103.

Kitanidis, P. K. and Bras, R. L., (1980). Real-time forecasting with a conceptual hydrologic model, 2, applications and results, ***Wat. Resour. Res.***, 16 (6), 1034–1044.

Kücüķ, M. and Aġiralİoğlu, N., (2006).Wavelet Regression Technique for Stream flow Prediction, ***Journal of Applied Statistics***, 33(9), 943–960.

Laio, F., Porporato, A., Revelli, R., and Ridolfi, L., (2003). A comparison of nonlinear flood forecasting methods, ***Water Resour. Res.***, 39(5), 1129, doi:10.1029/2002WR001551.

Lee, C.F., Lee, J.C. and Lee, A.C., (2000). ***Statistics for business and financial economics*** (2nd version), World Scientific, Singapore.

Legates, D.R., McCabe Jr., G.J., (1999). Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation, ***Water Resour. Res.*** 35 (1), 233–241.

Lisi, F., Nicolis, and Sandri, M., (1995). Combining singular-spectrum analysis and neural networks for time series forecasting, ***Neural Processing Letters***, 2(4), 6-10.

Mallat, S.G., (1989). A theory for multi-resolution signal decomposition: the wavelet representation. ***IEEE transactions on pattern analysis and machine intelligence***, 11(7), 674-692.

Mallat, S. G., (1998). ***A wavelet tour of signal processing***, Academic, San Diego, 577.

Marques, C.A.F., Ferreira, J., Rocha, A., Castanheira, J., Gonçalves, P., Vaz. N., and Dias, J.M., (2006). Singular spectral analysis and forecasting of hydrological time series, ***Physics and Chemistry of the Earth***, 31, .1172-1179.

Muttil, N. and Chau, K.W., (2006). Neural network and genetic programming for modelling coastal algal blooms. ***Int. J. Environment and Pollution***, 28, 3/4, 223–238.

Nash, J. E. and Sutcliffe, J. V., (1970). River flow forecasting through conceptual models; part I – a discussion of principles, ***Journal of Hydrology***, 10, 282–290, 1970.

Partal. T. and Cigizoglu, H.K., (2008). Estimation and forecasting of daily suspended sediment data using wavelet–neural networks, ***Journal of Hydrology,*** 358, (3-4), 317-331.

Partal, T. and Kişi, Ö., (2007). Wavelet and Neuro-fuzzy conjunction model for precipitation forecasting. ***Journal of Hydrology***, 342, (2), 199-212.

Newbold, P., Carlson, W. L., and Thorne, B.M. (2003). ***Statistics for business and economics*** (fifth version), Prentice Hall, Upper Saddle River, N.J.

Raman, H., and Sunilkumar, N., (1995). Multivariate modeling of water resources time series using artificial neural networks, ***Hydrological Sciences Journal***, 40(2):145-163.

Sivakumar, B., Jayawardena, A. W., and Fernando, T. M. K., (2002). River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches, ***Journal of Hydrology***, 265(1), 225-245.

Sivapragasam, C., Liong, S.Y., and Pasha, M.F.K., (2001). Rainfall and discharge forecasting with SSA-SVM approach, ***Journal of Hydroinformatics***, 3(7), 141–152.

Sudheer, K.P., Gosain, A.K., and Ramasastri, K.S., (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models, ***Hydrol. Processes***, 16, 1325–1330.

Thirumalaiah, K., and Deo, M. C., (2000). Hydrological forecasting using neural networks. ***Journalof Hydrologic Engineering***, 5(2), 180-189.

Torrence, C. and Compo, G.P., (1998). A practical guide to wavelet analysis, ***Bulletin of the American Meteorological Society***, 79, 61–78.

Vautard, R., Yiou, P. & Ghil, M., (1992). Singular-spectrum analysis: a toolkit for short, noisy and chaotic signals, ***Physica D*** 58, 95–126.

Wang, W., van Gelder, P.H.A.J.M. Vrijling, J.K., and Ma, J., (2006). Forecasting Daily Streamflow Using Hybrid ANN Models, ***Journal of Hydrology***, 324, 383-399.

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

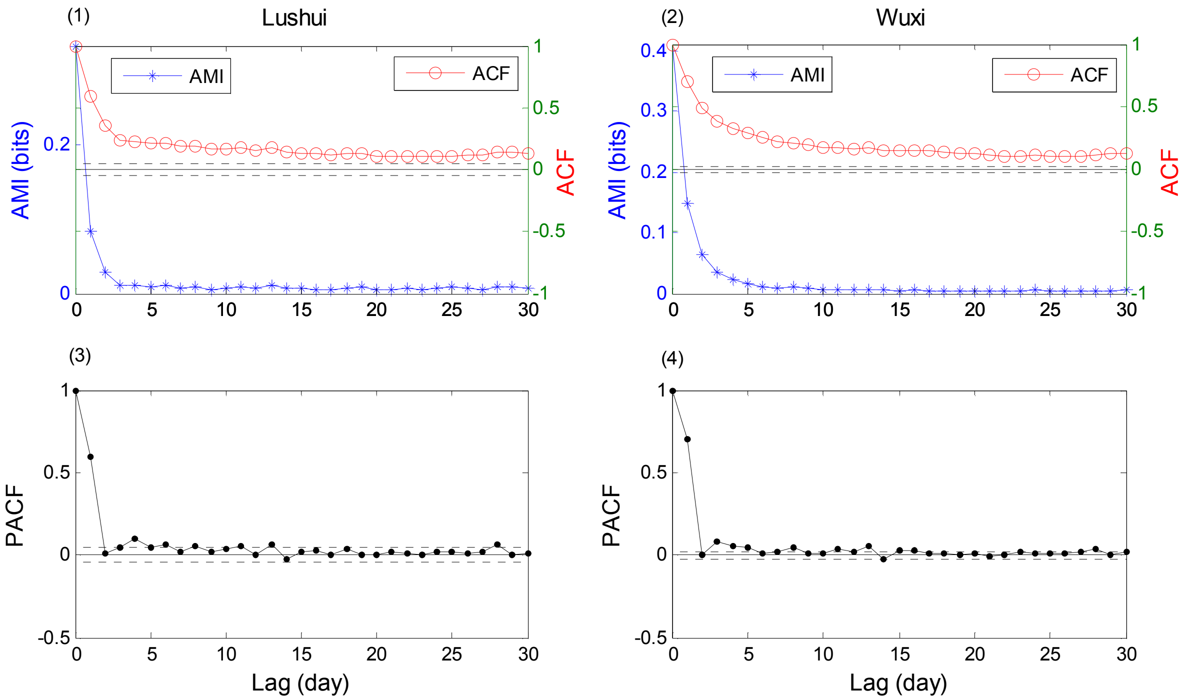
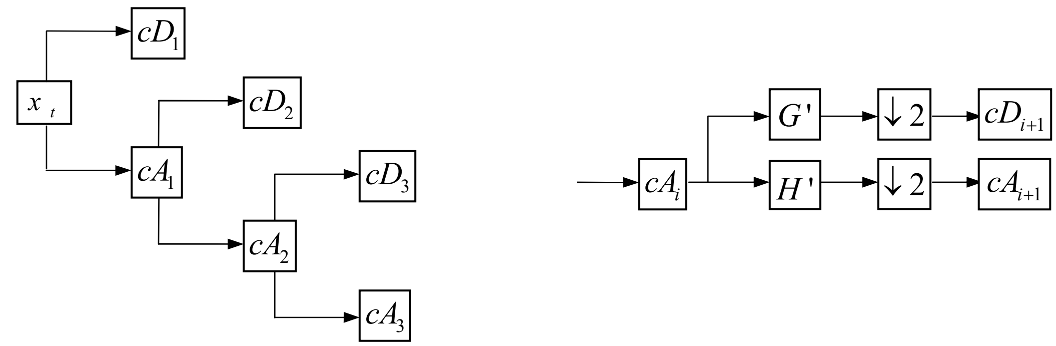


Figure 1. Plots of ACF, AMI, and PACF of the flow data ((1) and (3) for Lushui; (2) and (4) for Wuxi) where the dashed lines stand for 95% confidence bound.

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93



(a) (b)

Figure 2. Schematics of WMRA (a) perform decomposition of x_t at level 3 and (b) filter signal

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

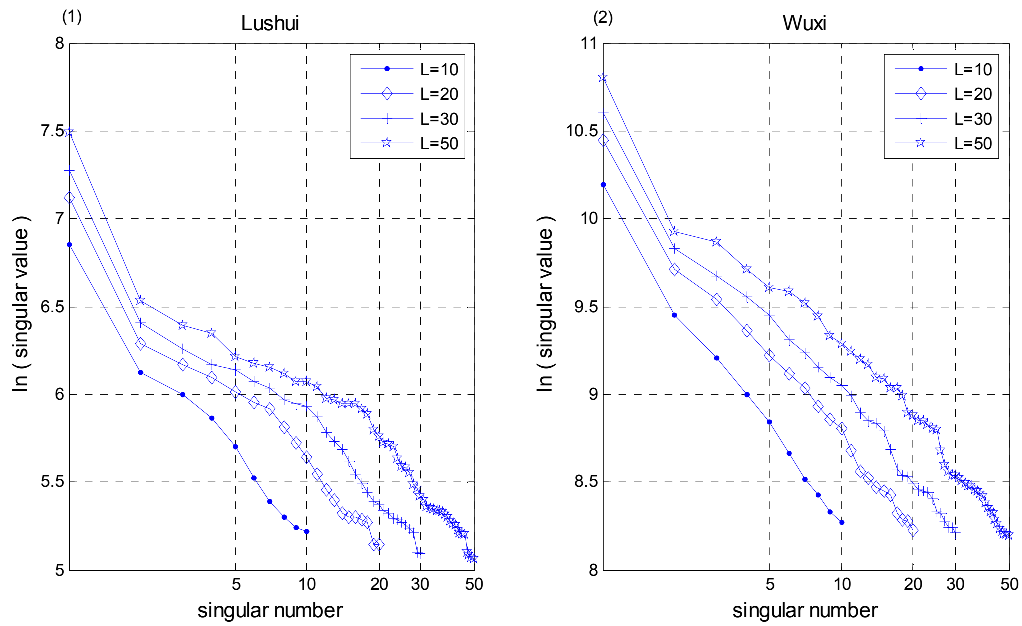


Figure 3. Singular Spectrum for (1) Lushui and (2) Wuxi with different L

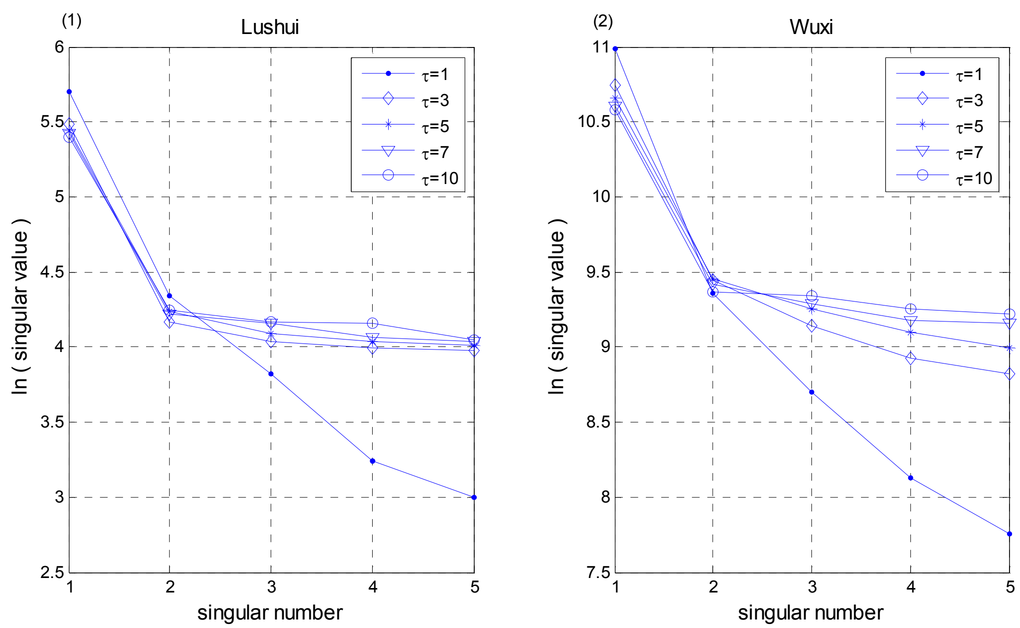


Figure 4. Singular Spectrum for (1) Lushui and (2) Wuxi with different τ

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

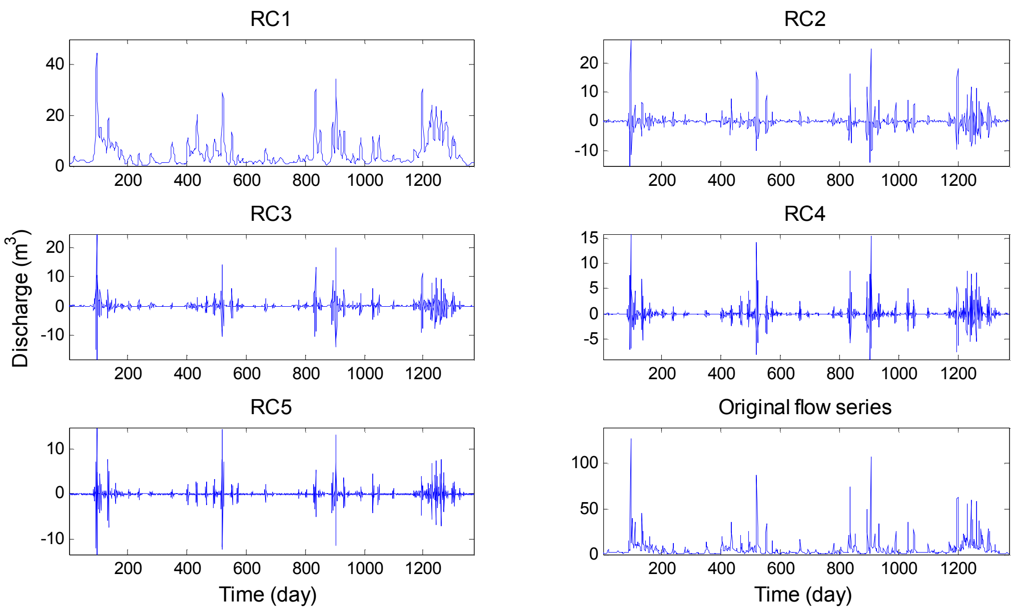


Figure 5. Reconstructed components (RCs) by SSA and original flow series of Lushui

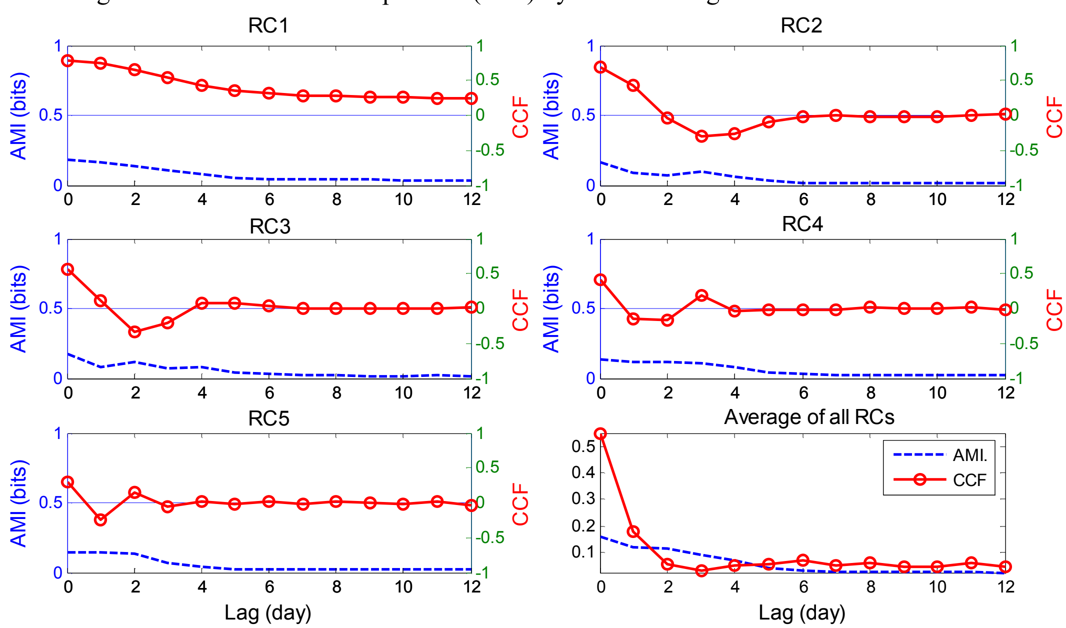


Figure 6. Plots of AMI and CCF between RC and the raw flow data of Lushui

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

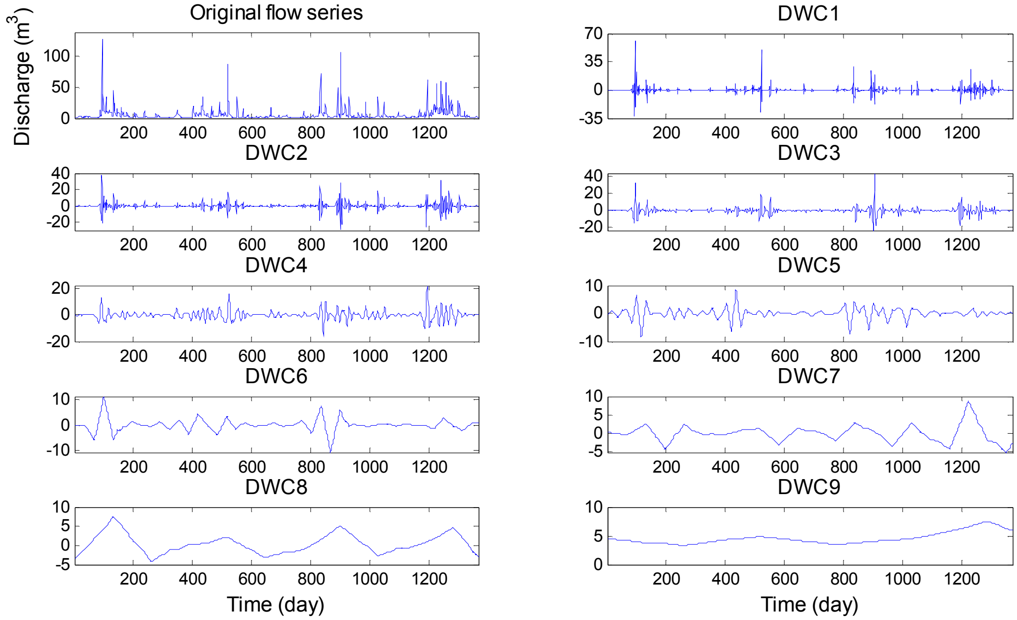


Figure 7. Discrete wavelet components (DWCs) and original flow series of Lushui

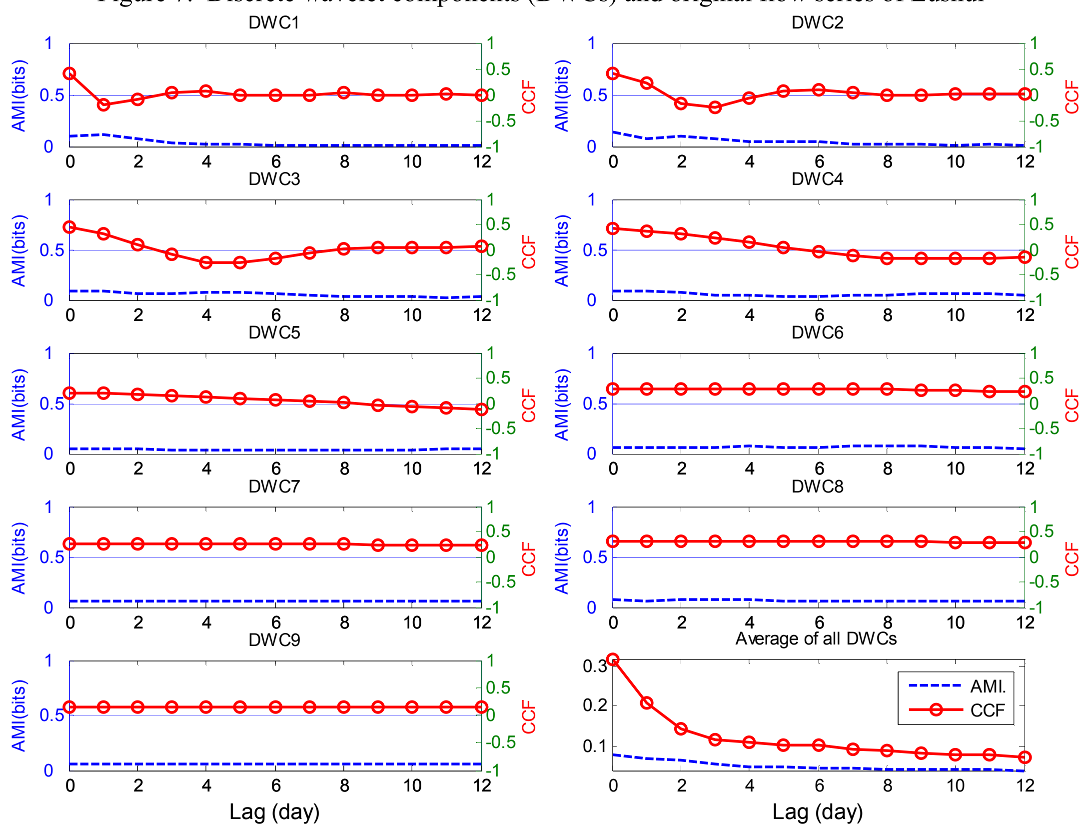


Figure 8. Plots of AMI and CCF between DWC and the raw flow data of Lushui

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

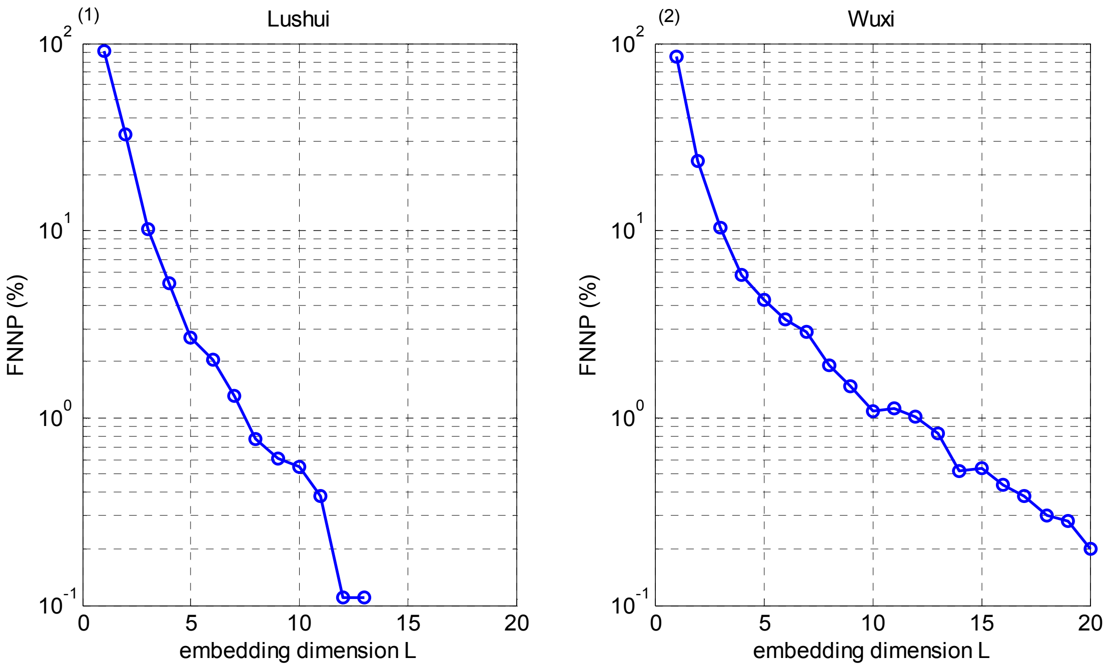


Figure 9. FNNP as a function of the embedding dimension for (1) Lushui and (2) Wuxi when $\tau = 1$ and $R_{tol} = 15$

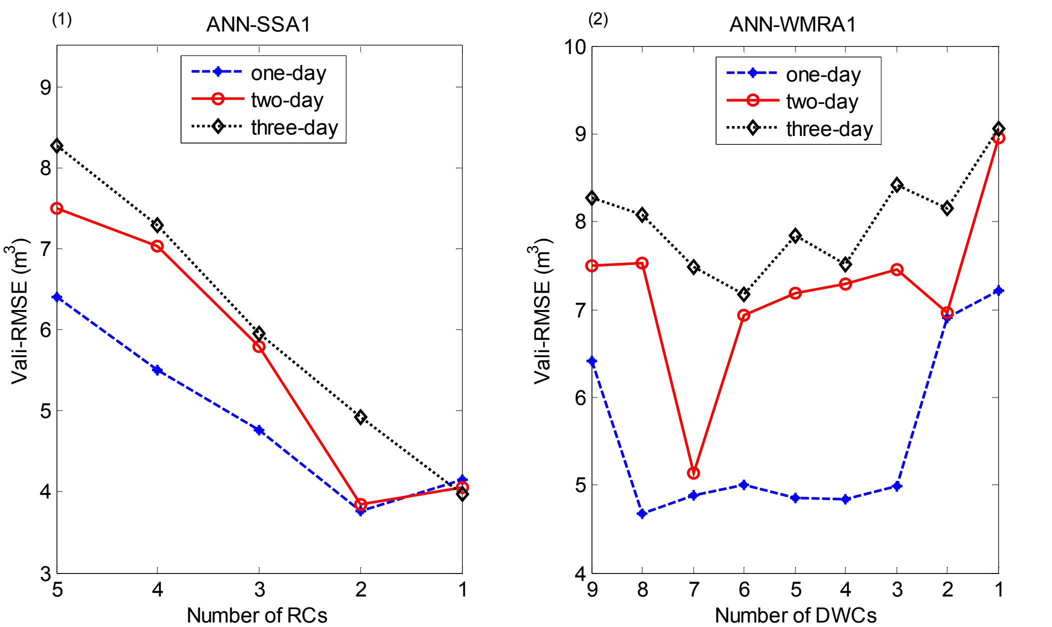


Figure 10. Performances of (1) ANN-SSA1 and (2) ANN-WMRA1 as a function of p ($\leq L$) at different prediction horizons (based on the Lushui flow data)

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

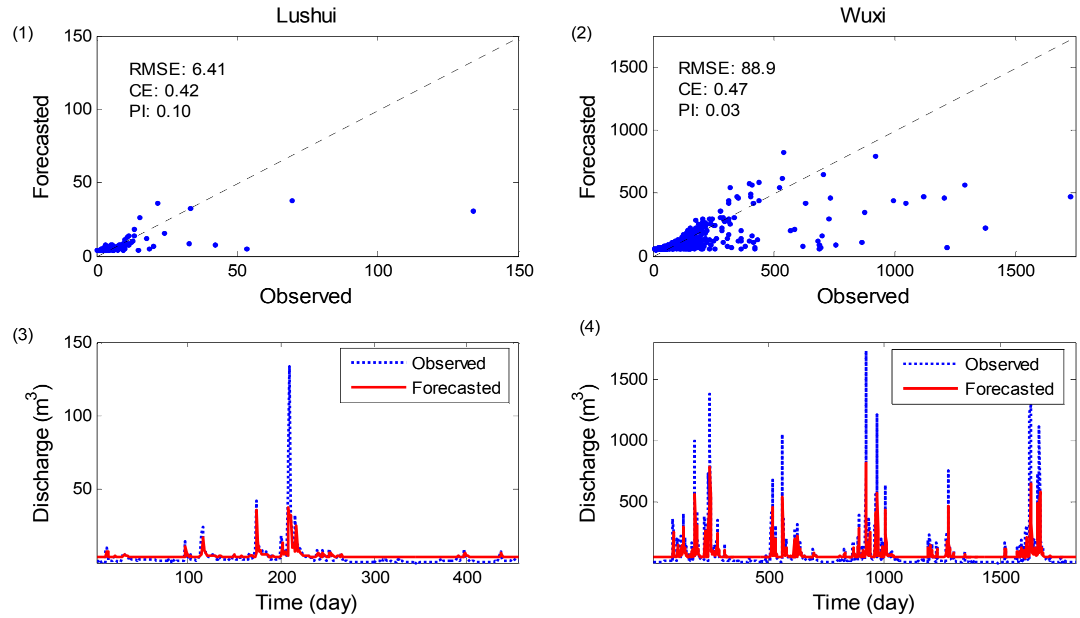


Figure 11. Scatter plots and hydrographs of the results of one-day-ahead forecast by the ANN model using the Lushui data ((1) and (3)) and Wuxi data ((2) and (4))

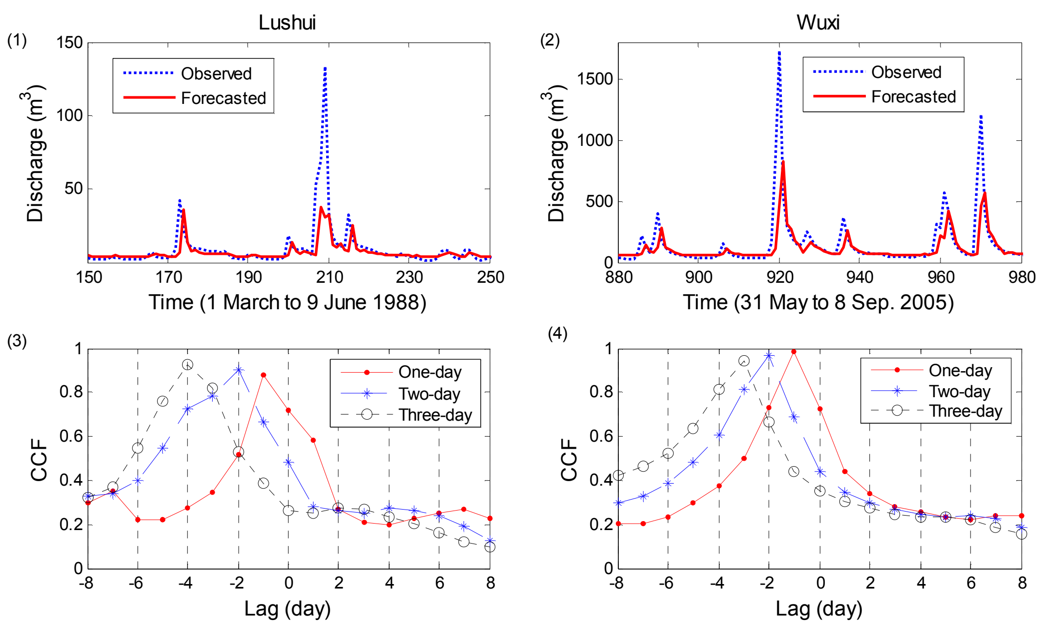


Figure 12. Representative detail of observed and forecasted discharges for one-day-ahead forecast and CCF between observed and forecasted dischargs at three forecast horizons from the ANN model ((1) and (3) for Lushui; (2) and (4) for Wuxi)

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

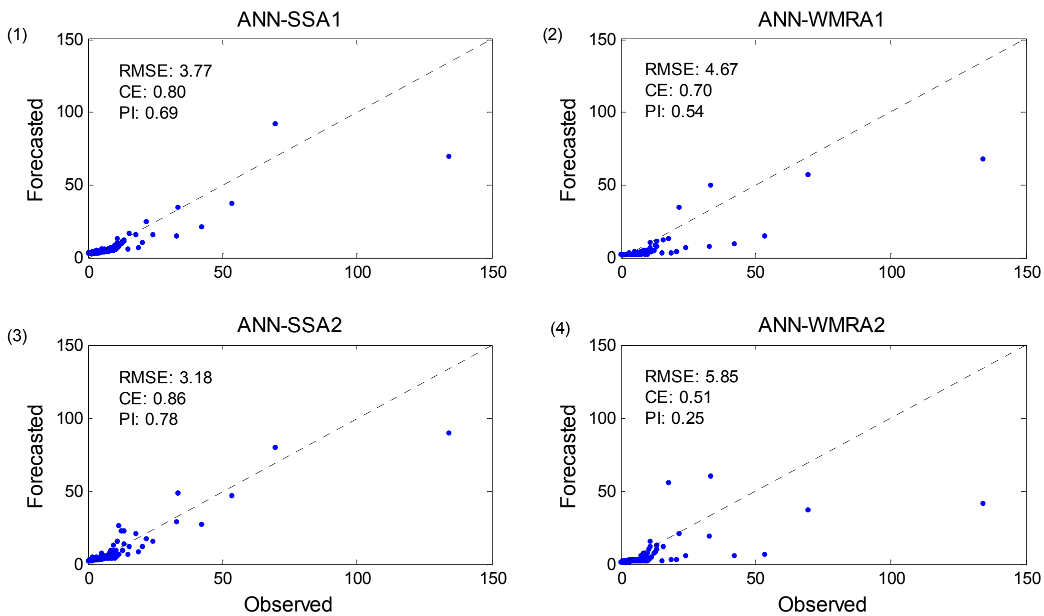


Figure 13. Scatter plots of observed and forecasted Luishui discharges for one-day-ahead forecast using (1) ANN-SSA1, (2) ANN-WMRA1, (3) ANN-SSA2, and (4) ANN-WMRA2

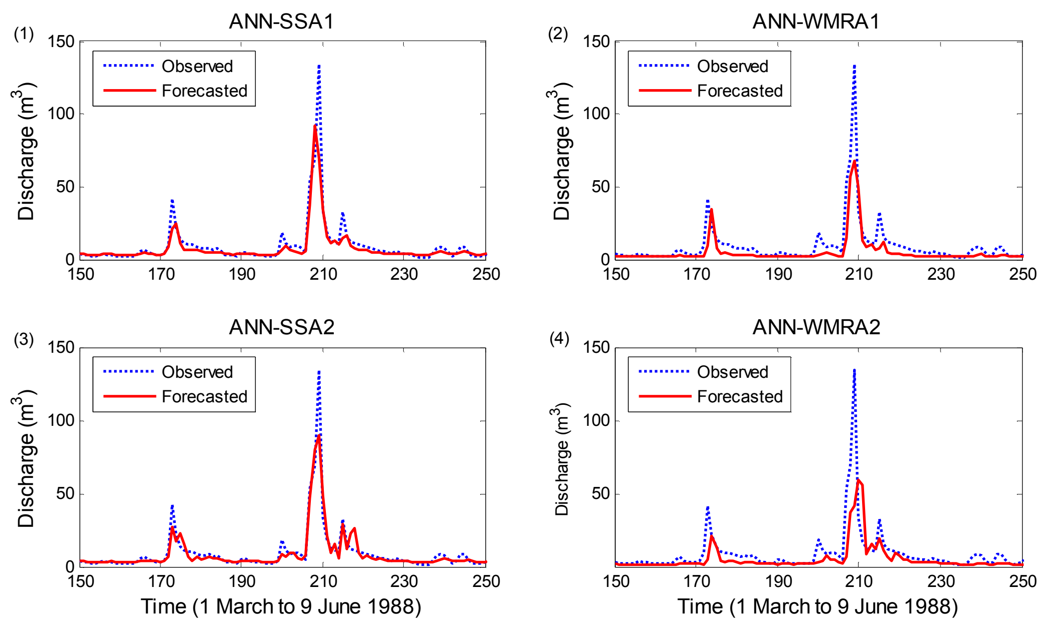


Figure 14. Representative detail of observed and forecasted Luishui discharges for one-day-ahead forecast using (1) ANN-SSA1, (2) ANN-WMRA1, (3) ANN-SSA2, and (4) ANN-WMRA2

Journal of Hydrology, Vol. 372, No. 1-4, 2009, pp 80–93

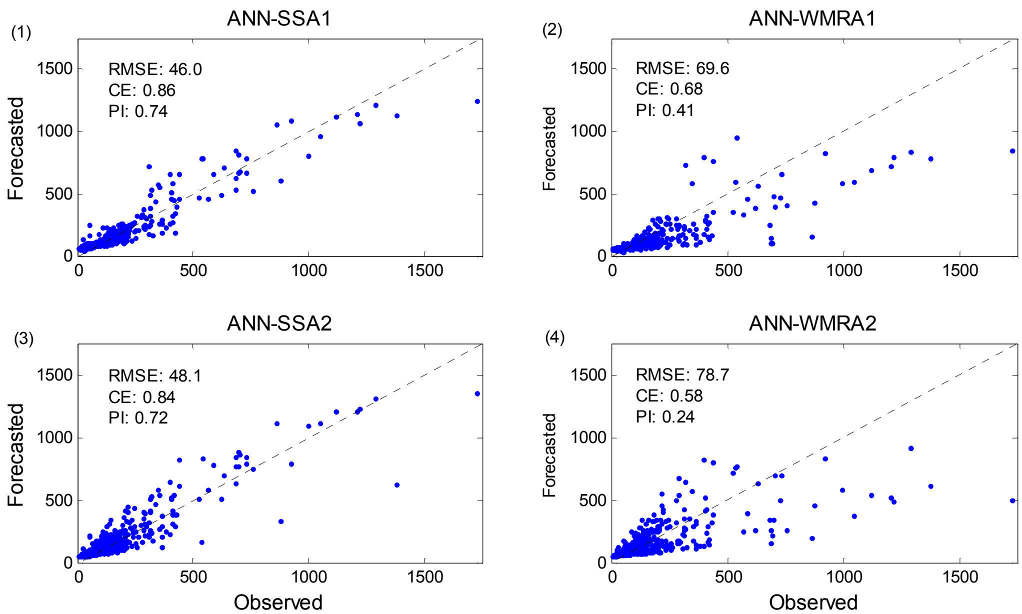


Figure 15. Scatter plots of observed and forecasted Wuxi discharges for one-day-ahead forecast using (1) ANN-SSA1, (2) ANN-WMRA1, (3) ANN-SSA2, and (4) ANN-WMRA2

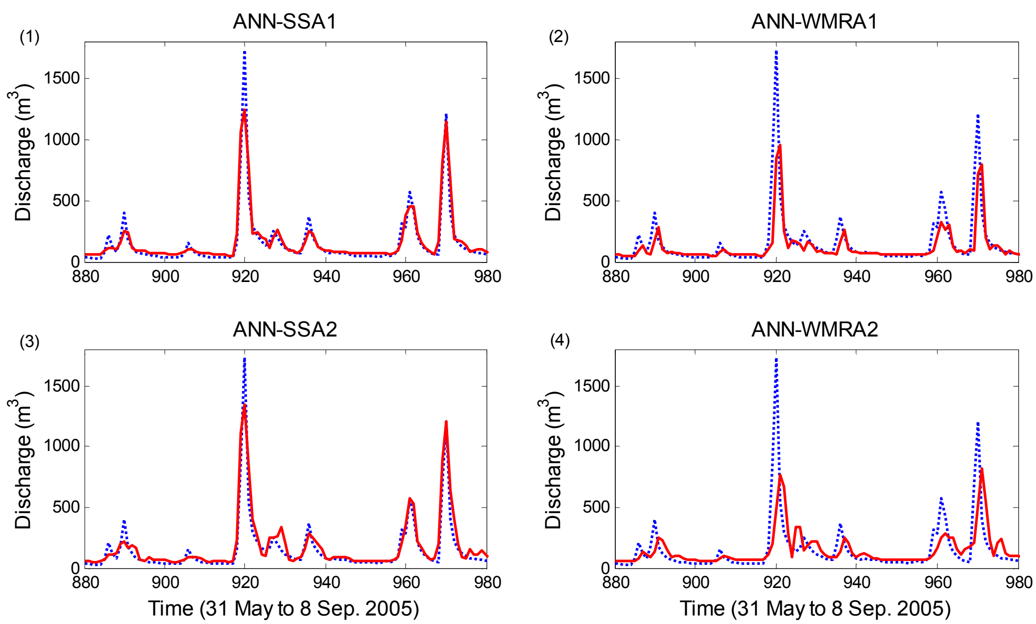


Figure 16. Representative detail of observed and forecasted Wuxi discharges for one-day-ahead forecast using (1) ANN-SSA1, (2) ANN-WMRA1, (3) ANN-SSA2, and (4) ANN-WMRA2

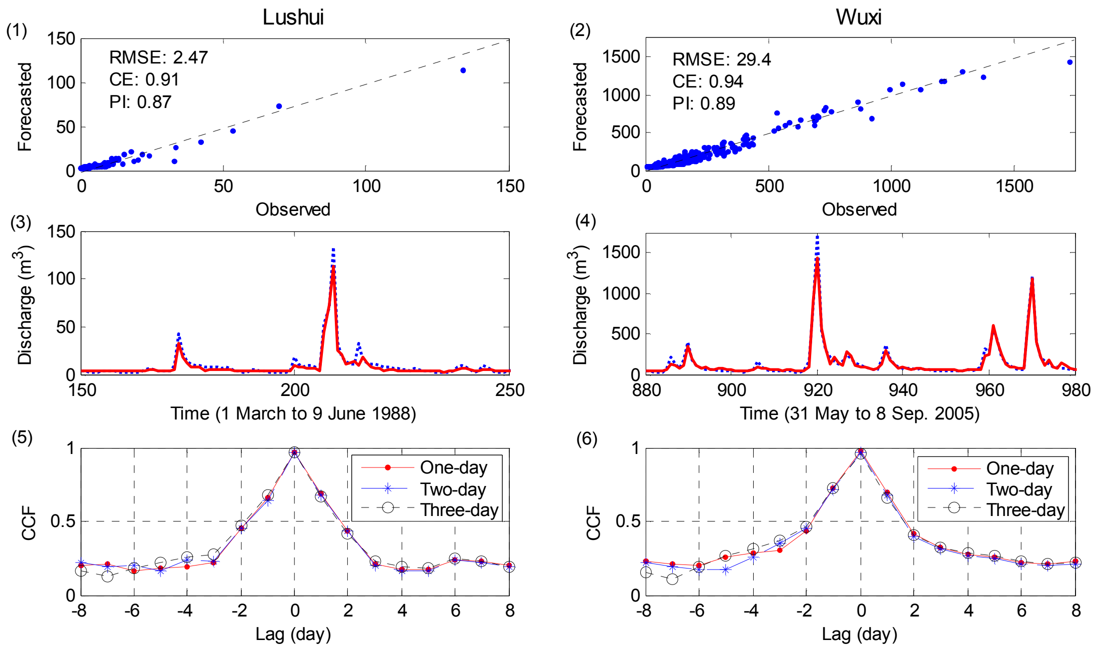


Figure 17. Forecast results from the ANN-MA model for Lushui ((1),(3), and (5)) and Wuxi ((2),(4), and (6)) where (1) and (2) denote scatter plots, (3) and (4) are representative details, and (5) and (6) show the CCF between forecasts and observed discharges at three prediction levels

Table 1 Related information for two rivers and the flow data

Watershed and datasets	Statistical parameters						Watershed area and data period
	μ (m^3)	S_x (m^3)	C_v	C_s	X_{min} (m^3)	X_{max} (m^3)	
Lushui							
Original data	4.63	8.49	0.55	7.38	0.02	134	Area:
Training	4.41	8.56	0.52	7.84	0.02	128	224 km^2
Cross-validation	5.78	8.35	0.69	3.84	0.05	63	Data period:
Testing	3.90	8.39	0.47	10.22	0.30	134	Jan., 1984- Dec., 1988
Wuxi							
Original data	61.9	112.6	0.55	7.20	6.0	2230	Area:
Training	60.6	95.6	0.63	5.90	7.6	1530	2 001 km^2
Cross-validation	60.7	132.2	0.46	8.35	6.0	2230	Data period:
Testing	66.0	122.1	0.54	6.30	10.1	1730	Jan., 1988- Dec., 2007

Table 2 Effective components for the ANN-SSA2 and ANN-WMRA2 inputs at various forecasting horizons (based on the flow data of Lushui)

Model	Model inputs	Effective components		
		One-day lead	Two-day lead	Three-day lead
ANN-SSA2				
	Q_{t-1}	1, 2*[a]	1, 5	1, 4
	Q_{t-2}	1, 5	1, 4	1
	Q_{t-3}	1, 4	1	1
	Q_{t-4}	1	1	1
	Q_{t-5}	1	1	1
	Q_{t-6}	1	1	1
ANN-WMRA2				
	Q_{t-1}	4, 8, 3, 6, 7, 2, 5, 9*[b]	8, 4, 6, 7, 5, 9	8, 6, 7, 4, 5, 9
	Q_{t-2}	8, 4, 6, 7, 5, 9, 3	8, 6, 7, 4, 5, 9	8, 6, 7, 9, 4, 5
	Q_{t-3}	8, 6, 7, 4, 5, 9, 1	8, 6, 7, 9, 4, 5	8, 6, 7, 9, 5
	Q_{t-4}	8, 6, 7, 9, 4, 5, 1	8, 6, 7, 9, 5	8, 6, 7, 9
	Q_{t-5}	8, 6, 7, 9, 5, 2, 4	8, 6, 7, 9	8, 6, 7, 9
	Q_{t-6}	8, 6, 7, 9, 2, 5	8, 6, 7, 9	8, 6, 7, 9

*[a] the numbers of '1, 2' denote RC1 and RC2;
*[b] the numbers of '4, 8, 3, 6, 7, 2, 5, 9' stand for DWC4, DWC8, DWC3, DWC6, DWC7, DWC2, DWC5, and DWC9, and the sequence of these numbers is in a descending order of their correlation coefficients.

Table 3 Model performances at various forecasting horizons using testing data of Lushui

Model	RMSE			CE			PI		
	1*	2*	3*	1	2	3	1	2	3
ANN	6.41	7.50	8.27	0.42	0.20	0.03	0.10	0.26	0.36
ANN-MA	2.47	2.63	2.60	0.91	0.90	0.90	0.87	0.91	0.94
ANN-SSA1	3.77	3.85	3.98	0.80	0.79	0.77	0.69	0.81	0.85
ANN-SSA2	3.18	3.47	3.85	0.86	0.83	0.79	0.78	0.84	0.86
ANN-WMRA1	4.67	5.14	7.17	0.70	0.62	0.27	0.54	0.65	0.52
ANN-WMRA2	5.85	6.26	6.91	0.51	0.44	0.32	0.25	0.48	0.55

* The number of '1, 2, and 3' denote one-, two-, and three-day-ahead forecasts

Table 4 Model performances at various forecasting horizons using testing data of Wuxi

Model	RMSE			CE			PI		
	1*	2*	3*	1	2	3	1	2	3
ANN	88.9	111.0	114.7	0.47	0.17	0.12	0.03	0.24	0.32
ANN-MA	29.4	39.4	41.5	0.94	0.90	0.88	0.89	0.90	0.91
ANN-SSA1	46.0	50.4	50.5	0.86	0.83	0.83	0.74	0.84	0.87
ANN-SSA2	48.1	45.3	50.5	0.84	0.86	0.83	0.72	0.87	0.87
ANN-WMRA1	69.6	82.3	92.2	0.68	0.55	0.43	0.41	0.59	0.56
ANN-WMRA2	78.7	86.9	94.1	0.58	0.49	0.41	0.24	0.54	0.54

* The number of '1, 2, and 3' denote one-, two-, and three-day-ahead forecasts