Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Prediction of Rainfall Time Series Using Modular Artificial Neural Networks Coupled with Data Preprocessing Techniques

C. L. Wu[1,2], K. W. Chau[1,*], and C. Fan[3]

[1] Dept. of Civil and Structural Engineering, Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong, People's Republic of China

[2] Changjiang Institute of Survey, Planning, Design and Research,
Changjiang Water Resources Commission,
430010, Wuhan, HuBei, People's Republic of China

[3] Department of Civil Engineering, Ryerson University,
350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3

*Email: cekwchau@polyu.edu.hk

ABSTRACT

This study is an attempt to seek a relatively optimal data-driven model for rainfall forecasting from three aspects: model inputs, modeling methods, and data preprocessing techniques. Four rain data records from different regions, namely two monthly and two daily series, are examined. A comparison of seven input techniques, either linear or nonlinear, indicates that linear correlation analysis (LCA) is capable of identifying model inputs reasonably. A proposed model, modular artificial neural network (MANN), is compared with three benchmark models, viz. artificial neural network (ANN), K-nearest-neighbors (K-NN), and linear regression (LR). Prediction is performed in the context of two modes including normal mode (viz., without data preprocessing) and data preprocessing mode. Results from the normal mode indicate that MANN performs the best among all four models, but the advantage of MANN over ANN is not significant in monthly rainfall series forecasting. Under the data preprocessing mode, each of LR, K-NN and ANN is respectively coupled with three data preprocessing techniques including moving average (MA), principal component analysis (PCA), and singular spectrum analysis (SSA). Results indicate that the improvement of model performance generated by SSA is considerable whereas those of MA or PCA are slight. Moreover, when MANN is coupled with SSA, results show that advantages of MANN over other models are quite noticeable, particularly for daily rainfall forecasting. Therefore, the proposed optimal rainfall forecasting model can be derived from MANN coupled with SSA.

KEYWORDS

Rainfall prediction, Modular artificial neural network, Moving Average, Principal component analysis, Singular spectral analysis, Fuzzy C-Means clustering, K-nearest-neighbors

1. Introduction

Accurate and timely rainfall forecasting is crucial for reservoir operation and flooding prevention because it can provide an extension of lead-time of the flow forecasting, larger than the response time of the watershed, in particular for small and medium-sized mountainous basins.

Many studies have been conducted for the quantitative precipitation forecasting using diverse techniques including numerical weather prediction models and remote sensing observations (Yates et al., 2000; Ganguly and Bras, 2003; Sheng, et al., 2006; Doomede et al., 2008), statistical models (Chu and He, 1995; Chan and Shi, 1999; DelSole and Shukla, 2002; Munot, 2007; Li and Zeng, 2008; Nayagam et al., 2008), chaos theory-based approach (Jayawardena and Lai, 1994), K-nearest-neighbor (K-NN) method (Toth et al, 2000), and soft computing methods including artificial neural network (ANN), support vectors regression (SVR) and fuzzy inference system (Venkatesan et al., 1997; Silverman and Dracup, 2000;

Toth et al., 2000; Pongracz et al., 2001; Sivapragasam et al., 2001; Brath et al., 2002; Lin and Chen, 2005; Chattopadhyay and Chattopadhyay, 2007; Guhathakurta, 2008; Lin et al., 2009). Venkatesan et al. (1997) employed ANN to predict all India summer monsoon rainfall with different meteorological parameters as model inputs. Toth et al. (2000) applied three data-driven models, auto-regressive moving average, ANN and K-NN, to short-term rainfall predictions. Results showed that ANN performed the best in terms of the accuracy of runoff forecasting when the predicted rainfalls by the three models were used as inputs of a rainfall-runoff model. Pongracz et al. (2001) applied fuzzy inference to monthly rainfall prediction. Chattopadhyay and Chattopadhyay (2007) constructed an ANN model to predict monsoon rainfall in India depending on the rainfall series alone.

Recently, the concept of coupling different models has attracted more attention in hydrologic forecasting. They can be broadly categorized into ensemble models and modular (or hybrid) models. The basic idea behind ensemble models is to build several different or similar models for the same process and to integrate them together (Shamseldin et al., 1997; Shamseldin and O'Connor, 1999; Xiong et al., 2001; Abrahart and See, 2002; Kim et al, 2006). For example, Xiong et al. (2001) used a Takagi-Sugeno-Kang fuzzy technique to couple several conceptual rainfall-runoff models. Coulibaly et al. (2005) employed an improved weighted-average method to coalesce forecasted daily reservoir inflows from K-NN, conceptual model and ANN. Kim et al. (2006) investigated five ensemble methods for improving stream flow prediction.

Physical processes in rainfall and/or runoff are generally composed of a number of sub-processes. Their accurate modeling by building of a single global model is sometimes not possible (Solomatine and Ostfeld, 2008). Modular models were therefore proposed where sub-processes were first of all identified and then separate models (also called local or expert model) were established for each of them (Solomatine and Ostfeld, 2008). Different modular models were proposed depending on the soft or hard splitting of training data. Soft splitting means the dataset can be overlapped and the overall forecasting output is the weighted-average of each local model (Zhang and Govindaraju, 2000; Shrestha and Solomatine, 2006; Wu et al., 2008). Zhang and Govindaraju (2000) examined the performance of modular networks in predicting monthly discharges based on the Bayesian concept. Wu et al. (2008) employed a distributed SVR for daily river stage prediction. On the contrary, there is no overlap of data in the hard splitting and the final forecasting output is explicitly from only one of local models (See and Openshaw, 2000; Hu et al., 2001; Solomatine and Xue, 2004; Sivapragasam and Liong, 2005; Jain and Srinivasulu, 2006; Wang et al., 2006; Corzo and Solomatine, 2007; Lin and Wu, 2009). Hu et al. (2001) developed a range-dependent network which employs a number of multilayer perceptron neural networks to model the river flow in different flow bands of magnitude (e.g. high, medium and low). Their results indicated that the range-dependent network performed better than the conventional global ANN. Solomatine and Xue (2004) used M5 model trees and neural networks in a flood-forecasting problem. Sivapragasam and Liong (2005) divided the flow range into three regions, and employed different SVR models to predict daily flows in high, medium and low regions. Wang et al. (2006) used a crisp modular ANN to make soft or crisp predictions for validation data where each local network was trained using the subsets achieved by either a threshold discharge value or a clustering of input spaces. Lin and Wu (2009) proposed a hybrid ANN model for event-based hourly rainfall prediction where self-organizing map networks are

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

used for data cluster analysis and multilayer conceptron networks are employed to serve each cluster to construct mapping between input and output.

A hydrological time series can be actually regarded as an integration of stochastic (or random) and deterministic components (Salas et al., 1985). Once the stochastic (noise) component is appropriately eliminated, the deterministic component can then be easily modeled. For the purpose of cleaning hydrological series, many data preprocessing techniques, including Principal component analysis (PCA), wavelet analysis (WA), and singular spectrum analysis (SSA), have been employed in hydrology field by researchers (Sivapragasam et al., 2001; Marques et al., 2006; Hu et al., 2007; Partal and Kişi, 2007; Sivapragasam et al., 2007;Wu et al., 2009). Hu et al. (2007) employed PCA as an input data preprocessing tool to improve the prediction accuracy of the rainfall-runoff neural network models. The use of WA to improve rainfall forecasting was conducted by Partal and Kişi (2007). Their results indicated that WA was promising. SSA has also been recognized as an efficient preprocessing algorithm to avoid the effect of discontinuous or intermittent signals, coupled with neural networks (or similar approaches) for time series forecasting (Lisi et al., 1995; Sivapragasam et al., 2001; Baratta et al., 2003). For example, Lisi et al. (1995) applied SSA to extract the significant components in their study on southern oscillation index time series and used ANN for prediction. They reconstructed the original series by summing up the first “p” significant components. Sivapragasam et al. (2001) proposed a hybrid model of support vector machine (SVM) and SSA for rainfall and runoff predictions. The hybrid model resulted in a considerable improvement in the model performance in comparison with the original SVM model. A comparison between WA and SSA in Wu et al. (2009) indicated that SSA performed better than WA. In addition, moving average (MA) is used for data preprocessing to improve the performance of ANN by de Vos and Rientjes (2005). They argued that one of reasons on lagged predictions of ANN was due to the use of previous observed data as ANN inputs and suggested that an effective solution was to obtain new model inputs by MA over the original data series.

In this paper, one of the main purposes is to develop a modular ANN (MANN) coupled with appropriate data-preprocessing techniques to improve the accuracy of rainfall forecasting. MANN consists of three local models which are associated with three subsets clustered by the fuzzy C-means (FCM) clustering method. To evaluate MANN, LR, K-NN and ANN are employed for comparison. ANN is first used to choose the best model inputs by seven candidate model inputs techniques. Once all forecasting models are established, three data-preprocessing methods (i.e., MA, PCA, and SSA) can be examined. To ensure wider application of the conclusions, four cases consisting of two monthly rainfall series and two daily rainfall series from India and China, are investigated. The remaining part is structured as follows. Methodology is detailed in Section 2 where case studies are first described, and then data-preprocessing techniques and forecasting models are introduced. Section 3 presents modeling methods and their applications to four rainfall series. The optimal model input method and the best data preprocessing can be identified. In Section 4, principal results are shown along with relevant discussions. The last section presents main conclusions.

2. Methodology

2.1 Study Area and Data

Two daily mean rainfall series from Daning and Zhenshui river basins of China, and two monthly mean rainfall series from India and Zhongxian of China, are analyzed.

The Daning River, a first-order tributary of the Yangtze River, is located at the northeastern side of Chongqing city. The daily rainfall data from Jan. 1, 1988 to Dec. 31, 2007 were measured at six raingauges located at the upstream of the study basin (Figure 1). The upstream part is controlled by Wuxi hydrology station, with a drainage area of around 2 000 km^2. The mean areal rainfall series is calculated by the Thiessen polygon method (hereafter the averaged rainfall series is referred to as Wuxi).

The Zhenshui basin is located at the northern side of Guangdong Province and adjoined by Hunan Province and Jianxi Province. The basin belongs to a second-order tributary of the Pearl River and has an area of 7,554 km^2. The daily rainfall time series of Zhenwan raingauge was collected between January 1, 1989 and December 31, 1998 (hereafter the averaged rainfall series is referred to as Zhenwan).

The all Indian average monthly rainfall is estimated from area-weighted observations at 306 land stations uniformly distributed over India. The data, with period spanning from January 1871 to December 2007, are available at the website http://www.tropmet.res.in run by the Indian Institute of Tropical Meteorology.

The other monthly rainfall series is from Zhongxian raingauge which is located at Chongqing city, China. The catchment containing this raingauge belongs to a first-order tributary of the Yangtze River. The monthly rainfall data were collected from January 1956 to December 2007.

Figure 2 shows hyetographs of four rainfall series. A linear fit to each hyetograph is denoted by the dashed line. All series appear stationary at least in a weak sense since these linear fits are close to horizontal.

In this study, each of data series is 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 has dual functions: one is to implement an early stopping approach in order to avoid overfitting of the training data and another is to select some best predictions from a number of ANN's runs. In the present study, 10 best predictions are selected from a total of 20 ANN's runs. The same data partition is adopted for each series: the first half of the entire 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 pertinent information about watersheds 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 data. Owing to the weak extrapolation ability of ANN, all data are scaled to the interval [-0.9, 0.9] instead of [-1, 1] whilst hyperbolic tangent sigmoid functions are employed as transfer functions in hidden and output layers.

2.2 Data preprocessing techniques

(1) MA

MA smoothes data by replacing each data point with the average of the k neighboring data points, where k may be termed 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 equally weighted MA is the most commonly-used, in which each value of the data carries the same weight in the smoothing process. There are three types of moving modes including centering, backward and forward. In a forecasting scenario, only the backward mode is used since the other two modes may necessitate future observed values. For a time series $\{x_1, x_2, \cdots, x_N\}$, when the backward moving mode is adopted (Lee et al., 2000), the k-term unweighted moving average y_t^* is written as

$$y_t^* = \left(\sum_{i=0}^{k-1} y_{t-i}\right) \Big/ k \tag{1}$$

where $t = k, \cdots, N$. The choice of the window length k is by a trial and error procedure with a minimization of the loss of the objective function.

(2) PCA

PCA was first introduced by Pearson (1901) and developed independently by Hotelling (1933), and has now well entrenched as an important technique in data analysis. The central idea is to reduce the dimensionality of a data set consisting of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. The PCA approach uses all of the original variables to obtain a smaller set of principal components (PCs) which can be used to approximate the original variables. PCs are uncorrelated and are ordered so that the first few retain most of the variation present in the original set.

Consider a data matrix $\mathbf{X}$ which has n rows (observations) and p column (variables). Let the covariance matrix of $\mathbf{X}$ be Σ, where $\Sigma = \text{cov}(\mathbf{X}) = E(\mathbf{X}^T\mathbf{X})$. The linear transformed orthogonal matrix $\mathbf{Z}$ is presented as

$$\mathbf{Z} = \mathbf{XA} \tag{2}$$

where $\mathbf{Z}$ is the PCs with elements (i, j) of ith observation and jth principal component; $\mathbf{A}$ is a ($p \times p$) matrix with eigenvector elements of the covariance of $\mathbf{X}$, and having $\mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}$.

Because matrix $\mathbf{X}^T\mathbf{X}$ is real and symmetric, it can be expressed as $\mathbf{X}^T\mathbf{X} = \mathbf{A\Lambda A^T}$ where $\mathbf{\Lambda}$ is a diagonal matrix whose nonnegative entries are the eigenvalues ($\lambda_i, i = 1, \cdots, p$) of $\mathbf{X}^T\mathbf{X}$. The total variance of the data matrix $\mathbf{X}$ is represented as

$$\text{trace}(\Sigma) = \text{trace}(\mathbf{A\Lambda A^T}) = \text{trace}(\mathbf{\Lambda}) = \sum_{i=1}^{p} \lambda_i \tag{3}$$

On the other hand, the covariance matrix of principal components $\mathbf{Z}$ is expressed as

$$\text{cov}(\mathbf{Z}) = E(\mathbf{Z}^T\mathbf{Z}) = E(\mathbf{A}^T\mathbf{X}^T\mathbf{X}\mathbf{A}) = \mathbf{\Lambda} \tag{4}$$

$$\text{trace}(\mathbf{Z}) = \text{trace}(\mathbf{\Lambda}) = \sum_{i=1}^{p} \lambda_i \tag{5}$$

Therefore, the total variance of the data matrix $\mathbf{X}$ is identical to the total variance after PCA transformation $\mathbf{Z}$.

The solution of PCA, using singular value decomposition (SVD) or determinants of the covariance matrix of $\mathbf{X}$, can provide the eigenvectors $\mathbf{A}$ with their eigenvalues, $\lambda_i, i = 1, \cdots, p$, representing the variance of each component after PCA transformation. If the eigenvalues are ordered by $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \cdots \geq \lambda_p \geq 0$, the first few PCs can capture most of the variance of the original data while the remaining PCs mainly represent the noise in the data. The percentage of total variance explained by the first mth PCs is

$$V = \sum_{i=1}^{m} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i \cdot 100\% \tag{6}$$

The higher is the selection of the total data variance, V, the better the properties of the data matrix are preserved. For the sake of the reduction of dimensionality, a small number of PCs are selected, though most of the data variance in selected components still remain. If the transformation is to prevent the collinearity of regression variables, the selected component number m in Eq. (6) can be set for a higher total variance, such as $V = 95\% \sim 99\%$ (Hsu et al., 2002).

The original data matrix $\mathbf{A}$ can be reconstructed by a reverse operation of Eq. (2) as

$$\mathbf{X} = \mathbf{Z}\mathbf{A}^T \tag{7}$$

By choosing suitable m ($\leq p$) PCs from $\mathbf{Z}$ and accompanying m eigenvectors from $\mathbf{A}$, the original data can be filtered.

(3) SSA

According to Golyandina et al. (2001), the basic SSA consists of two stages: decomposition and reconstruction. The decomposition stage involves two steps: embedding and SVD; the reconstruction stage also comprises two steps: grouping and diagonal averaging. Consider a real-valued time series $F = \{x_1, x_2, \cdots, x_N\}$ of length $N(> 2)$. Assume that the series is a nonzero series, viz. there exists at least one i such that $x_i \neq 0$. Four steps are briefly presented as follows.

1st step: embedding

The embedding procedure maps the original time series to a sequence of multi-dimensional lagged vectors. Let L be an integer (window length), $1 < L < N$, and τ be the delayed time at a multiple of the sampling period. The embedding procedure forms $n = N - (L-1)\tau$ lagged vectors $\mathbf{x}_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \cdots, x_{i+(L-1)\tau}\}^T$, where $\mathbf{x}_i \in \mathrm{R}^L$, and

$i = 1, 2, \cdots, n$. The 'trajectory matrix' of the time series is denoted by $\mathbf{X} = [\mathbf{x}_1 \ \cdots \ \mathbf{x}_i \ \cdots \ \mathbf{x}_n]$ having lagged vectors as its columns. In other words, the trajectory matrix is

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_n \\ x_{1+\tau} & x_{2+\tau} & x_{3+\tau} & \cdots & x_{n+\tau} \\ x_{1+2\tau} & x_{2+2\tau} & x_{3+2\tau} & \cdots & x_{n+2\tau} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{1+(L-1)\tau} & x_{2+(L-1)\tau} & x_{3+(L-1)\tau} & \cdots & x_N \end{pmatrix} \tag{8}$$

If $\tau = 1$, the matrix $\mathbf{X}$ is called Hankel matrix since it has equal elements on the 'diagonals' where the sum of subscripts of row and column is equal to constant. If $\tau > 1$, the equal elements in $\mathbf{X}$ are not definitely in the 'diagonals'.

2nd step: SVD

Let $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, $\lambda_1, \lambda_2, \cdots, \lambda_L$ denote the eigenvalues of $\mathbf{S}$ taken in the decreasing order of magnitude ($\lambda_1 \ge \lambda_2 \ge \lambda_3 \ge \cdots \ge \lambda_L \ge 0$) and $\mathbf{U}_1, \mathbf{U}_2, \cdots, \mathbf{U}_L$ denote the orthonormal system of the eigenvectors of the matrix $\mathbf{S}$ corresponding to these eigenvalues. If we denote $\mathbf{V}_i = \mathbf{X}_i^T \mathbf{U}_i \big/ \sqrt{\lambda_i}$ ($i = 1, \cdots, L$) (equivalent to the ith eigenvector of $\mathbf{X}^T\mathbf{X}$), then the SVD of the trajectory matrix $\mathbf{X}$ can be written as

$$\mathbf{X} = \mathbf{X}_1 + \cdots + \mathbf{X}_L \tag{9}$$

where $\mathbf{X}_i = \sqrt{\lambda_i}\mathbf{U}_i\mathbf{V}_i^T$. The matrices $\mathbf{X}_i$ have rank 1; therefore they are elementary matrices. The collection ($\lambda_i, \mathbf{U}_i, \mathbf{V}_i$) is called the ith eigentriple of the SVD. Note that $\mathbf{U}_i$ and $\mathbf{V}_i$ are also the ith left and right singular vectors of $\mathbf{X}$, respectively.

3rd step: grouping

The purpose of this step is to identify appropriately the trend component, oscillatory components with different periods, and structureless noises by grouping components. This step can be skipped if one does not want to precisely extract hidden information by regrouping and filtering of components.

The grouping procedure partitions the set of indices $\{1,\cdots,L\}$ into m disjoint subsets $I_1,\cdots,I_m$, so that the elementary matrix in Eq. (9) is regrouped into m groups. Let $I=\{i_1,\cdots,i_p\}$. Then the resultant matrix $\mathbf{X}_I$ corresponding to the group I is defined as $\mathbf{X}_I=\mathbf{X}_{i_1}+\cdots+\mathbf{X}_{i_p}$. These matrices are computed for $I_1,\cdots,I_m$. By substituting into the expansion (9), one obtains the new expansion

$$\mathbf{X}=\mathbf{X}_{I_1}+\cdots+\mathbf{X}_{\mathrm{I_m}} \tag{10}$$

The procedure of choosing the sets $I_1,\cdots,I_m$ is called the eigentriple grouping.

4th step: Diagonal averaging

The last step in the basic SSA is the transformation of each resultant matrix of the grouped decomposition (10) into a new series of length N. The diagonal averaging is to find equal elements in the resultant matrix and then to generate a new element by averaging over them. The new element has the same position (or index) as the corresponding elements in the original series. As mentioned in the step 1, the concept of 'diagonal' is not true for $\tau>1$. Regardless of the value of τ being larger than or equal to 1, the principle of reconstruction is the same. For $\tau=1$, the diagonal averaging can be carried out by the formula recommended by Golyandina et al. (2001). Let $\mathbf{Y}$ be a ($L\times n$) matrix with elements y_{ij}, $1\le i\le L$, $1\le j\le n$. Let $L^*=\min(L,n)$, $n^*=\max(L,n)$ and $N=n+(L-1)\tau$. Let $y_{ij}^*=y_{ij}$ if $L<n$ and $y_{ij}^*=y_{ji}$ otherwise. Diagonal averaging transfers matrix $\mathbf{Y}$ to a series $\{y_1,y_2,\cdots,y_N\}$ by the following equation

$$y_k = \begin{cases} \dfrac{1}{k}\sum\limits_{m=1}^{k} y^*_{m,k-m+1} & for\, 1 \le k < L^* \\ \dfrac{1}{L^*}\sum\limits_{m=1}^{L^*} y^*_{m,k-m+1} & for\, L^* \le k \le K^* \\ \dfrac{1}{N-k+1}\sum\limits_{m=k-K^*+1}^{N-K^*+1} y^*_{m,k-m+1} & for\, L^* < k \le N \end{cases} \quad (11)$$

Eq. (11) corresponds to the averaging of the matrix elements over the 'diagonals' $i + j = k + 1$. The diagonal averaging, applied to a resultant matrix $\mathbf{X}_{I_k}$, produces a $N-$length series F_k, and thus the original series F is decomposed into the sum of m series:

$$F = F_1 + \cdots + F_m \quad (12)$$

As mentioned above, these reconstructed components (RCs) can be associated with the trend, oscillations or noise of the original time series with proper choices of L and the sets of $I_1, \cdots, I_m$. Certainly, if the third step (namely, grouping) is skipped, F can be decomposed into L RCs.

2.3 Forecasting models

This section describes four candidate forecasting models. They are LR, K-NN, ANN and MANN. They are usually called data-driven models because they capture the mapping between input (e.g. antecedent rainfall) and output variables (forecasted rainfall) without directly considering the physical laws that underlie the mechanism of rainfall (or precipitation). These models are purely based on the information retrieved from the collected rainfall data.

(1) Construction of input/output pairs

Let $\{x_1, x_2, \cdots, x_N\}$ stand for a rainfall time series. It can be reconstructed into a series of delay vectors as $\mathbf{X}_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, \cdots, x_{t+(m-1)\tau}\}$, where $\mathbf{X}_t \in \mathrm{R}^m$, τ is the delay time as a multiple of the sampling period and m is the embedded dimension. Suppose that the rainfall $x_{t+T+(m-1)\tau}$ at T-step lead is related to the vector $\mathbf{X}_t$, the available historical data may be summarized into a set of pairs as $\{\mathbf{X}_t, x_{t+T+(m-1)\tau} : t = 1, \cdots, n\}$, where n stands for the number of pairs, and $n = N - (m-1)\tau$.

The functional relationship between the input vector $\mathbf{X}_t$ at time t and the predicted output $x^F_{t+T+(m-1)\tau}$ at time $t+T$ can be written as follows:

$$x^F_{t+T+(m-1)\tau} = f(\mathbf{X}_t) + e_t \tag{13}$$

where e_t is a typical noise term, $x^F_{t+T+(m-1)\tau}$ is the prediction of $x_{t+T+(m-1)\tau}$, and $f(\bullet)$ is the mapping function. The difference of various data-driven forecasting models used in the current study relies on the way of approximating $f(\bullet)$ once model inputs are attained with the appropriate selection of (τ, m).

(2) LR

The linear regression model herein is actually called stepwise linear regression (SLR) model because the forward stepwise regression is used to determine optimal input variables. The basic idea of SLR is to start with a function that contains the single best input variable and to subsequently add potential input variables to the function one at a time in an attempt to improve the model performance. The order of addition is determined by using the partial F- test values to select which variable should enter next. The high partial F- value is compared to a (selected or default) F-to-enter value. After a variable has been added, the function is examined to see if any variable should be deleted. Interested readers are referred to Draper and Smith (1998) and McCuen (2005) for more details.

(3) K-NN

The prediction of $x_{t+T+(m-1)\tau}$ by the K-NN method is formulated as:

$$x^F_{t+T+(m-1)\tau} = \frac{1}{K} \sum_{t \in S(\mathbf{X},n)} x_{t+T+(m-1)\tau} \tag{14}$$

where $\mathrm{S}(\mathbf{X}, \mathrm{n})$ denotes the set of indices t of the K nearest neighbors to the feature vector $\mathbf{X}(\mathrm{n})$. The meaning of "nearest neighbors" is generally interpreted in a Euclidean sense. Therefore, if i belongs to $\mathrm{S}(\mathbf{X}, \mathrm{n})$ and j is not in $\mathrm{S}(\mathbf{X}, \mathrm{n})$, then according to Euclidean distance $\|\mathbf{X}_n - \mathbf{X}_i\| \le \|\mathbf{X}_n - \mathbf{X}_j\|$. Intuitively speaking, the forecast $x^F_{t+T+(m-1)\tau}$ in Eq. (14) is the sample average of output rainfall of the K nearest neighbors to $\mathbf{X}(\mathrm{n})$. Obviously, a key task is to determine the parameter K in the K-NN method.

(4) ANN

The multilayer perceptron network is by far the most popular ANN paradigm, which usually uses the technique of error back propagation to train the network configuration. The architecture of the ANN consists of a number of hidden layers and a number of neurons in the 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. Based on Eq. (13), the ANN forecasting model is formulated as

$$x^F_{t+T+(m-1)\tau} = f(\mathbf{X}_t, w, \theta, m, h) = \theta_0 + \sum_{j=1}^{h} w_j^{out} \varphi\left(\sum_{i=1}^{m} w_{ji} x_{t+(i-1)\tau} + \theta_j\right) \tag{15}$$

where φ denotes transfer functions; w_{ji} are the weights defining the link between the ith node of the input layer and the jth node of the hidden layer; θ_j are biases associated to the jth node of the hidden layer; w_j^{out} are the weights associated to the connection between the jth node of the hidden layer and the node of the output layer; and θ_0 is the bias at the output node. To apply Eq. (15) to rainfall predictions, appropriate training algorithm is required to optimize w and θ.

(5) MANN

To construct MANN, the training data have to be divided into several clusters according to cluster analysis techniques, and then each single model is applied to each cluster. The FCM clustering technique is adopted in the present study (e.g., Bezdek, 1981, Wang et al., 2006). It is able to generate either soft or crisp clusters. ANN (or similar techniques) is unable to extrapolate beyond the range of the data used for training. Otherwise, poor forecasts or predictions can be expected when a new input data is outside the range of those used for training. Hard forecasting is, therefore, taken into consideration in this study.

Figure 3 displays the schematic diagram of MANN where the training data is partitioned into three clusters which are based on an assumption that three magnitudes of rainfall (i.e., low, medium, and high) may be derived from different mechanisms. According to this flow chart, once input-output pairs are obtained, they are first split into three subsets by the FCM technique, and then each subset is approximated by a single ANN. The final output of the modular model results directly from the output of one of three local models.

2.4 Implementation framework of rainfall forecasting

Figure 4 illustrates the implementation framework of rainfall forecasting where four prediction models can be conducted in two modes: without/with three data preprocessing methods (dashed box). These acronyms in the column of "methods for model inputs" represent seven methods to determine model inputs: LCA (linear correlation analysis, Sudheer et al., 2002), AMI (average mutual information, Fraser and Swinney, 1986), PMI (partial mutual information, May et al., 2008), FNN (false nearest neighbors, Kennel et al., 1992), CI (correlation integral, Theiler, 1986), SLR, and MOGA (ANN based on multi-objective genetic algorithm, Giustolisi and Simeone, 2006).

2.5 Evaluation of model performances

The Pearson'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., root mean square error (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 are therefore used in this study. They are listed below.

$$\mathrm{CE} = 1 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{16}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{17}$$

$$\mathrm{PI} = 1 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} (y_i - y_{i-l})^2 \tag{18}$$

In these equations, n is the number of observations, $\hat{y}_i$ stands for the forecasted flow, y_i represents the observed flow, $\bar{y}$ denotes the average observed flow, and y_{i-l} is the flow estimate from a so-call persistence model (or termed 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. A small value of PI may imply occurrence of lagged prediction.

3. Applications of Models

3.1 Determination of model inputs

ANN, equipped with the Levernberg-Marquardt (L-M) training algorithm and hyperbolic tangent sigmoid transfer functions, is used as the benchmark model to examine aforementioned seven model input methods in terms of RMSE. Depending on the simplified algorithm from Yu et al. (2000) (downloaded at http://small.eie.polyu.edu.hk/), the four rainfall series are identified as non-chaotic since the correlation dimension does not display the property of convergence, in particular, for daily rainfall series. Results from remaining six methods are presented in Table 2. These results are based on one step lead prediction and let X_{t+1} be the target value at one-step prediction horizon. It can be seen from RMSE that most of these methods tend to be mutually alternative because their RMSE are close. Owing to the convenience of operation, the LCA method is preferred in this study. Furthermore, Figure 5 shows identification of effective inputs in Table 2 for the LCA method. Taking Wuxi and Zhenwan as examples, model inputs should take the previous 5-day and 7-day rainfalls for them respectively because the partial auto-correlation function (PACF) value decays within the confidence band around these time lags.

3.2 Identification of models

The model identification is to determine the structure of a candidate model by using training data to optimize relevant parameters of model control once model inputs have been obtained. The LR model is built by the SLR technique. In terms of one step prediction (viz., $T = 1$), input variables can be found in Table 2. For example, the LR model for Wuxi can be expressed as

$$x_{t+1}^F = 0.421x_t - 0.043x_{t-1} + 0.044x_{t-2} + 0.025x_{t-4} + 0.036x_{t-7} + 0.03x_{t-11} \quad (19)$$

With respect to K-NN, the model identification consists in finding the optimal K if the $m-$dimensional input vector is determined. Sugihara and Mary (1990) suggested that the value of K was taken as $K = m+1$. On the other hand, the choice of K should ensure the reliability of the forecasting (Fraser and Swinney, 1986). The check of robustness of $K = m+1$ in terms of RMSE is presented in Figure 6, where K is in the interval of [2, 40]. Adopting the value of K as $m+1$ seems reasonable for the current study because the difference between its RMSE and the minimum RMSE is only 2.9% for Wuxi, 2.9% for Zhenwan, 2.6% for India, and 2.0% for Zhongxian, respectively. Consequently, the value of K is 6 for Wuxi ($m = 5$), 8 for Zhenwan ($m = 7$), 13 for India ($m = 12$), and 14 for Zhongxian ($m = 13$), respectively.

Based on Eq. (14), the formula for one-step lead prediction in the context of K-NN can be defined as

$$x_{t+1}^F = \frac{1}{K}\sum_{i=1}^{K} x_{t_i+1} \quad (20)$$

where X_{t_i+1} stands for an observed value associated with a neighbor of the current state. For a $T-$step lead prediction, Eq. (20) becomes

$$x_{t+T}^{F} = \frac{1}{K}\sum_{i=1}^{K} x_{t_i+T} \tag{21}$$

The identification of ANN structure is to optimize the number of hidden nodes h in the hidden layer with the known model inputs and output. The optimal size h of the hidden layer is found by systematically increasing the number of hidden neurons from 1 to 10 until the network performance on the cross-validation set no longer improves significantly. Based on the L-M training algorithm and hyperbolic tangent transfer functions, the identified configurations of ANN are 5-5-1 for Wuxi, 7-4-1 for Zhenwan, 12-5-1 for India, and 13-3-1 for Zhongxian, respectively. The same method is used to identify the structure of MANN, and the only difference is that the identification is repeated three times, with each time being for a local ANN. Consequently, MANN is obtained as 5-5/7/9-1 for Wuxi, 7-4/8/4-1 for Zhenwan, 12-3/2/5-1 for India, and 13-1/1/1-1 for Zhongxian, respectively.

It is worthwhile to notice that the standardization/normalization of the training data is very crucial in the improvement of the model performance. Two methods can be found in the literature (Dawson and Wilby, 2001; Cannas et al., 2002; Rajurkar et al, 2002; Campolo et al., 2003; Wang et al., 2006). The standardization (also termed rescaling in some papers) method, as adopted above for model input determination, is to rescale the training data to [-1, 1], [0, 1] or even more narrow interval depending on what kinds of transfer functions are employed in ANN. The normalization method is to rescale the training data to a Gaussian function with a mean of 0 and unit standard deviation, which is by subtracting the mean and dividing by the standard deviation. When the normalization approach is adopted, ANN uses the linear function (e.g. purelin) instead of the hyperbolic tangent sigmoid transfer function in the output layer. In addition, some studies have indicated that considerations of statistical principles may improve ANN model performance (e.g. Cheng and Titterington, 1994). For example, the training data was recommended to be normally distributed (Fortin et al., 1997). Sudheer et al. (2002) suggested that the issue of stationarity should be considered in the ANN development because the ANN cannot account for trends and heteroscedasticity in the data. Their results showed that data transformation to reduce the skewness of data was capable of significantly improving the model performance. For the purpose of obtaining better model performance, four data-transformed schemes are examined:

- Standardizing the raw data (referred to as Std_raw);
- Normalizing the raw data (referred to as Norm_raw);
- Standardizing the n-th root transformed data (referred to as Std_nth_root);
- Normalizing the n-th root transformed data (referred to as Norm_nth_root).

Table 3 compares the ANN model performance of the four schemes in terms of RMSE and CE. The Norm_raw scheme is, on the whole, slightly more effective than the Std_raw method. It can also be seen that the effect of the n-th root scheme (3 is taken after trial and error) on the improvement of the performance is basically negligible. Therefore, the Norm_raw scheme is adopted for the later rainfall prediction in the present study.

3.3 Rainfall data preprocessing

(1) MA

The MA operation entails the window length k in Eq. (1) to smooth the raw rainfall data. An appropriate k can be found by systematically increasing k from 1 to 10. The smoothed data is then used to feed into each forecasting model. The targeted value of k corresponds to the optimal model performance in terms of RMSE.

(2) PCA

PCA is employed in two ways: one for reduction of the dimensionality or preventing collinearity (depending on Eq. (2)); second for noise reduction by choosing leading components (contributing most of the variance of the original rainfall data) to reconstruct rainfall series (depending on Eq. (7)). The percentage V of total variance (see Eq. (6)) is set at three horizons, 85%, 90%, and 95% for principal component selection.

(3) SSA

This approach of filtering a time series to retain desired modes of variability is based on the idea that the predictability of a system can be improved by forecasting the important oscillations in time series taken from the system. The general procedure is to filter the original record first and then to build the forecasting model based on the filtered series. To filter the raw rainfall series, the series needs to be decomposed into components with the aid of SSA. The decomposition by SSA requires identifying the parameter pair (τ, L). The value of an appropriate L should be able to clearly resolve different oscillations hidden in the original signal. However, the present study does not require accurately resolving the raw rainfall signal into trends, oscillations, and noises. A rough resolution can be adequate for the separation of signals and noises where some leading eigenvalues should be identified.

To select L, a small interval of [3, 10] is examined in the present study. Figure 7 shows the relation between singular spectrum (namely, a set of singular values) and singular number L for Wuxi, Zhenwan, India, and Zhongxian. It can be observed that the curve of singular values in each case except for Wuxi tends to level off with the increase of L. Generally, extraction of high-frequency oscillations becomes more difficult with the increase of singular number L (or mode). L is selected empirically by following the criterion that the singular spectrum can be distinguished markedly under that L. According to this criterion, L is set the value of 7 for India and Zhenwan, 6 for Zhongxian. For Wuxi, since all values in the interval satisfy the criterion, in order to reduce computational load in later filtering operation, L is set a small value of 5. The singular spectrum associated with the selected L is highlighted by the dotted solid line in Figure 7.

As regards τ, Figure 8 presents the results of sensitivity analysis of singular spectrum on the lag time τ using SSA with the determined L. For daily rainfall series, the singular spectrum can be distinguished only when $\tau = 1$. In contrast, the singular spectrum is insensitive to τ in the case of monthly rainfall series. The final parameter pair (τ, L) in SSA are set as (1, 5) for Wuxi, (1, 7) for Zhenwan, (1, 7) for India, (1, 6) for Zhongxian, respectively.

3.4 Filtering of RCs

The subsequent task is to reconstruct a new rainfall series as model inputs by finding contributing RCs so as to improve the predictability of the rainfall series. There is no practical guide on how to identify a contributing or noncontributing component to the improvement of accuracy of prediction. Two proposed filtering methods, supervised and unsupervised, are herein examined.

(1) Supervised filtering (denoted by SSA1)

Figure 9 depicts cross-correlation function (CCF) between RCs and the original Zhenwan rainfall series. The last plot in this figure presents the average of CCFs from all 7 RCs. The average indicates an overall correlation between input and output at various lags (also termed prediction horizons). The plot of average CCF shows that the best correlation is positive and occurs at lag 1. Among all 7 RCs, RC1 exhibits the best positive correlation with the original rainfall series. The CCF values for other RCs change alternatively between positive and negative with the increase of the lag. From the perspective of linear correlation, the positive or negative CCF value 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. With the assumption, deleting RCs, which have negative correlations with the model output if the average CCF is positive, may improve the performance of the forecast model. This is the basic idea behind the supervised method.

The procedure of the supervised method coupled with ANN is depicted in Figure 10. The aim is to find the optimal p ($\leq L$) RCs from all L RCs for each prediction horizon. The procedure can be summarized into three steps: SSA decomposition, correlation coefficients sorting, and reconstructed components filtering. Operation in each step is bounded by the dashed box. It is worth noting that the filtering method is based on assumption that combination of components with the same sign in CCF ($+$ or -) can strengthen the correlation with the model output.

(2) Unsupervised filtering (denoted by SSA2)

There are some drawbacks on the supervised method. The salient one is that this method relies on linear correlation analysis, which disregards the existence of nonlinearity in meteorological processes. Also, random combinations among all RCs are not taken into account. To overcome these drawbacks, an unsupervised filtering method (also termed enumeration) is recommended where all input combinations are examined. There are 2^L combinations for L RCs. The unsupervised method may be computationally intensive if L is large.

4. Results and Discussions

This section presents predictions using various models under two types of modes, namely "normal" and "data preprocessing". The "data preprocessing" mode is separately described by MA, PCA, and SSA. To extend one-step-ahead prediction to multi-step-ahead prediction, a direct multi-step prediction method (by directly having the multi-step-ahead prediction as output, also termed static prediction method) is adopted in this study to perform two- and three-step-ahead predictions.

4.1 Forecasting with normal mode

Table 4 shows results of three prediction horizons by applying five models including naïve model to each case study. The naïve model is used as the benchmark in which the forecasted value is directly equal to the last observed value (namely, no change). The naïve model presents the poorest forecasting which can be explained by the fact that it is unlikely to capture any dependence relation. From the perspective of rainfall series, the monthly rainfall can be better predicted than the daily rainfall. Generally, a daily rainfall series, in particular in a semi-humid and semi-dry or dry region, tends to be intermittent and discontinuous due to a large number of no rain periods (dry periods). Two global modeling methods, LR and ANN, mainly capture the zero-zero (or similar extreme low-intensity) rainfall patterns in daily rainfall series because the type of pattern is overwhelmingly dominant in the daily rainfall series. As a consequence, poor performance indices in terms of RMSE, CE, and PI can be observed (depicted in Table 4 for Wuxi and Zhenwan). Nevertheless, Table 4 also shows that MANN performs the best in each case study. MANN adopts three local ANN models, one for each cluster generated by FCM, which can better capture the mapping relation than using a single global ANN. It can be noticed that MANN is more effective for daily rainfall series than monthly rainfall data, which can be because daily rainfall data is more irregular (or non-periodic) than monthly rainfall series. The use of K-NN for daily rainfall forecasting is even worse than LR although it employs a local prediction approach. Apart from the issue of the selection of K, the performance of K-NN is also influenced by the similarity of input-output patterns. The smooth monthly rainfall series easily construct similar patterns so that they are well predicted by K-NN. It is worth noting that negative values occasionally appear in the forecasts of ANN or MANN whereas this situation does not happen in the K-NN method.

Take Wuxi and India data as representative examples, Figure 11 shows the scatter plots and hyetographs of the results at one-day-ahead prediction of ANN and MANN using the rainfall data of Wuxi, where the hyetograph is plotted in a selected range for better visual inspection. ANN seriously underestimates a number of moderate- and high-intensity rainfalls. The low values of CE and PI demonstrate that time shift between the forecasted and observed rainfall may occur, which is further verified by the hyetograph. MANN improves noticeably the accuracy of forecasting in terms of CE and PI. As shown by the scatter plots, the medium-intensity rainfall can be simulated better by MANN although high-intensity rainfalls (or peak values) are still underestimated. Figure 12 shows scatter plots and hyetographs of results at one-day-ahead prediction of ANN and MANN using the rainfall data of India. It can be seen from hyetograph graphs that both ANN and MANN reproduce well the corresponding observed rainfall data, which is further revealed by the scatter plots with a low dispersion around the exact fit line.

Figure 13 shows the analysis of the lag effect between forecasted and observed rainfall series. The value of CCF at zero lag corresponds to the actual performance (i.e. correlation coefficient) of the model. A target lag is associated with the maximum value of

CCF, and is an expression for the mean lag for the forecast. It can be seen from Figure 16 that ANN makes fairly obvious lagged predictions for daily rainfall series, and the lag effect can be overcome by MANN. There are 1, 2, and 3 days lag for Wuxi, which are respectively associated with one-, two-, and three-day-ahead forecasting. In contrast, there is no lag effect in monthly rainfall predictions of ANN or MANN.

4.2 Forecasting with MA

Table 5 presents forecasted results of ANN with the "backward" MA (hereafter referred to as ANN-MA) using the Wuxi rainfall data. The performance indices corresponding to $k=1$ are associated with the normal ANN. Results at each prediction horizon seem to be insensitive to the window length k in view of slight differences among each performance index for k from 1 to 10. Considering the fact that ANNs tend to generate unstable outputs, the influence of MA on the performance of ANN is negligible. Small values of PI also imply that MA cannot eliminate the lagged forecast from ANN.

4.3 Forecasting with PCA

As mentioned previously, PCA is used in two ways: one (denoted by PCA1) for reduction of dimensionality (also termed principal component regression) and the other one (denoted by PCA2) for noise reduction. Results from PCA1 are presented in Table 6. The scenario of $\mathrm{V}=100\%$ stands for forecasting using models with the normal mode. Results show that PCA1 cannot improve the model performances in terms of RMSE, CE, and PI, which means that the reduction of dimensionality is unnecessary for the present case studies. Actually, the original inputs are characterized by a low dimension. Table 7 describes the results from PCA2. According to results from LR and K-NN (because results from ANN tend to be unstable), a marginal improvement in the model performances can be observed for the

Wuxi watershed whereas the model performances deteriorate for the India watershed with the decrease of the value of V.

4.4 Forecasting with SSA

Following the procedure in Figure 10, the supervised filtering (SSA1) using ANN for RCs of Wuxi and India is illustrated in Figure 14. The RMSE associated with the maximum number of p (for instance $p = 5$ for Wuxi) represents the performance of ANN with the normal mode. The optimal p corresponds to the minimum RMSE, which can be found by systematically deleting RCs one at a time. Consequently, numbers of chosen optimal p RCs in three forecasting horizons are 3, 2, and 1 for Wuxi, and 1, 3, and 5 for India, respectively. However, the unsupervised filtering method (SSA2) is based on enumeration of combinations of all RCs. Selection of the optimal p RCs cannot be presented in a graphical form.

Table 8 shows selected p at various prediction horizons using LR, K-NN, and ANN in conjunction with SSA1 and SSA2. A large amount of information can be extracted from this Table. First of all, a considerable improvement in the model performance is achieved by each forecasting model in conjunction with SSA1 or SSA2, compared with results in Table 4. From the perspective of rainfall series, the accuracy of daily rainfall prediction is improved significantly in comparison to that in the normal mode. Secondly, as expected, results from SSA2 are superior to or at least equivalent to those from SSA1 since the former examines each combination of RCs in search of the optimal p. SSA2 is therefore considered as an efficient and effective method if the number L of RCs is small. For the present four cases, SSA2 method is appropriate due to the small number of RCs. Once L is large, say 40 or 50, SSA1 may be a good alternative where a relative optimal forecasting can be guaranteed.

Additionally, it should be noted that the optimal p are different at three forecast horizons. Finally, Table 8 also shows that, among the three models, ANN performs the best with SSA1 or SSA2, which is consistent with results in the normal mode.

In the normal mode, MANN has been proved to be superior to ANN, in particular, for daily rainfall forecasting. As an attempt to improve the accuracy of rainfall forecasting, MANN is also coupled with SSA2. Table 9 demonstrates results in terms of RMSE, CE, and PI using MANN compared with those of ANN. Good accuracies of forecasting are made by both MANN and ANN. It can be seen from values of PI that the prediction lag effect is completely eliminated. The model performance does not deteriorate markedly with the increase of the forecasting lead. Results also show that MANN still maintains a salient superiority over ANN in the SSA2 mode for both daily and month rainfall series.

One-step lead estimates of MANN and ANN with the help of SSA2 are shown in Figure 15 (Wuxi) and Figure 16 (India) in the form of hyetographs and scatter plots (the former is plotted in a selected range for better visual inspection). Compared with Figure 11, each scatter plot in Figure 15 is closer to the exact line, which means that the daily rainfall process is fitted appropriately. Nevertheless, some peak values still remain mismatched although MANN shows a better ability to capture the peak value than ANN. Regarding the monthly rainfall series, the scatter plots with perfect match of the diagonal indicates that the rainfall process is perfectly reproduced. The representative hyetograph shows that the peak times and peak values are also accurately predicted.

Figure 17 presents the correlation analysis between observed and forecasted rainfall from ANN and MANN using the Wuxi and India series, respectively. Compared with Figure

13, the lagged prediction of ANN is completely eliminated by SSA2 since the maximum CCF occurs at zero lag. The larger the CCF at zero lag is, the better the model performance is.

4.5 Discussions

Some discussions regarding forecasting models and the effects of the SSA technique are made in the following.

(1) About the investigation of effects of SSA

Figure 18 shows that a large number of zeros and near zeros occur in the original Wuxi rainfall which makes the series discontinuous. Using the intermittent series is difficult to reconstruct similar input patterns for a forecasting model. Thus, depending on those reconstructed input patterns, data-driven models based on pattern training, for example, ANNs, tend to be unfeasible. In contrast, rainfall series preprocessed by SSA becomes smoother where most of zeros are replaced by nonzero values. New input vectors from the reconstructed rainfall series are characterized by better repetition of patterns so that they are easier reproduced.

To investigate the influence of SSA on the ANN's performance, correlation analyses between inputs and output of ANN and ANN-SSA2 are compared using the Wuxi data and are depicted in Figure 19. As the input and output series in ANN are both the raw rainfall series, the cross-correlation analysis is equivalent to the autocorrelation analysis of the raw rainfall series. At all three prediction horizons, cross-correlation coefficients (CCs) between reconstructed inputs by SSA2 and the raw rainfall data are improved significantly at most lags except for the lag of 3 at one-step lead. It should be recalled that model inputs are the previous five rainfall data for Wuxi. The "starting point" in Figure 19 represents the first previous rainfall of the five inputs, and the remaining inputs consist of four points after the starting point. It can be observed that the CCF value between each new model input and output are far larger than that between the raw model input and output (seen at the same lag). Therefore, the improvement of a model's performance by the SSA technique may be owing to the enhancement of the mapping relation of model input and output by deleting noises hidden in the raw signals.

(2) About parameter L in SSA

The parameter L *in SSA* has a significant impact on the performance of a forecasting model since the optimal p RCs may be different with the change of L when using the same forecasting model at the same prediction horizon. The selection of L in this study is based on the interval of [3, 10] in conjunction with an empirical criterion (namely, a particular L is selected only if the singular spectrum can be distinguished markedly under that L). To check the robustness of the empirical method, each L in [3, 10] is examined by the LR model with SSA2 using the Wuxi and Zhenwan data and presented in Table 10. As mentioned previously, the target L for Wuxi and Zhenwan are 5 and 7, respectively. The RMSE associated with them at each prediction horizon is highlighted in bold (shown in Table 10). In terms of Wuxi, the difference between the target RMSE and the minimum RMSE at the same prediction horizon is only 9.2% for one-step prediction, 1.4% for two-step prediction, 0.0% for three-step prediction, respectively. Regarding Zhenwan, the three values are respectively 5.2%, 7.5%, and 0.0%. These changes are slight and cannot influence the conclusions drawn previously. Therefore, the empirical method for the present rainfall data should be appropriate.

5. Conclusion

This study suggests the use of modular artificial neural network (MANN) coupled with data preprocessing techniques for improving four rainfall predictions from India and China consisting of two monthly and two daily series. To reasonably evaluate MANN's performance, three models, LR, K-NN and ANN, are used for the purpose of comparison. In the process of model development, model inputs and data preprocessing techniques are carefully analyzed and discussed. The following conclusions are reached based on this study:

1. LCA is regarded as an effective and efficient method among all seven input techniques due to its simplicity of computation and comparable capability of forecasting.

2. In the normal mode (without data preprocessing), MANN distinguishes from the other three models for both monthly and daily rainfall series forecasting. Whilst all four models reasonably forecast two monthly rainfall series, only MANN is able to simulate each daily rainfall series without obvious lag effect.

3. In the data preprocessing mode, the effect of MA is negligible for the improvement of each forecasting model.

4. PCA as a data preprocessing technique is discussed in two forms, i.e. PCA1 for the purpose of dimension reduction, and PCA2 for the purpose of noise reduction. Results show that PCA1 cannot improve model's performance and PCA2 marginally improve model's performance.

5. Two filtering method, i.e. supervised (SSA1) and unsupervised (SSA2), are examined for SSA when coupled with forecasting models. It can be found that each model achieves considerable improvement in performance with the aid of SSA1 or SSA2. In terms of forecasting models, MANN still outperforms all other models.

6. As far as two filtering methods are concerned, SSA2 tends to be better if the number of raw RCs is small. Otherwise, SSA1 is a good alternative.

7. A further discussion reveals that the essence of SSA in improving model performance is to strengthen the mapping relation of model input and output by deleting noises in the raw signal.

8. There is still considerable room for improving forecasting of peak values although MANN coupled with SSA has made perfect overall predictions for daily rainfall series.

Nomenclature

ACF	Auto Correlation Function
AMI	Average Mutual Information
ANN	Artificial Neural Networks
CC	Cross-correlation Coefficient
CCF	Cross Correlation Function
CE	Coefficient of Efficiency
CI	Correlation Integral
FCM	Fuzzy C-means
FNN	False Nearest Neighbor
K-NN	K-Nearest-Neighbors
LCA	Linear Correlation Analysis
L-M	Levenberg-Marquart
LR	Linear Regression
MA	Moving Average

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

MANN	Modular Artificial Neural Networks
MOGA	ANN based on Multi-objective Genetic Algorithm
PACF	Partial Auto Correlation Function
PC	Principal Component
PCA	Principal Component Analysis
PMI	Partial Mutual Information
PI	Persistence Index
RC	Reconstructed Component
RMSE	Root Mean Square Error
SLR	Stepwise Linear Regression
SSA	Singular Spectrum Analysis
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vectors Regression
WA	Wavelet Analysis

References:

Abrahart, R.J. and See, L.M. (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.

Baratta et al., Baratta, D., Cicioni, G., Masulli, F. and Studer, L. (2003), Application of an ensemble technique based on singular spectrum analysis to daily rainfall forecasting. Neural Networks, 16, 375-387.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Bezdek, J.C. (1981), Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York.

Brath, A., Montanari, A., and Toth, E. (2002), Neural networks and non-parametric methods for improving realtime flood forecasting through conceptual hydrological models. Hydrology and Earth System Sciences, 6(4), 627-640.

Campolo, M., Andreussi, P. and Soldati, A. (2003), Artificial neural network approach to flood forecasting in the river Arno. Hydrological Sciences Journal, 48(3), 381-398.

Cannas, B., Fanni, A., Pintus, M., and Sechi, G. M. (2002), Neural Network Models to Forecast Hydrological Risk. IEEE IJCNN, 2002.

Chan, J.C.L., and Shi, J.E. (1999), Prediction of the summer monsoon rainfall over South China. International Journal of Climatology, 19 (11), 1255-1265.

Chattopadhyay,S., and Chattopadhyay, G. (2007), Identification of the best hidden layer size for three-layered neural net in predicting monsoon rainfall in India. Journal of Hydroinformatics, 10(2), 181-188.

Cheng, B., and Titterington, D. M. (1994), Titterington Neural networks: A review from a statistical perspective. Statistical Science, 9(1), 2-54.

Chu, P.S., and He, Y.X. (1994), Long-Range Prediction of Hawaiian Winter Rainfall Using Canonical Correlation-Analysis). International Journal of Climatology, 14(6), 659-669.

Corzo, G., and Solomatine, D. (2007), Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. Hydrological Sciences–Journal–des Sciences Hydrologiques, 52(3), 491-507.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Coulibaly, P., Haché, M., Fortin, V., and Bobée, B. (2005), Improving daily reservoir inflow forecasts with model combination. Journal of Hydrologic Engineering, 10(2), 91-99.

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

DelSole, T., and Shukla, J. (2002), Linear prediction of Indian monsoon rainfall. Journal of Climate, 15 (24), 3645-3658.

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.

Diomede, T., Davolio, S., Marsigli, C., Miglietta, M. M., Moscatello, A., Papetti, P., Paccagnella, T., Buzzi, A., and Malguzzi, P. (2008), Discharge prediction based on multi-model precipitation forecasts. Meteorology and Atmospheric Physics, 101 (3-4), 245-265.

Draper, N. R. and Smith, H. (1998), Applied regression analysis, 3rd ed. New York: Wiley.

Fortin, V., Ouarda, T.B.M.J., and Bobe´e, B., (1997), Comment on 'The use of artificial neural networks for the prediction of water quality parameters' by H.R. Maier and G.C. Dandy. Water Resources Research, 33 (10), 2423-2242.

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

Ganguly, A.R., and Bras, R.L. (2003), Distributed quantitative precipitation forecasting (DQPF) using information from radar and numerical weather prediction models. Journal of Hydrometeorology, 4 (6), 1168-1180.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Giustolisi, O., and Savic, D. A. (2006), A symbolic data-driven technique based on evolutionary polynomial regression, Journal of Hydroinformatics, 8(3), 207-222.

Golyandina, N. Nekrutkin,V., and Zhigljavsky, A. (2001), Analysis of Time Series Structure: SSA and Related Techniques, Chapman & Hall/CRC.

Guhathakurta, P. (2008), Long lead monsoon rainfall prediction for meteorological sub-divisions of India using deterministic artificial neural network model. Meteorology and Atmospheric Physics, 101 (1-2), 93-108.

Guhathakurta, P., Rajeevan, M., and Thapliyal, V. (1999), Long range forecasting Indian summer monsoon rainfall by a hybrid Principal Component neural network model. Meteorology and atmospheric physics, 71, (3-4), 255-266.

Hsu, K.L., Gupta, H.V., Gao, X.G., Sorooshian, S., and Imam, B. (2002), Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis. Water Resources Research, 38(12), 1302,doi:10.1029/2001WR000795.

Hu, T.S., Lam, K.C., Ng, S.T., (2001), River flow time series prediction with a range-dependent neural network. Hydrological Science Journal, 46 (5), 729-745.

Hotelling, H., (1933), Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24, 417-441.

Hu, T.S.,Wu, F.Y., and Zhang, X. (2007), Rainfall-runoff modeling using principal component analysis and neural network. Nordic Hydrology, 38(3), 235-248.

Jain, A., and Srinivasulu, S. (2006), Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. Journal of Hydrology, 317, 291-306.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Jayawardena, A.W., and Lai, F. (1994), Analysis and prediction of chaos in rainfall and stream flow time series. Journal of hydrology, 153, 23-52.

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

Kim, T., Heo, J. H. and Jeong, C. S. (2006), Multireservoir system optimization in the Han River basin using multi-objective genetic algorithms. Hydrological Processes, 20 (9), 2057-2075.

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

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., and McCabe, Jr, G. J. (1999), Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation, Water Resources. Research, 35(1), 233-241.

Li, F., and Zeng, Q.C. (2008), Statistical prediction of East Asian summer monsoon rainfall based on SST and sea ice concentration. Journal of the meteorological society of Japan, 86 (1), 237-243.

Lin, G.F., and Chen, L.H. (2005), Application of artificial neural network to typhoon rainfall forecasting. Hydrological Processes, 19 (9), 1825-1837.

Lin, G.F., Chen, G.R., Wu, M.C., and Chou, Y.C. (2009), Effective forecasting of hourly typhoon rainfall using support vector machines. Water Resources Research, 45, W08440.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

---

doi:10.1029/2009WR007911.

Lin, G.F., and Wu, M.C. (2009), A hybrid neural network model for typhoon-rainfall forecasting. Journal of Hydrology, 375 (3-4), 450-458.

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.

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.

May, R.J., Maier, H.R., Dandy, G.C., and Fernando, T.M.K. (2008), Non-linear variable selection for artificial neural networks using partial mutual information. Environmental Modeling & Software, 23, 1312-1328.

McCuen, R. H. (2005), Hydrologic analysis and design (3rd ed.), Upper Saddle River, NJ: Pearson/Prentice Hall.

Munot, A. A., and Kumar, K.K. (2007), Long range prediction of Indian summer monsoon rainfall. Journal of Earth System Science, 116 (1), 73-79.

Nash, J. E. and Sutcliffe, J. V. (1970), River flow forecasting through conceptual models part I — A discussion of principles. Journal of Hydrology, 10 (3), 282-290.

Nayagam, L.R., Janardanan, R., and Mohan, H.S.R. (2008), An empirical model for the seasonal prediction of southwest monsoon rainfall over Kerala, a meteorological subdivision of India. International Journal of Climatology, 28 (6), 823-831.

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

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

Pearson, K., (1901), On lines and planes of closest fit to systems of points in space. Philosophical Magazine, 2, 559-572.

Pongracz, R., Bartholy, J., and Bogardi, I. (2001), Fuzzy rule-based prediction of monthly precipitation. Physics and Chemistry of the Earth Part B-Hydrology Oceans and Atmosphere, 26 (9), 663-667.

Rajurkar, M.P., Kothyari, U.C., and Chaube, U.C. (2002), Artificial neural networks for daily rainfall-runoff modeling. Hydrological Sciences-Journal, 47(6), 865-876.

Salas, J.D., Delleur, J.W., Yevjevich, V., and Lane, W.L. (eds) (1985), Applied Modeling of Hydrologic Time Series. Water Resources Publications: Littleton, Colorado.

See, L., and Openshaw, S. (2000), A hybrid multi-model approach to river level forecasting. Hydrology Science Journal, 45(4), 523-536.

Shamseldin, A.Y. and O’Connor, K.M. (1999), A real-time combination method for the outputs of different rainfall-runoff models. Hydrological Sciences Journal, 44(6), 895-912.

Shamseldin, A.Y., O'Connor, K.M. and Liang, G.C., (1997), Methods for combining the outputs of different rainfall runoff models, Journal of Hydrology, 197, 203–229.

Sheng, C., Gao, S., and Xue, M. (2006), Short-range prediction of a heavy precipitation event by assimilating Chinese CINRAD-SA radar reflectivity data using complex cloud analysis. Meteorology and Atmospheric Physics, 94 (1-4), 167-183.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Shrestha, D.L., and Solomatine, D.P. (2006), Machine learning approaches for estimation of prediction interval for the model output. Neural Networks, 19(2), 225-235.

Silverman, D., and Dracup, J.A. (2000), Artificial neural networks and long-range precipitation in California. Journal of Applied Meteorology 31 (1), 57-66.

Sivapragasam, C., and Liong, S. Y. (2005), Flow categorization model for improving forecasting. Nordic Hydrology 36 (1), 37-48.

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

Sivapragasam, C., Vincent and, P., and Vasudevan, G. (2007), Genetic programming model for forecast of short and noisy Data. Hydrological Processes, 21, 266-272.

Solomatine, D. P., and Ostfeld, A. (2008), Data-driven modelling: Some past experiences and new approaches. Journal of Hydroinformatics, 10(1), 3-22.

Solomatine, D. P. and Xue, Y. I. (2004), M5 model trees and neural networks: application to flood forecasting in the upper reach of the Huai River in China. Journal of Hydrological Engineering, 9(6), 491-501.

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

Sugihara, G and May, R.M. (1990), Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature, 344, 734- 741.

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

Theiler, J., (1986), Spurious dimension from correlation algorithms applied to limited time-series data. Physical Review A, 34 (3), 2427-2432.

Toth, E, Brath, A., and Montanari, A. (2000), Comparison of short-term rainfall prediction models for real-time flood forecasting. Journal of Hydrology, 239, 132-147.

Venkatesan, C., Raskar, S.D., Tambe, S.S., Kulkarni, B.D., and Keshavamurty, R.N. (1997), Prediction of all India summer monsoon rainfall using error-back-propagation neural networks. Meteorology and Atmospheric Physics, 62 (3-4), 225-240.

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.

Wu, C.L., Chau, K.W., and Li, Y.S., (2009), Methods to improve neural network performance in daily flows prediction. Journal of Hydrology, 372(1-4), 80-93.

Wu, C.L., Chau, K.W. and Li, Y.S. (2008), River stage prediction based on a distributed support vector regression. Journal of Hydrology, 358, 96-111.

Xiong, L.H., Shamseldin, A.Y., and O'Connor, K.M. (2001), A non-linear combination of forecasts of rainfall-runoff models by the first-order TS fussy system for forecast of rainfall-runoff model. Journal of Hydrology, 245, 196-217.

Yates, D.N., Warner, T.T., and Leavesley, G.H. (2000), Prediction of a flash flood in complex terrain. Part II: A comparison of flood discharge simulations using rainfall input from radar, a dynamic model, and an automated algorithmic system. Journal of Applied Meteorology, 39 (6), 815-825.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Yu, D.L., Gomm, J.B., and Williams, D. (2000), Neural model input selection for a MIMO chemical process. Engineering Applications of Artificial Intelligence, 13(1), 15-23.

Zhang, B., and Govindaraju, R. S. (2000), Prediction of watershed runoff using Bayesian concepts and modular neural networks. Water Resources Research, 36(3), 753-762.

Journal of Hydrology, Vol. 389, No. 1-2, 2010, pp. 146-167

Figure Captions

Figure1. Location of Daning river basin (Map of Chongqing in the left panel, and Daning watershed in the dashed box)

Figure 2. Rainfall series of (a) Wuxi, (b) India, (c) Zhenwan, and (d) Zhongxian

Figure 3. Flow chart of hard forecasting using a modular model

Figure 4. Implementation framework of forecast models with/without data preprocessing

Figure 5. Plots of ACF and PACF of the rainfall series with the 95% confidence bounds (the dashed lines), (a) and (c) for Wuxi, and (b) and (d) for Zhenwan

Figure 6. Check of robustness of K in KNN method for (a) Wuxi, (b) India, (c) Zhenwan, and (d) Zhongxian.

Figure 7. Singular Spectrum as a function of lag using various window lengths L for (a) Wuxi, (b) India, (c) Zhenwan, and (d) Zhongxian

Figure 8. Sensitivity analysis of singular Spectrum on varied τ for (a) Wuxi, (b) India, (c) Zhenwan, and (d) Zhongxian

Figure 9. Plots of CCF between each RC and the raw rainfall data for Zhenwan

Figure 10. Supervised procedure for a forecast model with SSA

Figure 3. Scatter plots and hyetographs of one-step-ahead forecast using ANN and MANN for Wuxi ((a) and (c) from ANN, and (b) and (d) from MANN)

Figure 4. Scatter plots and hyetographs of one-step-ahead forecast using ANN and MANN for India ((a) and (c) from ANN, and (b) and (d) from MANN)

Figure 5. CCFs at three forecast horizons for various lags in time of observed and forecasted rainfall of ANN and MANN: Wuxi (left column) and India (right column)

Figure 6. Performance of ANN with SSA1 in terms of RMSE as a function of p ($\leq L$)

selected RCs at various prediction horizons: (a) Wuxi and (b) India

Figure 7. Scatter plots and hyetographs of one-step-ahead forecast using ANN and MANN with SSA2 for Wuxi ((a) and (c) from ANN, and (b) and (d) from MANN)

Figure 8. Scatter plots and hyetographs of one-step-ahead forecast using ANN and MANN with SSA2 for India ((a) and (c) from ANN, and (b) and (d) from MANN)

Figure 17. CCFs at three forecast horizons using ANN and MANN with Wuxi and India ((a) and (c) for Wuxi and (b) and (d) for India)

Figure 18. Hyetographs (detail; from 900^{th} to 950^{th}) of Wuxi: (1) the raw rainfall series and (2) reconstructed rainfall series for one-step prediction of ANN.

Figure 19. CCF between inputs and output for ANN and ANN-SSA2 at three prediction horizons using the Wuxi data

Table captions

Table 1. Pertinent information for four watersheds and the rainfall data

Table 2. Comparison of methods to determine mode inputs using ANN model

Table 3. Performance comparison of ANN with different data-transformed methods

Table 4. Model performances at three forecasting horizons under normal mode

Table 5. Model performances of ANN-MA using the Wuxi data

Table 6. Multiple-step predictions for Wuxi and India series using LR, K-NN, and ANN with PCA1

Table 7. Multiple-step predictions for Wuxi and India series using LR, K-NN, and ANN with PCA2

Table 8. Optimal p RCs for model inputs at various forecasting horizons

Table 9. Model performances at three forecasting horizons using MANN and ANN with the SSA2

Table 10. RMSE of the LR model coupled with SSA2 using various L

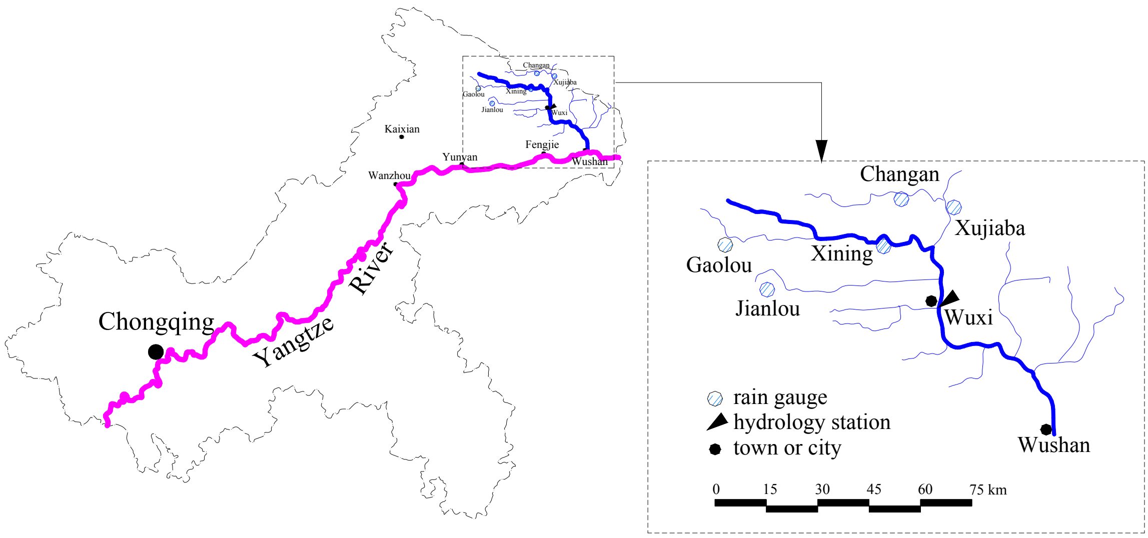

Chongqing
Yangtze
River
Kaixian
Wanzhou
Yunyan
Fengjie
Wushan
Changan
Xujiaba
Gaolou
Xining
Jianlou
Wuxi
Wushan
rain gauge
hydrology station
town or city
0
15
30
45
60
75 km

Figure 1.

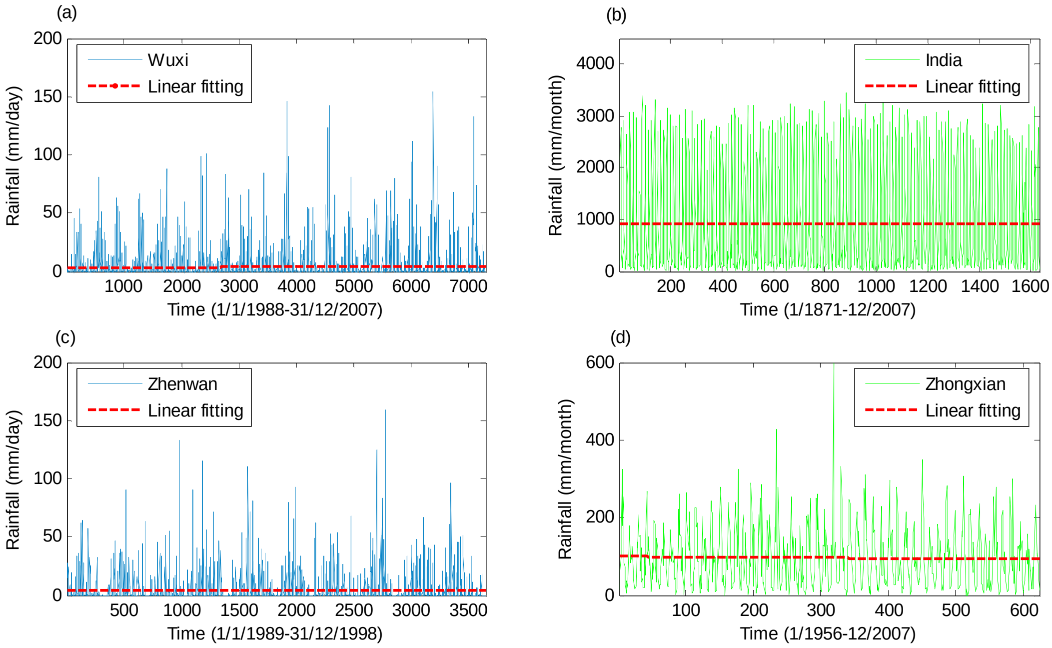

(a)
Wuxi
Linear fitting
Rainfall (mm/day)
Time (1/1/1988-31/12/2007)
(b)
India
Linear fitting
Rainfall (mm/month)
Time (1/1871-12/2007)
(c)
Zhenwan
Linear fitting
Rainfall (mm/day)
Time (1/1/1989-31/12/1998)
(d)
Zhongxian
Linear fitting
Rainfall (mm/month)
Time (1/1956-12/2007)

Figure 2.

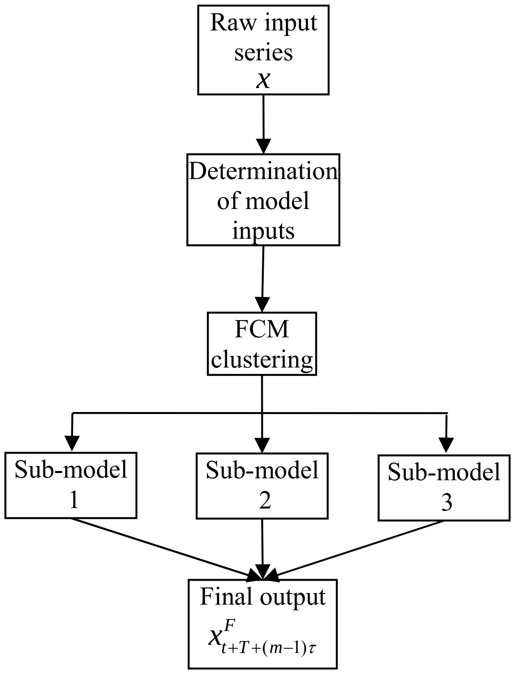

Raw input series x
Determination of model inputs
FCM clustering
Sub-model 1
Sub-model 2
Sub-model 3
Final output $x^F_{t+T+(m-1)\tau}$

Figure 3.

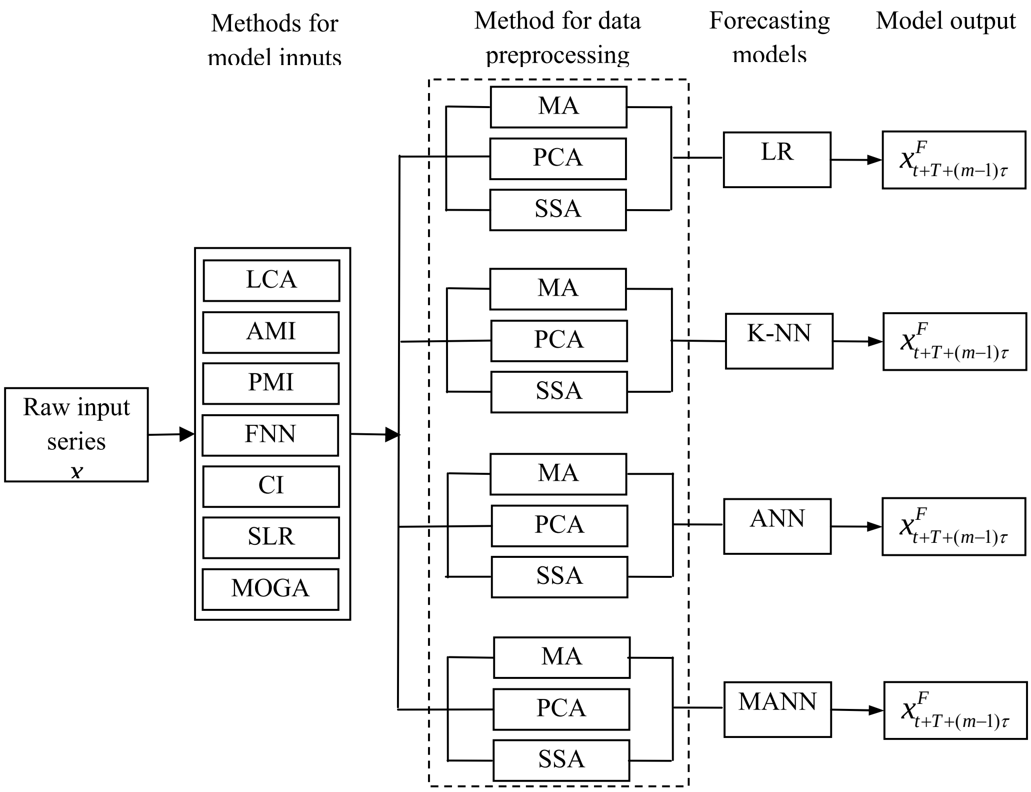
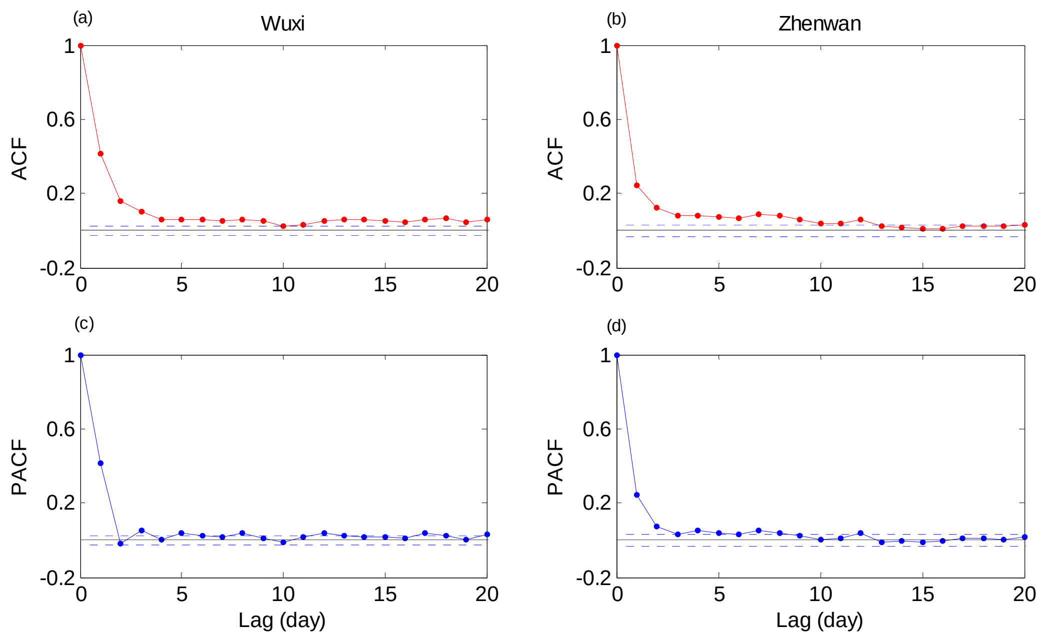

Methods for model inputs
Method for data preprocessing
Forecasting models
Model output
Raw input series
x
LCA
AMI
PMI
FNN
CI
SLR
MOGA
MA
PCA
SSA
LR
K-NN
ANN
MANN
$x^F_{t+T+(m-1)\tau}$

Figure 4.

1010

1011



1012

1013 Figure 5.

1014

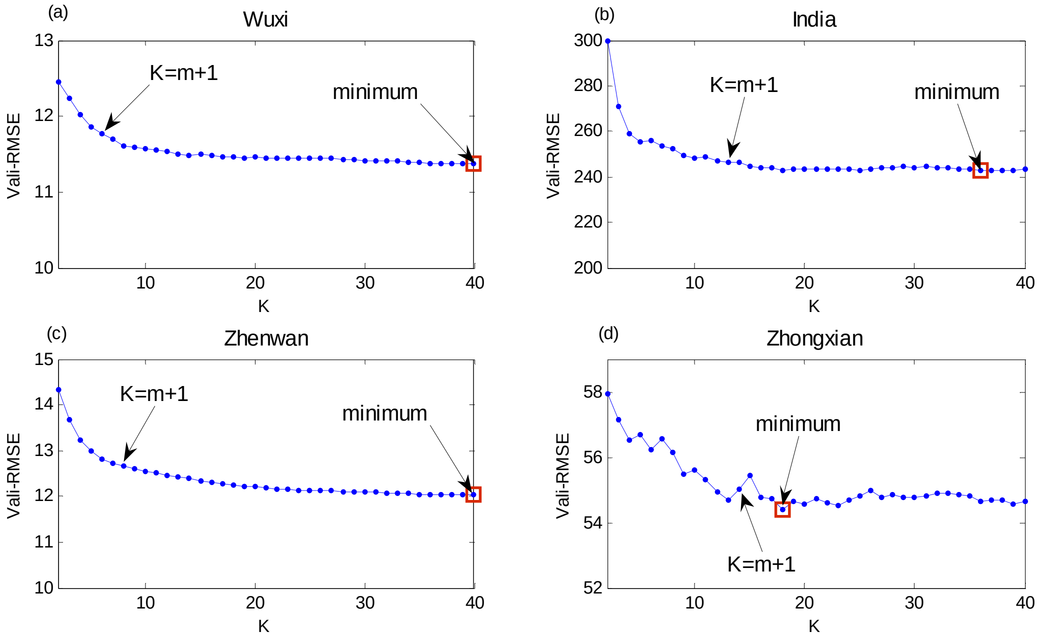
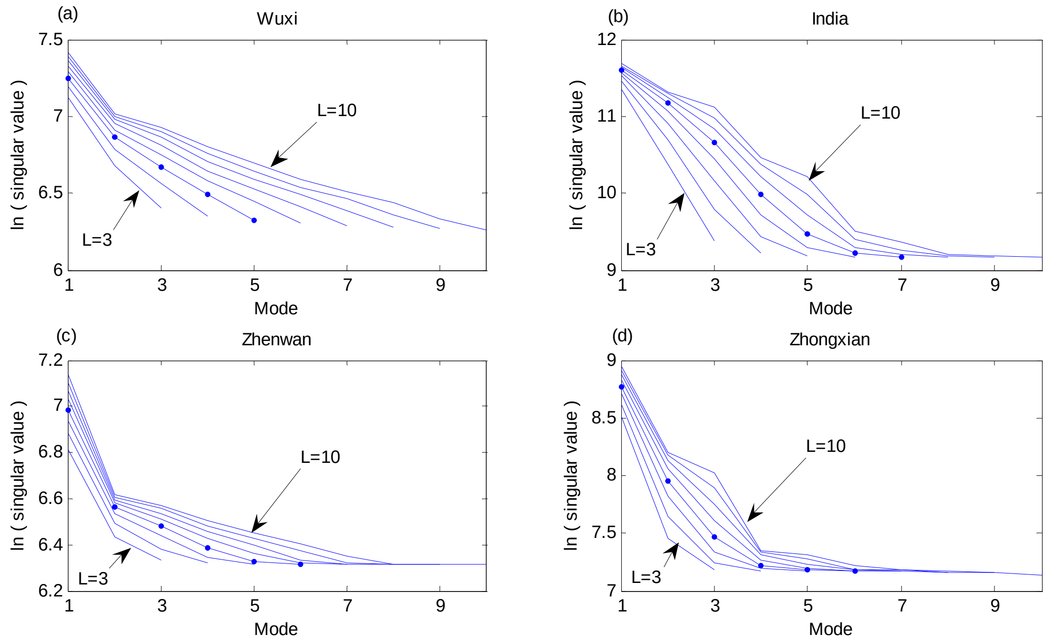

(a)
Wuxi
Vali-RMSE
K=m+1
minimum
K
(b)
India
Vali-RMSE
K=m+1
minimum
K
(c)
Zhenwan
Vali-RMSE
K=m+1
minimum
K
(d)
Zhongxian
Vali-RMSE
minimum
K=m+1
K

Figure 6.



1019

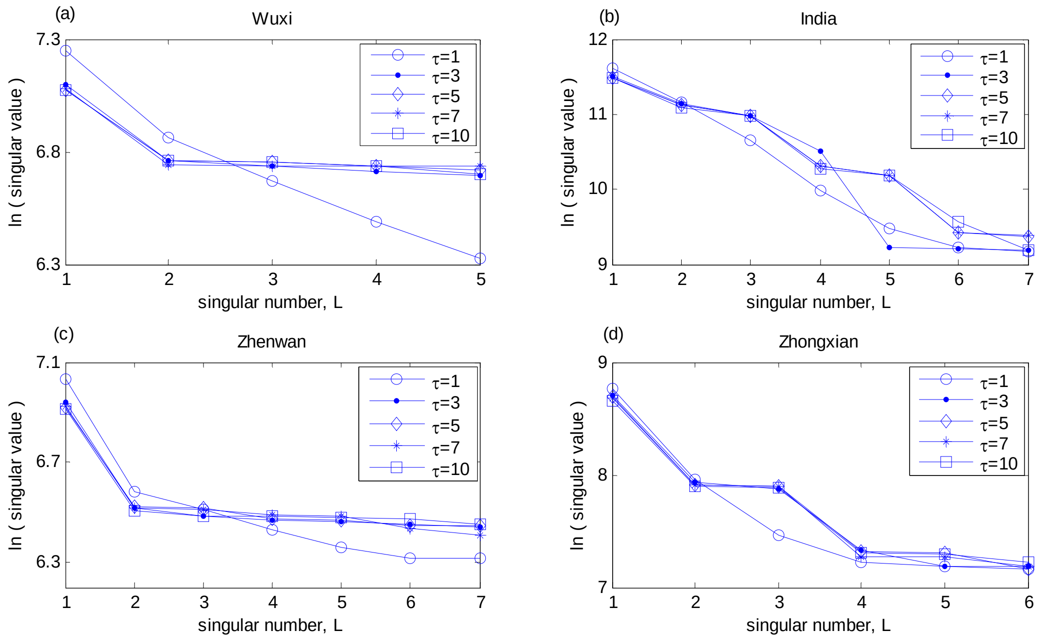
1020 Figure 7.

1021

1022

1023

1024



1025

1026 Figure 8.

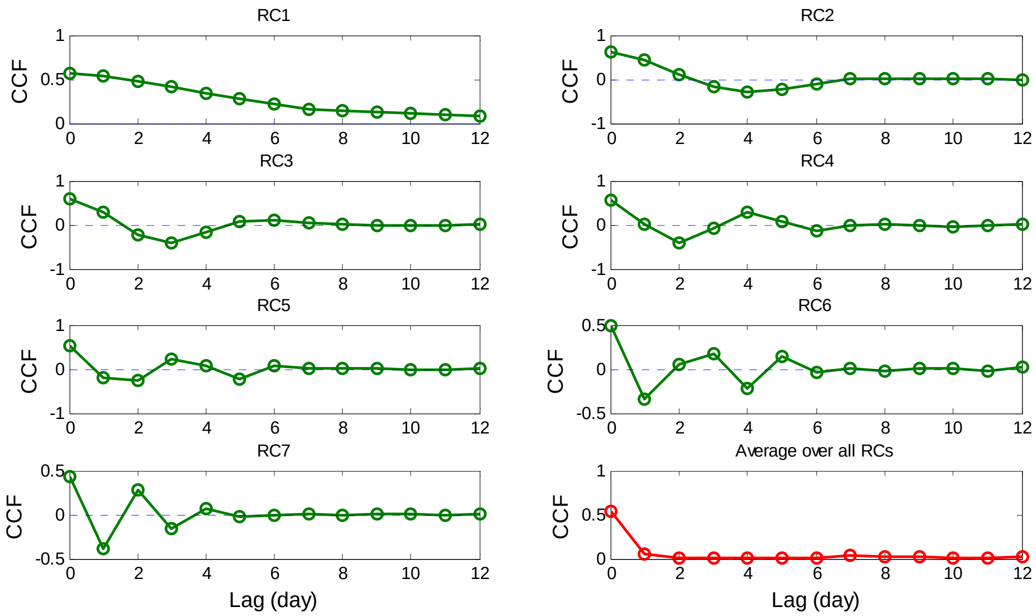


Figure 9.

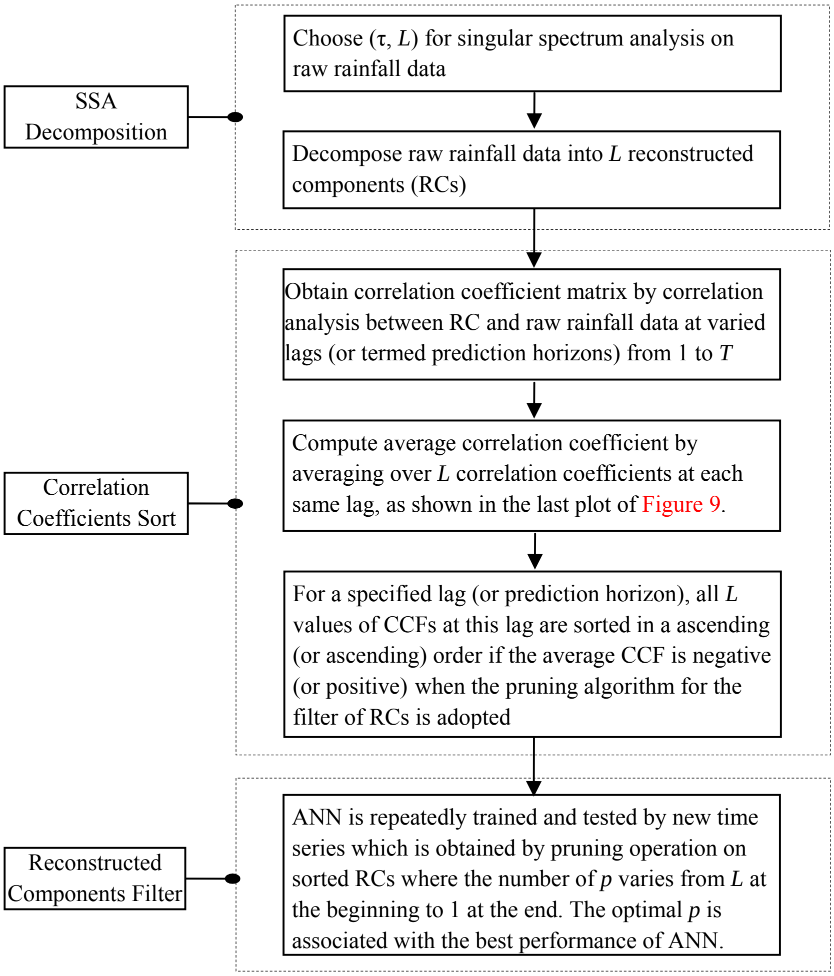
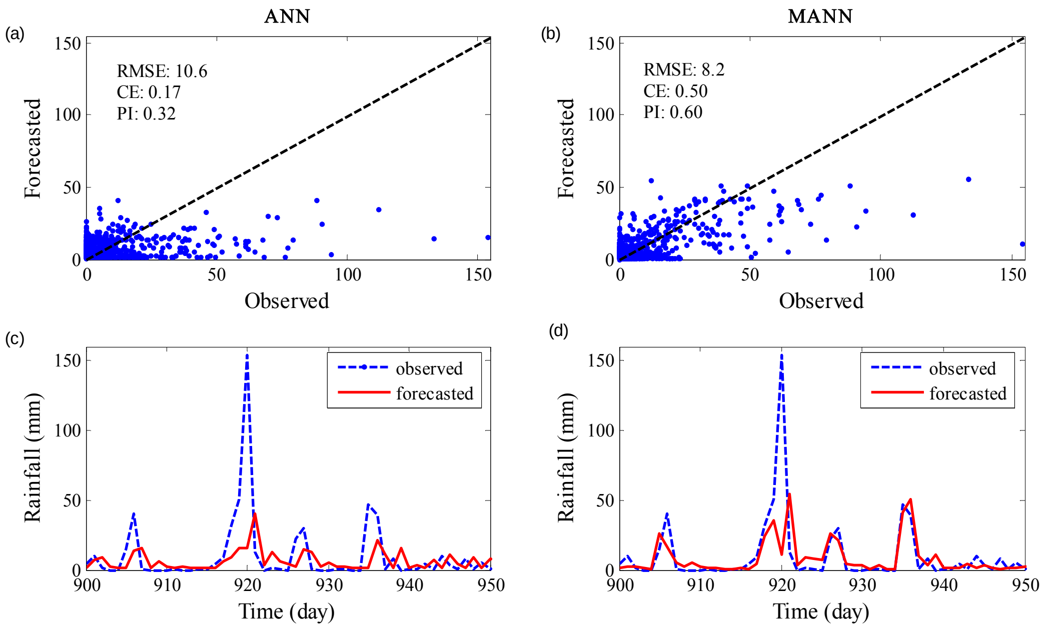

SSA Decomposition
Choose (τ, *L*) for singular spectrum analysis on raw rainfall data
Decompose raw rainfall data into *L* reconstructed components (RCs)
Correlation Coefficients Sort
Obtain correlation coefficient matrix by correlation analysis between RC and raw rainfall data at varied lags (or termed prediction horizons) from 1 to *T*
Compute average correlation coefficient by averaging over *L* correlation coefficients at each same lag, as shown in the last plot of Figure 9.
For a specified lag (or prediction horizon), all *L* values of CCFs at this lag are sorted in a ascending (or ascending) order if the average CCF is negative (or positive) when the pruning algorithm for the filter of RCs is adopted
Reconstructed Components Filter
ANN is repeatedly trained and tested by new time series which is obtained by pruning operation on sorted RCs where the number of *p* varies from *L* at the beginning to 1 at the end. The optimal *p* is associated with the best performance of ANN.

Figure 10.

1035



1036

1037 Figure 11.

1038

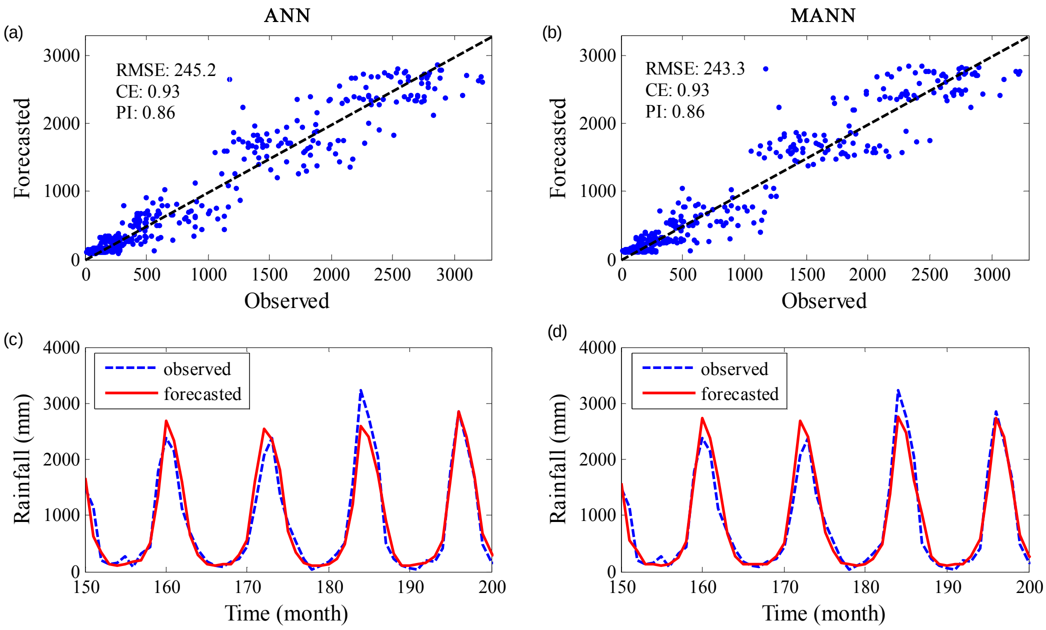
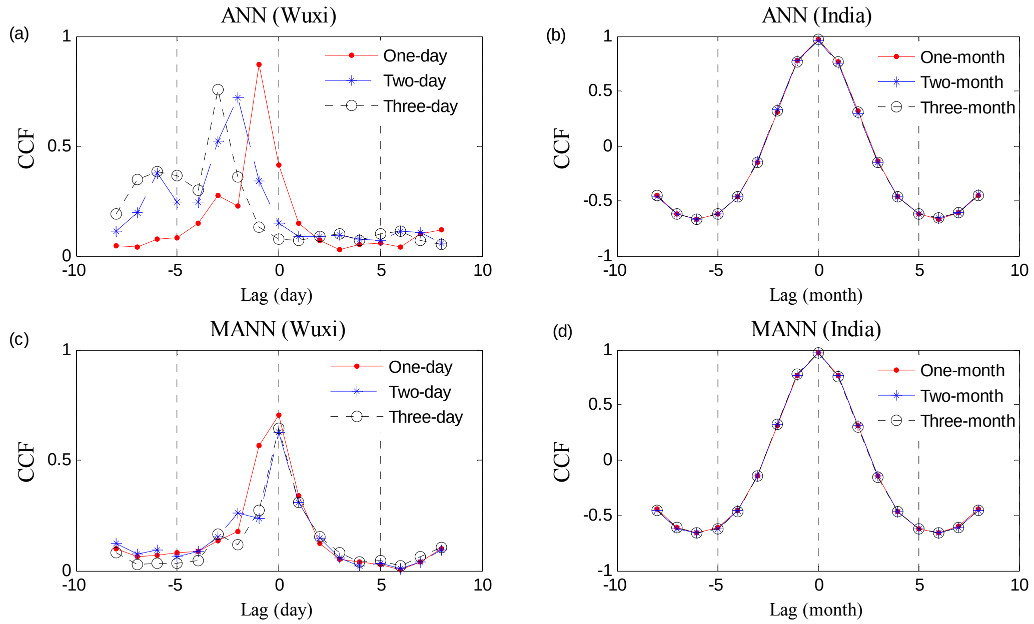

(a)
ANN
RMSE: 245.2
CE: 0.93
PI: 0.86
Forecasted
Observed
(b)
MANN
RMSE: 243.3
CE: 0.93
PI: 0.86
Forecasted
Observed
(c)
observed
forecasted
Rainfall (mm)
Time (month)
(d)
observed
forecasted
Rainfall (mm)
Time (month)

Figure 12.

1045



1046

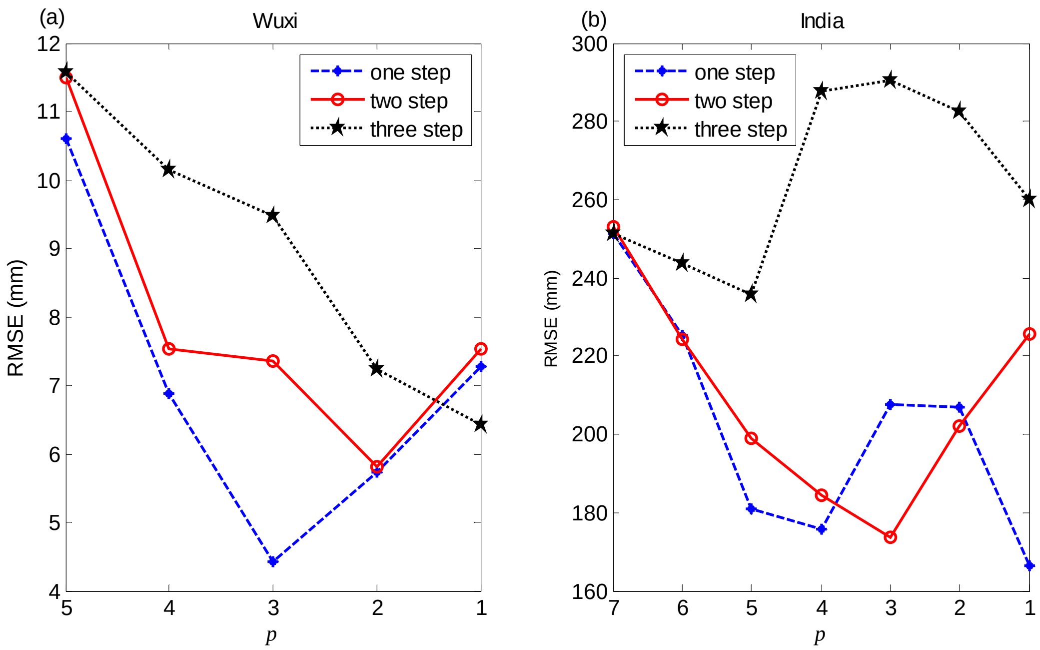
1047 Figure 13.

1048

1049

1050

1051



1052

1053 Figure 14.

1054

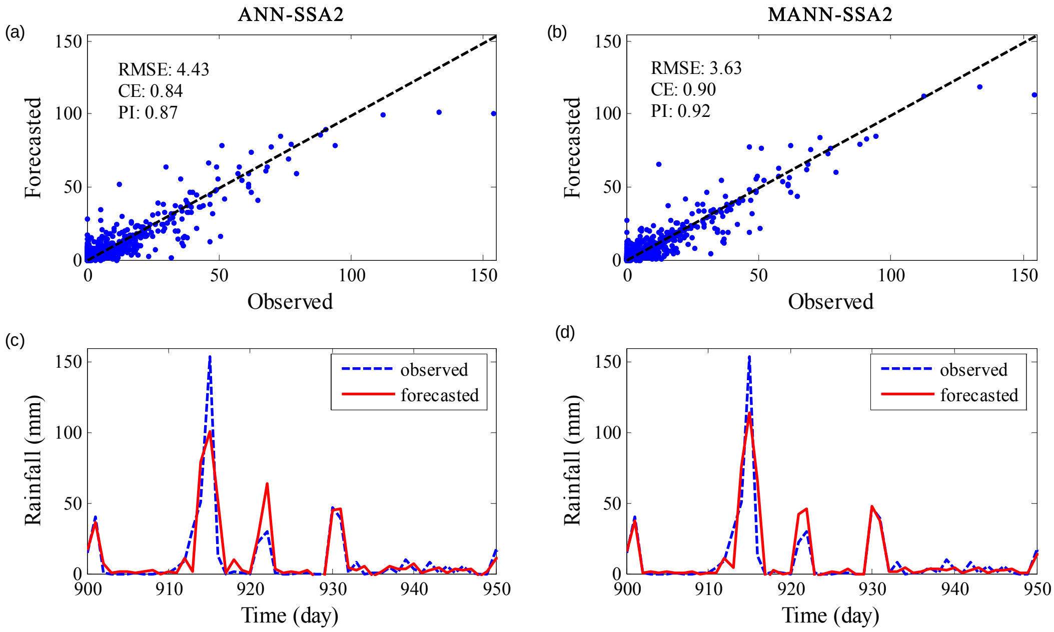

ANN-SSA2
MANN-SSA2
(a)
(b)
RMSE: 4.43
CE: 0.84
PI: 0.87
RMSE: 3.63
CE: 0.90
PI: 0.92
Forecasted
Observed
(c)
(d)
observed
forecasted
Rainfall (mm)
Time (day)

Figure 15.

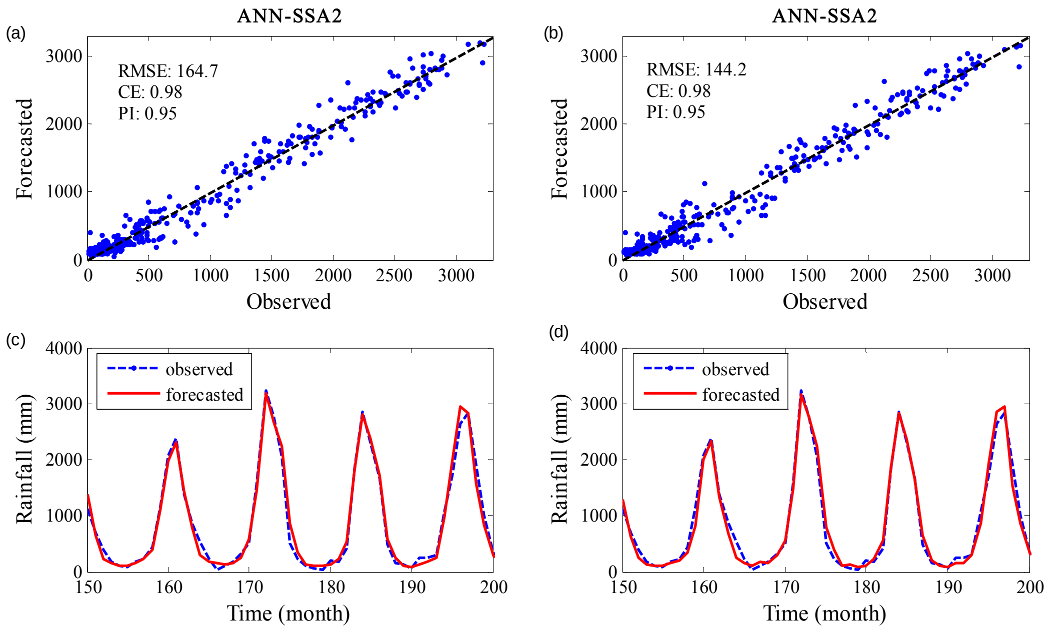
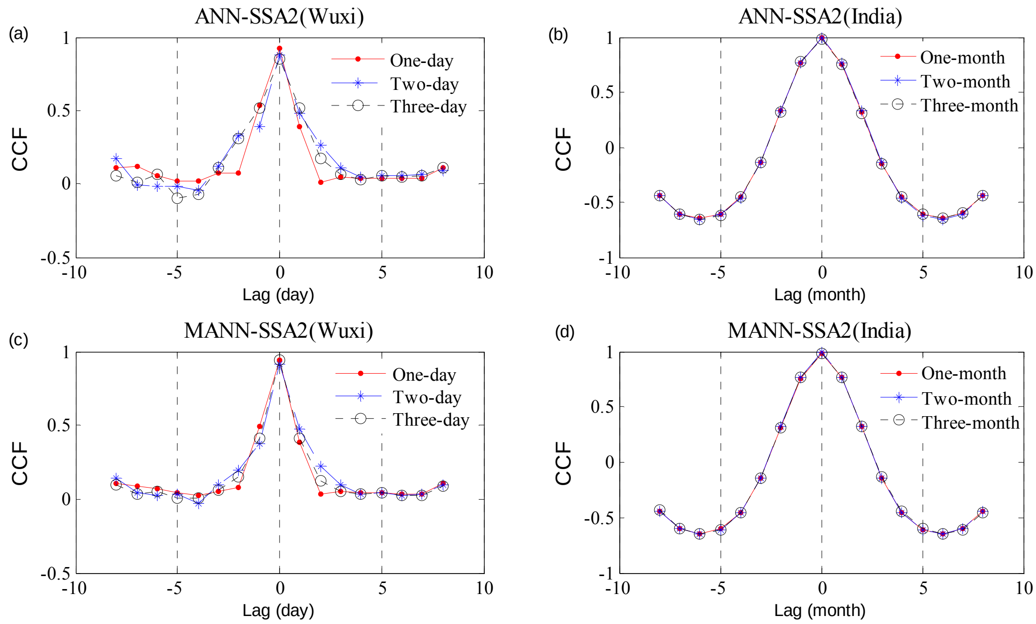

(a)
ANN-SSA2
RMSE: 164.7
CE: 0.98
PI: 0.95
Forecasted
Observed
(b)
ANN-SSA2
RMSE: 144.2
CE: 0.98
PI: 0.95
Forecasted
Observed
(c)
observed
forecasted
Rainfall (mm)
Time (month)
(d)
observed
forecasted
Rainfall (mm)
Time (month)

Figure 16.

1064

1065

1066



1067

1068 Figure 17.

1069

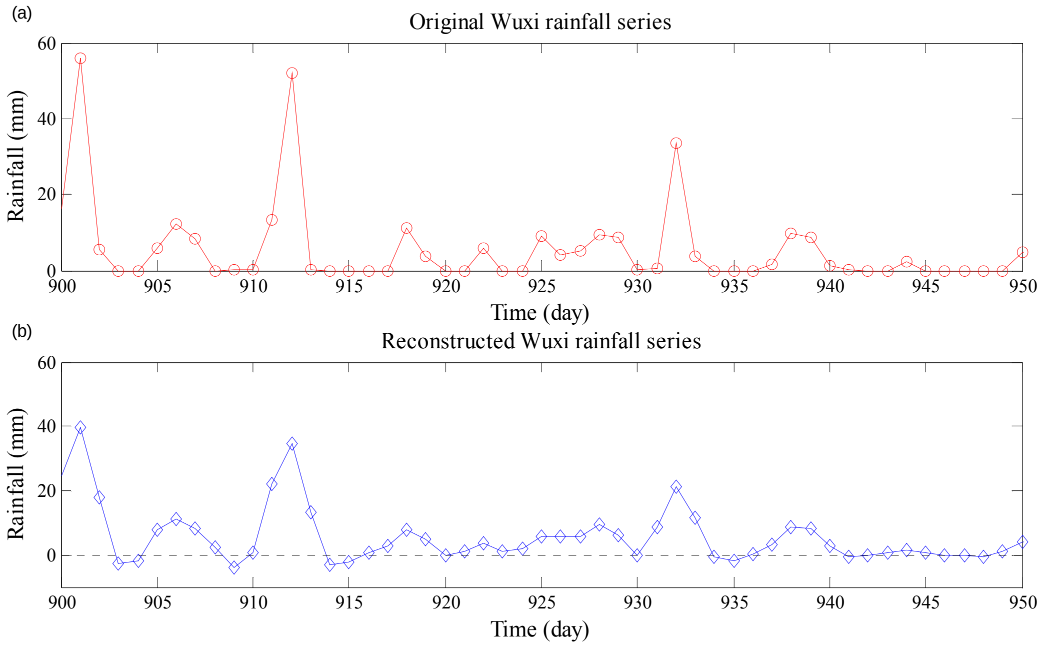
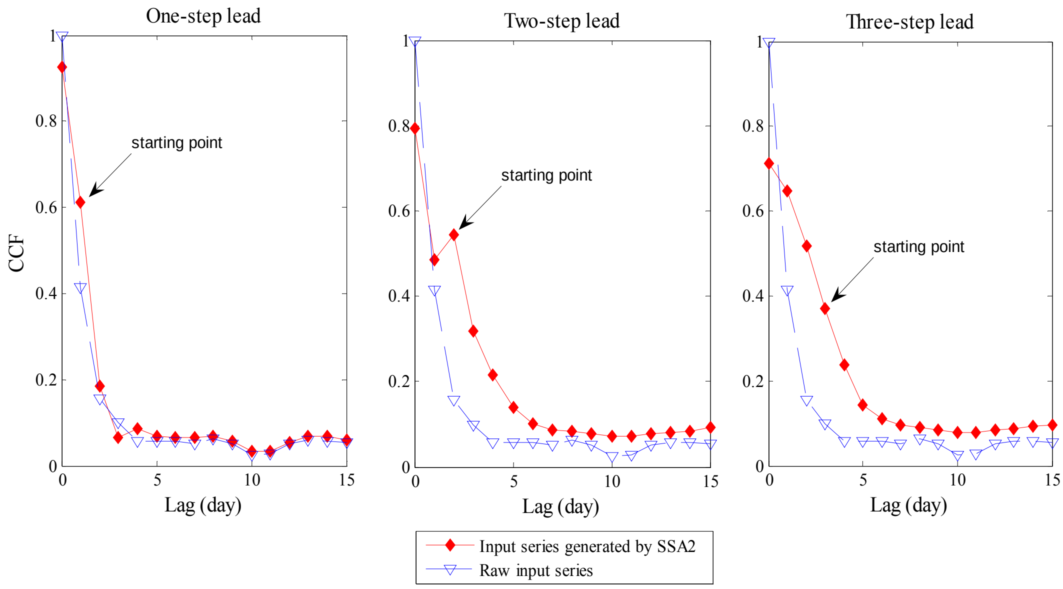

(a)
Original Wuxi rainfall series
Rainfall (mm)
Time (day)
(b)
Reconstructed Wuxi rainfall series
Rainfall (mm)
Time (day)

Figure 18.

1075



1076

1077 Figure 19.

1078

Table 11. Pertinent information for four watersheds and the rainfall data

Watershed and datasets	Statistical parameters						Watershed area and data period
	μ (mm)	S_x (mm)	C_v	C_s	X_{min} (mm)	X_{max} (mm)	
Wuxi							
Original data	3.67	10.15	0.36	5.68	0.00	154	Area:
Training	3.81	10.94	0.35	6.27	0.00	147	2 000 km^2
Cross-validation	3.42	8.87	0.39	4.96	0.00	102	Data period:
Testing	4.03	11.60	0.35	5.46	0.00	154	Jan., 1988- Dec., 2007
Zhenwan							
Original data	4.3	11.0	0.39	4.94	0.0	159	Area:
Training	4.3	11.2	0.38	5.60	0.0	159	7554 km^2
Cross-validation	4.7	11.2	0.42	4.22	0.0	125	Data period:
Testing	4.0	10.9	0.37	4.97	0.0	133	Jan., 1989- Dec., 1998
India							
Original data	906.7	951.6	1.0	0.9	3.0	3460	Area:
Training	904.8	955.7	0.9	0.9	3.0	3393	all India
Cross-validation	918.2	969.5	0.9	1.0	8.0	3460	Data period:
Testing	898.9	927.4	1.0	0.9	16.0	3232	Jan., 1871- Dec., 2007
Zhongxian							
Original data	96.2	79.2	1.2	1.2	0.0	599	Area:
Training	97.2	77.5	1.3	0.9	0.0	429	
Cross-validation	98.6	86.8	1.1	1.9	0.0	599	Data period:
Testing	91.8	74.9	1.2	0.8	0.0	306	Jan., 1956- Dec., 2007

Table 12. Comparison of methods to determine mode inputs using ANN model

Watershed	Methods	τ	m	Effective inputs[a]	Identified ANN	RMSE
Wuxi						
	LCA	1	20	The last 5	(5-5-1)	10.74
	AMI	1	12	Except for Xt-10,t-9	(10-3-1)	10.91
	PMI[b]	1	12	Xt,t-1,t-3,t-5,t-7,t-10,t-4	(7-8-1)	10.85
	FNN	1	20	The last 14	(14-3-1)	11.02
	CI	4	20	Nil		
	SLR	1	12	Xt-11,t-7,t-4,t-2,t-1,t	(6-3-1)	10.94
	MOGA	1	12	Xt, t-1	(2-6-1)	10.55
Zhenwan						
	LCA	1	20	The last 7	(7-4-1)	11.03
	AMI	1	12	Except for Xt-11,t-10,t-9,t-8,t-2	(7-5-1)	10.95
	PMI	1	12	Xt-4,t,t-1,t-3,t-11,t-5,t-10,t-6,t-9	(10-4-1)	10.98
	FNN	1	20	Last 14	(14-3-1)	11.08
	CI	3	20	Nil		
	SLR	1	12	Xt-11,t-7,t-6,t-3,t-1,t	(6-3-1)	11.01
	MOGA	1	12	Xt,t-4,t-7,t-9,t-11	(5-8-1)	10.43
India						
	LCA	1	20	the last 12	(12-5-1)	256.22
	AMI	1	12	the last 12	(12-5-1)	256.22
	PMI	1	12	Xt-11,t-10,t-5,t	(4-5-1)	275.06
	FNN	1	20	the last 5	(5-9-1)	286.04
	CI	4	20	nil		
	SLR	1	12	except for Xt-4	(11-9-1)	258.13
	MOGA	1	12	Xt-11,t-9,t-7,t-5,t-4,t-1,t	(7-1-1)	277.57
Zhongxian						
	LCA	1	20	the last 13	(13-3-1)	51.70
	AMI	1	12	Xt-11,t-10,t-6,t-5,t-4,t	(6-5-1)	54.67
	PMI	1	12	Xt-11,t,t-9,t-7,t-7	(5-9-1)	55.39
	FNN	1	20	the last 4	(4-7-1)	59.78
	CI	3	20	nil		
	SLR	1	12	Xt-11,t-7,t-6,t-5,t-3,t	(6-6-1)	55.47
	MOGA	1	12	Xt-11,t-10,t-6,t-3,t	(5-2-1)	53.93

Note:[a] for the convenience of writing down effective inputs, “Xt, t-1”stands for Xt, Xt-1; [b] effective inputs from PMI are in descending order of priority.

Table 13. Performance comparison of ANN with different data-transformed methods

Watershed	Data Transformation	RMSE			CE		
		1	2	3 [a]	1	2	3
Wuxi							
	Std_raw	10.77	11.54	11.62[b]	0.14	0.01	0.00
	Norm_raw	10.57	11.49	11.59	0.17	0.02	0.00
	Std_nth_root	11.00	12.02	12.10	0.10	-0.07	-0.09
	Norm_nth_root	11.15	12.01	12.09	0.08	-0.07	-0.09
Zhenwan							
	Std_raw	11.03	11.11	11.16	0.03	0.02	0.01
	Norm_raw	10.72	11.06	11.14	0.09	0.03	0.02
	Std_nth_root	11.25	11.68	11.75	-0.01	-0.09	-0.10
	Norm_nth_root	11.34	11.70	11.74	-0.02	-0.09	-0.09
Wuxi							
	Std_raw	256.22	250.51	249.46	0.92	0.93	0.93
	Norm_raw	251.74	246.48	250.99	0.93	0.93	0.93
	Std_nth_root	259.81	253.42	256.43	0.92	0.93	0.92
	Norm_nth_root	252.75	251.95	259.00	0.93	0.93	0.92
Zhongxian							
	Std_raw	54.26	54.23	53.91	0.48	0.48	0.48
	Norm_raw	52.91	53.10	52.78	0.50	0.50	0.51
	Std_nth_root	52.15	53.44	53.17	0.52	0.49	0.50
	Norm_nth_root	52.27	53.37	54.30	0.51	0.49	0.48

[a] Numbers of "1, 2, and 3" denote one-, two-, and three-day-ahead forecasting;
[b] Result is average over 10 best runs from total 20 runs;

Table 14. Model performances at three forecasting horizons under the normal mode

Watershed	Model	RMSE			CE			PI		
		1*	2*	3*	1	2	3	1	2	3
WuXi										
	Naïve	12.2	16.0	16.5	0.05	-0.61	-0.72	0.00	0.00	0.00
	LR	10.9	11.9	12.0	0.12	-0.05	-0.07	0.28	0.41	0.43
	K-NN	11.8	12.4	12.6	-0.03	-0.14	-0.18	0.16	0.36	0.38
	ANN	10.6	11.5	11.6	0.17	0.02	0.00	0.32	0.45	0.47
	MANN	8.2	9.2	9.0	0.50	0.38	0.40	0.60	0.65	0.68
Zhenwan										
	Naïve	12.4	12.2	13.6	-0.53	-0.49	-0.84	0.00	0.00	0.00
	LR	11.1	11.3	11.4	0.02	-0.02	-0.03	0.39	0.42	0.45
	K-NN	12.7	12.7	12.8	-0.27	-0.28	-0.30	0.21	0.28	0.31
	ANN	10.7	11.1	11.1	0.09	0.03	0.02	0.43	0.45	0.48
	MANN	7.9	9.6	9.9	0.50	0.27	0.23	0.69	0.59	0.59
India										
	Naïve	643.1	1084.5	1399.2	0.52	-0.37	-1.28	0.00	0.00	0.00
	LR	286.8	301.6	302.7	0.90	0.89	0.89	0.80	0.92	0.95
	K-NN	246.6	257.3	251.2	0.93	0.92	0.93	0.85	0.94	0.97
	ANN	245.2	245.9	247.2	0.93	0.93	0.93	0.86	0.95	0.97
	MANN	243.3	241.8	244.4	0.93	0.93	0.93	0.86	0.95	0.97
Zhongxian										
	Naïve	75.7	91.9	109.2	-0.03	-0.51	-1.13	0.00	0.00	-0.02
	LR	56.1	57.7	58.4	0.44	0.41	0.39	0.46	0.60	0.71
	K-NN	55.0	56.0	57.2	0.46	0.44	0.42	0.48	0.63	0.72
	ANN	52.5	54.4	54.3	0.51	0.48	0.48	0.52	0.65	0.75
	MANN	50.3	50.2	53.2	0.55	0.55	0.50	0.56	0.70	0.76

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

Table 15. Model performances of ANN-MA using the Wuxi data

Prediction horizons	Performance index	Window length (k) for MA									
		1	2	3	4	5	6	7	8	9	10
One-step											
	RMSE	10.60	10.70	10.63	10.72	10.77	10.70	10.83	10.83	10.72	10.68
	CE	0.17	0.15	0.16	0.15	0.14	0.15	0.13	0.13	0.15	0.15
	PI	0.32	0.31	0.32	0.31	0.30	0.31	0.29	0.29	0.31	0.31
Two-step											
	RMSE	11.50	11.51	11.51	11.50	11.53	11.56	11.55	11.47	11.45	11.45
	CE	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.02	0.03	0.03
	PI	0.45	0.44	0.44	0.45	0.44	0.44	0.44	0.45	0.45	0.45
Three-step											
	RMSE	11.60	11.58	11.58	11.55	11.61	11.58	11.56	11.51	11.52	11.52
	CE	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.02	0.01	0.01
	PI	0.47	0.47	0.47	0.48	0.47	0.47	0.48	0.48	0.48	0.48

Table 6. Multiple-step predictions for Wuxi and India series using LR, K-NN, and ANN with PCA1

Watershed	Performance index	V (%)[1]	LR			K-NN			ANN		
			1*	2*	3*	1	2	3	1	2	3
Wuxi											
	RMSE	85	11.0	11.9	12.0	12.1	12.5	12.7	10.6	11.5	11.6
		90	11.0	11.9	12.0	12.1	12.5	12.7	10.7	11.5	11.6
		95	10.9	11.9	12.0	11.8	12.4	12.6	10.6	11.5	11.6
		100	10.9	11.9	12.0	11.8	12.4	12.6	10.6	11.5	11.6
	CE	85	0.12	-0.05	-0.07	-0.04	-0.15	-0.19	0.16	0.01	0.00
		90	0.12	-0.05	-0.07	-0.04	-0.15	-0.19	0.16	0.02	0.00
		95	0.12	-0.05	-0.07	-0.03	-0.14	-0.18	0.16	0.02	0.00
		100	0.12	-0.05	-0.07	-0.03	-0.14	-0.18	0.17	0.02	0.00
	PI	85	0.27	0.41	0.43	0.12	0.34	0.36	0.32	0.44	0.47
		90	0.27	0.41	0.43	0.12	0.34	0.36	0.31	0.45	0.47
		95	0.28	0.41	0.43	0.16	0.36	0.38	0.32	0.45	0.47
		100	0.28	0.41	0.43	0.16	0.36	0.38	0.32	0.45	0.47
India											
	RMSE	85	410.9	320.6	457.7	275.9	262.7	281.1	250.4	256.1	247.0
		90	294.9	307.8	311.1	260.3	256.2	276.8	248.1	252.3	249.0
		95	291.6	305.0	304.6	255.3	256.5	265.0	247.7	254.2	251.4
		100	286.8	301.6	302.7	246.6	257.3	251.2	245.2	245.9	247.2
	CE	85	0.81	0.89	0.78	0.91	0.91	0.90	0.93	0.92	0.93
		90	0.90	0.89	0.89	0.92	0.91	0.91	0.93	0.93	0.93
		95	0.90	0.89	0.89	0.92	0.91	0.91	0.93	0.92	0.93
		100	0.90	0.89	0.89	0.93	0.92	0.93	0.92	0.92	0.93
	PI	85	0.61	0.92	0.90	0.81	0.93	0.96	0.85	0.94	0.97
		90	0.79	0.92	0.95	0.83	0.94	0.96	0.85	0.95	0.97
		95	0.80	0.92	0.95	0.84	0.94	0.96	0.85	0.95	0.97
		100	0.80	0.92	0.95	0.85	0.94	0.97	0.84	0.94	0.97

Note: * "1, 2, and 3" denote one-, two-, and three-step-ahead forecasts; [1] "V" stands for the percentage of total variance.

Table 16. Multiple-step predictions for Wuxi and India series using LR, K-NN, and ANN with PCA2

Watershed	Performance index	V (%)[1]	LR			K-NN			ANN		
			1*	2*	3*	1	2	3	1	2	3
Wuxi											
	RMSE	85	10.8	11.6	11.6	11.5	12.2	12.7	10.6	11.5	11.6
		90	10.8	11.6	11.6	11.5	12.2	12.7	10.6	11.5	11.6
		95	10.9	11.9	12.0	11.8	12.4	12.6	10.6	11.5	11.6
		100	10.9	11.9	12.0	11.8	12.4	12.6	10.6	11.5	11.6
	CE	85	0.14	0.01	0.00	0.02	-0.11	-0.19	0.16	0.02	0.00
		90	0.14	0.01	0.00	0.02	-0.11	-0.19	0.16	0.02	0.00
		95	0.12	-0.05	-0.07	-0.03	-0.14	-0.18	0.17	0.02	0.00
		100	0.12	-0.05	-0.07	-0.03	-0.14	-0.18	0.16	0.02	0.00
	PI	85	0.30	0.44	0.47	0.20	0.37	0.37	0.32	0.44	0.47
		90	0.30	0.44	0.47	0.20	0.37	0.37	0.32	0.45	0.47
		95	0.28	0.41	0.43	0.16	0.36	0.38	0.32	0.45	0.47
		100	0.28	0.41	0.43	0.16	0.36	0.38	0.32	0.45	0.47
India											
	RMSE	85	482.9	413.0	800.6	254.7	248.9	253.1	241.7	247.2	249.8
		90	352.8	357.8	600.6	253.1	249.4	250.4	245.0	247.9	247.8
		95	326.5	331.9	376.1	247.8	246.1	246.1	243.5	249.0	244.8
		100	286.8	301.6	302.7	246.6	257.3	251.2	247.6	252.3	247.1
	CE	85	0.73	0.80	-2.74	0.92	0.93	0.93	0.93	0.93	0.93
		90	0.86	0.85	0.58	0.93	0.93	0.93	0.93	0.93	0.93
		95	0.88	0.87	0.84	0.93	0.93	0.93	0.93	0.93	0.93
		100	0.90	0.89	0.89	0.93	0.92	0.93	0.93	0.93	0.93
	PI	85	0.44	0.86	-1.75	0.84	0.95	0.97	0.86	0.95	0.97
		90	0.70	0.89	0.81	0.85	0.95	0.97	0.86	0.95	0.97
		95	0.74	0.91	0.93	0.85	0.95	0.97	0.86	0.95	0.97
		100	0.80	0.92	0.95	0.85	0.94	0.97	0.85	0.95	0.97

Table 17. Optimal *p* RCs for model inputs at various forecasting horizons

Watershed	Model	Prediction horizons	Supervised filter (SSA1)		Unsupervised filter (SSA2)	
			Optimal *p* RCs	RMSE	Optimal *p* RCs	RMSE
Wuxi						
	LR	1	1* 2*	6.01	1 2	6.01
		2	1 5	7.73	1 5	7.73
		3	1	8.40	1	8.40
	K-NN	1	1	8.02	2 3	7.17
		2	1	8.41	2 4	8.03
		3	1	9.99	2	9.69
	ANN	1	1 2 3	4.43	1 2 3	4.43
		2	1 5	5.82	1 5	5.57
		3	1	6.42	1	6.25
Zhenwan						
	LR	1	1 2 3	7.19	1 2 3	7.19
		2	1 7 2	7.99	1 2 7	7.99
		3	1 5	8.81	1 5	8.81
	K-NN	1	1 2 3 4 5	9.84	3 6	7.64
		2	1	9.72	3 6	8.95
		3	1	10.33	2 5	10.24
	ANN	1	1 2 3 4	5.55	5 6 7	5.02
		2	1 7 2	5.84	3 7	5.51
		3	1 5	6.58	3 7	5.56
India						
	LR	1	2 1 3 4	185.95	1 2 3 4	185.95
		2	1 2	237.85	1 2	237.85
		3	3 4 2 7 1	299.14	1 2 6	287.16
	K-NN	1	2 1 3 4 5	236.90	1 2 5 6	236.39
		2	1 2	247.44	1 2 5	242.65
		3	3 4 2 7	249.43	1 2 5 6	243.86
	ANN	1	2	166.58	1 7	164.70
		2	1 2 7	173.57	1 2 7	166.30
		3	3 4 2 7 1	235.59	1 5 7	172.56
Zhongxian						
	LR	1	1 2	40.06	1 2	40.06
		2	1 2 6	41.87	1 6	39.44
		3	3 6 2 4 1	58.29	1 5	41.53
	K-NN	1	1 2	51.78	1 2	51.78
		2	1 2	53.86	1 2 3	53.32
		3	3 6 2	54.15	2 3	53.16
	ANN	1	1 2 3	38.39	1 5 6	34.09
		2	1	44.49	1 5 6	39.45
		3	3 6	46.14	1 5	34.94

Note:* the numbers of "1, 2" stand for RC1 and RC2, and these numbers in the SSA1 column is in a descending order of CCFs shown in Figure 6.12.

Table 18. Model performances at three forecasting horizons using MANN and ANN with the SSA2

Watershed	Model	RMSE			CE			PI		
		1	2	3	1	2	3	1	2	3
WuXi										
	ANN	4.43	5.57	6.25	0.84	0.78	0.70	0.87	0.88	0.84
	MANN	3.63	4.32	3.93	0.90	0.86	0.89	0.92	0.92	0.94
Zhenwan										
	ANN	5.02	5.51	5.56	0.81	0.75	0.71	0.88	0.85	0.84
	MANN	3.18	3.20	3.31	0.92	0.92	0.91	0.95	0.95	0.95
India										
	ANN	164.7	166.3	172.6	0.97	0.97	0.97	0.95	0.98	0.99
	MANN	144.2	145.1	157.4	0.98	0.98	0.97	0.95	0.98	0.99
Zhongxian										
	ANN	34.09	39.45	34.94	0.84	0.71	0.83	0.84	0.80	0.92
	MANN	28.58	32.24	32.69	0.86	0.82	0.82	0.86	0.88	0.91

Table 19. RMSE of the LR model coupled with SSA2 using various L

Watershed	Prediction horizons	L in SSA							
		3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Wuxi									
	1	6.13	5.94	**6.01**	6.41	5.83	5.81	5.61	5.51
	2	7.79	7.62	**7.73**	8.14	7.71	7.75	7.62	7.66
	3	11.76	8.61	**8.40**	9.04	9.23	8.72	8.56	8.62
Zhenwan									
	1	7.74	7.49	7.31	7.29	**7.19**	6.99	7.08	6.84
	2	10.27	8.42	9.04	8.19	**7.99**	7.67	7.43	7.61
	3	11.28	10.66	10.06	9.33	**8.81**	8.91	9.42	9.28