Data-driven models for Monthly Streamflow Time Series Prediction

C. L. Wu and K. W. Chau*

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

*Email: cekwchau@polyu.edu.hk

ABSTRACT

Data-driven techniques such as Auto-Regressive Moving Average (ARMA), K-Nearest-Neighbors (KNN), and Artificial Neural Networks (ANN), are widely applied to hydrologic time series prediction. This paper investigates different data-driven models to determine the optimal approach of predicting monthly streamflow time series. Four sets of data from different locations of People's Republic of China (Xiangjiaba, Cuntan, Manwan, and Danjiangkou) are applied for the investigation process. Correlation integral and False Nearest Neighbors (FNN) are first employed for Phase Space Reconstruction (PSR). Four models, ARMA, ANN, KNN, and Phase Space Reconstruction-based Artificial Neural Networks (ANN-PSR) are then compared by one-month-ahead forecast using Cuntan and Danjiangkou data. The KNN model performs the best among the four models, but only exhibits weak superiority to ARMA. Further analysis demonstrates that a low correlation between model inputs and outputs could be the main reason to restrict the power of ANN. A Moving Average Artificial Neural Networks (MA-ANN), using the moving average of streamflow series as inputs, is also proposed in this study. The results show that the MA-ANN has a significant improvement on the forecast accuracy compared with the original four models. This is mainly due to the improvement of correlation between inputs and outputs depending on the moving average operation. The optimal memory lengths of the moving average were three and six for Cuntan and Danjiangkou respectively when the optimal model inputs are recognized as the previous twelve months.

Keywords: Hydrologic time series, Auto-Regressive Moving Average, K-Nearest-Neighbors, Artificial Neural Networks, Phase Space Reconstruction, False Nearest Neighbors, dynamics of chaos.

1. Introduction

Many data-driven models, including linear, nonparametric or nonlinear approaches, are developed for hydrologic discharge time series prediction in the past decades (*Marques et al., 2006*). Generally, there are two basic assumptions underlay different model techniques. The first assumption suggests that a time series is originated from a stochastic process with an infinite number of degrees of freedom. Under this assumption, linear models such as AutoRegressive (AR), AutoRegressive Moving Average (ARMA), AutoRegressive Integrated Moving Average (ARIMA), and Seasonal ARIMA (SARIMA) had made a great success in river flow prediction (*Carlson et al., 1970; Salas et al., 1985; Haltiner and Salas, 1988; Yu and Tseng, 1996; Kothyari and Singh, 1999; Huang et al., 2004; María et al., 2004*).

The second assumption is that a random-looking hydrologic time series is derived from a deterministic dynamic system such as chaos. In the past two decades, chaos-based streamflow prediction techniques have been increasingly obtaining interests of the hydrology community (*Jayawardena and Lai, 1994; Jayawardena and Gurung, 2000; Elshorbagy et al., 2002; Wang et al, 2006b*) although some doubts have been raised in terms of the existence of chaos in hydrologic data (*Ghilardi and Rosso, 1990; Koutsoyiannis and Pachakis, 1996; Pasternack, 1999; Schertzer et al., 2002; Wang et al., 2006a*). Generally, the prediction techniques for a dynamic system can be roughly divided into two approaches: local and global. Local approach uses only nearby states to make predictions whereas global approach involves all the states. K-Nearest-Neighbors (KNN) algorithm, Artificial Neural Networks (ANN) and Support Vectors Machine (SVM) are some typical forecast methods for dynamic systems (*Sivapragasam et al., 2001; Laio et al. ,2003; Wang et al., 2006b*). Phase-Space-Reconstruction (PSR) is a

precondition before performing any predictions of the dynamic system. Typical methods involved in PSR are correlation integral, singular-value decomposition of the sample covariance matrix, False Nearest Neighbors (FNN), and true vector fields (*Grassberger and Procaccia, 1983;Abarbanel et al., 1993*).

Comparative studies on the above prediction techniques have been further carried out by some researchers. Sivakumar et al. (*2002*) found that the performance of the KNN approach was consistently better than ANN in short-term river flow prediction. Laio et al. (*2003*) carried out a comparison of KNN and ANN for flood predictions and found that KNN performed slightly better at short forecast time while the situation was reversed for longer time. Similarly, Yu et al. (*2004*) proposed that KNN performed worse than ARIMA on the basis of daily streamflow prediction. The conclusions in literature are very inconsistent. It is difficult to justify which modeling technique is more suitable for a streamflow forecast.

The above two assumptions are in the extremes of a hydrologic streamflow series. Salas et al. (*1985*) suggested that a streamflow process should be treated as an integration of stochastic (or random) and deterministic components. Describing it as either a totally linear stochastic process or fully nonlinear deterministic chaos is not a practical approach (*Elshorbagy et al. 2002*). Therefore, the model based on either of two assumptions may not be the most suitable. An investigation on an optimal prediction model is worthy to further study with different real monthly streamflow data (Xiangjiaba, Cuntan, Manwan, and Danjiangkou).

The scope of this study is to compare four forecast models, ARMA, ANN, KNN, and ANN-PSR and develop an optimal model for monthly streamflow prediction. This paper is organized in the following manner. Section 2 presents the four sets of streamflow data used in this study. Section 3 first describes the principles of PSR and then identifies its parameters using the correlation integral approach and the FNN approach. The implementation of the forecast models, including data preparation and selection of parameters, is discussed in Section 4. Forecast results are described in Section 5 and conclusions of the study are presented in Section 6.

2. Streamflow Data

Monthly streamflow series of three watersheds and one river, i.e. Xiangjiaba, Manwan, Danjiangkou, and Yangtze River, were analyzed in this study.

The largest watershed, Xiangjiaba, is at the upstream of Yangtze river with average yearly discharge of 4538 m^3/s. Monthly streamflow series were taken from the hydrological station near the Xiangjiaba Dam site located in Sichuan Province. The basin area contributed to the streamflow series is around 45.88×10^4 km^2. The period of the data was from January 1940 to December 1997.

The medium watershed, Manwan, is located in the Lancang River which originates from the Qinghai-Tibet Plateau. Monthly streamflow series were taken from the hydrological station near the Manwan Dam site located in Sichuan Province. The catchment area controlled by the station is 11.45×10^4 km^2, and the average yearly discharge is 1230 m^3/s based on a statistic of 30-year data (January 1974 to December 2003).

The smallest watershed, Danjiangkou, lies at the upstream of Han river with average yearly discharge of 1203 m^3/s. Monthly streamflow data came from the hydrology station at the Danjiangkou Dam site which is located in Hubei Province. The catchment area at the dam site is around 9.5×10^4 km^2. The data range was from January 1930 to December 1981.

1   The last streamflow series is Yangtze River, the largest river in China. The selected
2   monthly streamflow data were from the hydrology station of Cuntan located in the middle
3   stream of the river. The stream flow series spanned from January 1893 to December 2007.
4   Four monthly streamflow series are shown in Fig. 1. Monthly streamflow data in
5   Xiangjiaba, Manwan, and Cuntan are characterized by a smooth process whereas monthly
6   streamflow data in Danjiangkou exhibits complex oscillations. The linear fits (dotted lines in
7   Fig. 1) verify the consistency of the streamflow series. All series exhibit good consistency
8   because the linear fits are closed to horizontal. Since there was no large-scale hydraulic works
9   such as dams built during the data collection period, the streamflow process is fairly pristine in
10 each case.

11 ## 3. Reconstruction of Dynamics

12 ### 3.1. Phase Space Reconstruction
13   To describe the temporal evolution of a dynamical system in a multi-dimensional phase
14 space with a scale time series, it is essential to employ some techniques to unfold the multi-
15 dimensional structure using the available data (*Wang et al., 2006a*). The most frequently used
16 reconstruction method for a univariate or multivariate time series is the delay-time method
17 (*Takens, 1981; Farmer and Sidorowich ,1987; Sauer et al. ,1991; Jayawardena and Lai, 1994*).
18 A dynamic univariate time series $\{x_1, x_2, \cdots, x_N\}$, may be reconstructed into a series of delay

19 vectors of the type $\mathbf{Y}_t = \{x_t, x_{t+\tau}, x_{t+2\tau}, \cdots, x_{t+(m-1)\tau}\}$, $t = 1, 2, \cdots, N - (m-1)\tau$, where $\mathbf{Y}_t \in \mathrm{R}^m$, τ is the

20 delay time as a multiple of the sampling period and m is the embedding dimension. Under ideal
21 conditions of time series of infinite length, all the reconstructions would be analogous and
22 topologically equivalent to the real system. Owing to the shortness of real time series and the
23 inevitable presence of dynamical noise, optimal reconstruction is involved in the choice of
24 m and τ (*Laio et al., 2003*).

25   The time evolution of the dynamic system is given as a mapping $\mathbf{Y}(t) \mapsto \mathbf{Y}(t+T)$

26 (or $\mathbf{Y}_t \mapsto \mathbf{Y}_{t+T}$). The function relationship between the current state $\mathbf{Y}(t)$ at time t and the

27 predicted state $\mathbf{Y}^F(t+T)$ at time $t + T$ can be written as follows:

28 $$\mathbf{Y}^F(t+T) = f(\mathbf{Y}(t)) + e_t \qquad (1)$$

29 where e_t is a typical noise term. In the form of time series, it can be expressed

30 as $[x_{t+T}^F, x_{t+T+\tau}^F, \cdots, x_{t+T+(m-2)\tau}^F, x_{t+T+(m-1)\tau}^F] = f([x_t, x_{t+\tau}, \cdots x_{t+(m-2)\tau}, x_{t+(m-1)\tau}]) + e_t$.   Therefore,

31 predicting future trajectory by current trajectory becomes viable once the function $f(\bullet)$ is
32 determined. In practice, the expression is often defined as:

33 $$x_{t+T+(m-1)\tau}^F = f(\mathbf{Y}(t)) + e_t \qquad (2)$$

34 where only the last component in $\mathbf{Y}^F(t+T)$ is indicated since normally the prediction of this last
35 component is of concern (*Laio et al., 2003*). For more details and examples of phase space
36 reconstruction, please refer to *Laio et al. (2003).*
37   Both global and local methods can be applied to estimate the function of $f(\bullet)$ in Eq. (2).
38 The global approach depends on the observations at all points whereas the local approach
39 depends on observations that are in some finite neighborhood of the point of estimate. The
40 KNN algorithm is a widely-used local method, which was originally developed by Farmer and
41 Sidorowich (*1987*). The basic idea behind KNN is that only nearby states are used to make

1 predictions for the local approximation. This method is similar to the Nearest Neighbor Method
2 (NNM) (*Yakowitz and Karlsson, 1987; Toth et al., 2000; Solomatine et al., 2008*) once an
3 attractor is correctly unfolded or the phase space of the dynamical system is correctly
4 reconstructed. Attractor reconstruction refers to those methods for inference of geometrical and
5 topological information about a dynamical attractor from observations. An attractor is the point
6 where the dynamics are discontinuous, through which the minimum embedding dimension can
7 be determined. For the purpose of comparisons, a local approach of KNN algorithm (*Farmer*
8 *and Sidorowich, 1987; Jayawardena and Lai, 1994*) and a global approach of ANN were
9 discussed in this study.

10 **3.2. Determination of Parameters (τ , m)**

11   Many methods, as mentioned in Section 1, are able to reconstruct the phase space by
12 identifying the parameter pair (τ , m). The correlation integral and FNN were employed to the
13 following discussion. The correlation integral identifies the two parameters for the perspective
14 of verifying the existence of chaos.

15 *(1) Correlation exponent*

16   The correlation exponent method is commonly used to investigate the existence of
17 chaos in hydrology community (*Jayawardena and Lai, 1994; Sivakumar et al., 1998;*
18 *Sivakumar et al., 2001*). Correlation integral is generally applied to verify the existence of
19 chaotic dynamics using the saturation value of the correlation dimension. The diagnosis of the
20 existence of chaos can begin if the phase space has been reconstructed. The PSR requires two
21 parameters (τ , m), where m can be identified from the plot of the correlation dimension (d_2)
22 versus m when it reaches a saturation value (D_2) (*Tsonis, 1992*). In other words, the saturation
23 value must be obtained prior the achievement of m . However, m is subject to τ (often called
24 decorrelation time) and should be first determined. Also, some researchers thought that τ and m
25 should not be determined separately. The length of the embedding window, $\tau_w = (m-1)\tau$
26 (*Broomhead and King, 1986; Mees et al., 1987; Martinerie et al., 1992*) and optimizing the
27 triplet (m, τ, k) using Genetic Algorithm (GA) (*Liong et al., 2002*) are typical examples of
28 determining τ and m together . In the present study, we tend to adopt a widely accepted
29 method to obtain τ .

30   Determination of τ and m have been reported in numerous works (*Grassberger and*
31 *Procaccia,1983; Fraser and Swinney, 1986; Casdagli et al., 1991; Tsonis, 1992; Abarbanel et*
32 *al., 1993; Kugiumtzis, 1996; Hegger et al., 1999; and Kantz and Schreiber, 2004*). Generally,
33 τ can be defined when the AutoCorrelation Function (ACF) attains the value of zero or below a
34 small value, or the Average Mutual Information (AMI) reaches the first minimum. The
35 calculations of ACF and AMI are discussed in detail in the works of Fraser and Swinney (*1986*),
36 Tsonis (*1992*), and Abarbanel et al. (*1993*). Fig. 2 displays the ACF and AMI for the four
37 monthly streamflow series. The ACF for all series are first attained zeros at the same lag time
38 of 3. Since the AMI gives the same estimates of τ as the ACF, τ was consistently chosen at the
39 lag time of 3 for all cases.

40   It is also suggested that a relationship of $\tau = T/m$ (where T represents the dominant
41 periodicity of the original series as revealed by Fourier analysis) exists if ACF is periodic
42 (*Tsonis, 1992*). Fig. 2 also demonstrates that the present cases suffice the condition. If all
43 streamflow series are chaotic, the potential m are therefore around 4 (i.e. $m = 12/3$) with all the
44 predominant periodicity of 12 months and τ of 3.

1   After the primary determination of τ, the correlation dimension can be computed by the
2 correlation integral according the formula of Grassberger-Procaccia algorithm (*Grassberger*
3 *and Procaccia,1983*). This original formula was modified by Theiler (*1986*) for the estimation
4 of the correlation integral in a time series which poses serious problems of temporal
5 correlations. Thus, the modified correlation integral $C(r)$ for a m-dimension phase space is
6 defined as:

7
$$C(r) = \frac{2}{N_{\text{pairs}}} \sum_{i=1}^{N} \sum_{j=i+w+1}^{N-i} \text{H}(r - \left\| \mathbf{Y}_i\text{-}\mathbf{Y}_j \right\|) \tag{3}$$

8 where $N_{\text{pairs}} = (N - w + 1)(N - w)$, w is the Theiler window excluding those points which are
9 temporally correlated, $\mathbf{Y}_i$ and N are already discussed in Section 3.1, r is the radius of a ball
10 centered on $\mathbf{Y}_i$, H is the Heaviside step function with $\text{H}(u) = 1$ if $u > 0$ and $\text{H}(u) = 0$ if $u \leq 0$.
11 The correlation integral only counts the pairs $(\mathbf{Y}_i, \mathbf{Y}_j)$ whose distance, in a Euclidean sense, is
12 smaller than r. In the limit of an infinite amount of data ($N \rightarrow \infty$) and sufficiently small r, the
13 relation of $C(r) \propto r^{D_2}$ between $C(r)$ and r is expected when m exceeds the correlation
14 dimension of the chaos system. The correlation exponent ν and the correlation dimension D_2
15 can be defined as:

16
$$\nu = \frac{\partial \ln C(r)}{\partial \ln r} \tag{4}$$

17
$$D_2 = \lim_{\substack{r \to 0 \\ N \to \infty}} \nu \tag{5}$$

18 Since D_2 is unknown before conducting the computation, the convergence of the correlation
19 dimension D_2 in m must be examined.
20   The procedure of the computation is first to plot $\ln C(r)$ versus $\ln r$ with a given m.
21 Then, the potential scaling region is determined wherever the slope (i.e. the correlation
22 exponent ν) of the curve for the given m is approximately constant. The constant slope can be
23 estimated by a straight line fitting of the scaling region. In general, the best way to define the
24 scaling region is to produce another figure which demonstrates the slope of the $\ln C(r)$ as a
25 function of $\ln r$. If a scaling region exists, a plateau should be shown in the figure. This plateau
26 provides an estimate for d_2, a correlation dimension of the possible attractor for the present m.
27 If d_2 converges to a finite value D_2 (i.e. saturation value) after repeating the above procedure for
28 successively higher m, a true attractor of dimension D_2 is formed and the system may be
29 considered as chaos. Meanwhile, m can be identified as the value that corresponds to the first
30 occurrence of the saturation value D_2 in the plot of d_2 versus m.
31   For a finite dataset, there is a radius r below which there are no pairs of points (i.e.
32 depopulation zone). Conversely, when r approaches the diameter of the cloud of points, the
33 number of pairs will increase no further as r increases (i.e. saturation zone). The scaling region
34 would be found somewhere between the depopulation and the saturation zones. In view
35 of $C(r) \propto r^{D_2}$, the population of pairs of points for a finite data set on small scales is smaller
36 than the population of pairs on large scales. This leads to poor statistics at small r and the
37 function $C(r)$ may be distorted. Nevertheless, the scaling region over large r's should remain
38 unchanged if there are sufficient points available. On the contrary, the scaling region may be
39 completely masked if there are inadequate points. Also, the scaling region can be "lost"
40 between the depopulation and the saturation zones by increasing the m while the number of

1 points remain the same positions. Thus, an accurate estimation of d_2 requires a minimum
2 number of points.
3   Many literatures focus on the quantity of data required to determine the reliable value of
4 d_2 (*Abarbanel et al., 1993 and Wang et al., 2006a*). Some researches claim that the size should
5 be 10^A (*Procaccia, 1988*) or $10^{2+0.4m}$ (*Tsonis, 1992*), where A is the greatest integer smaller than
6 d_2 and m ($m < 20$) is the embedding dimension used for estimating d_2 with an error less than
7 5%. Whereas other researches found that smaller data size is needed. For instance, the
8 minimum data points for reliable d_2 is $10^{d_2/2}$ (*Ruelle, 1990; Essex and Nerenberg, 1991*), or
9 $\sqrt{27.5}^{d_2}$ (*Hong and Hong, 1994*) and empirical results of dimension calculations are not
10 substantially altered by going from 3000 or 6000 points to subsets of 500 points (*Abraham et
11 al., 1986*).
12   The $\ln C(r)$ versus $\ln r$ graphs and the correlation exponent ν versus $\ln r$ for the four
13 streamflow series are shown in Figs. 4 and 5 respectively. Fig. 5 demonstrates that the scaling
14 region cannot be identified for any m ($m < 20$) in Danjiangkou catchment whereas the scaling
15 region can be determined in other three cases. Obviously, Danjiangkou series cannot be totally
16 defined as non-chaotic or random process because its number of data is as small as 624. A fact
17 should be noted for the other three streamflow series is that the identified scaling regions
18 become narrow or even "lost" as the saturation zone occurs over larger scales with the increase
19 of m. If the number of data in the streamflow series decreases, the speed for the scaling region
20 to narrow or "lost" with the increase of m will become faster. A representative detail of these
21 curves in Fig. 5 is presented in Fig. 6 by taking curves at $m = 1, 4, 8$ and 20. The scaling region
22 becomes ambiguous at $m = 8$ for Manwan and Xiangjiaba whereas a narrow scaling region can
23 still be defined even at $m = 20$ for Cuntan which has 1380 points from data of 114 years.
24 However, the length of data size in Danjiangkou, Cuntan and Manwan is adequate in defining
25 the scaling region when $m \leq 4$. Based on the results of Fig. 5, the relationship between
26 correlation dimension d_2 and embedding dimension m is depicted in Fig. 7. The saturation
27 values D_2 for the three streamflow series are at interval of (1.5, 2). Generally, a sufficient
28 condition for the smallest m is that m is an integer larger than $2D_2$. The associated m is
29 therefore set the value of 4 for the three series. A test for the robustness of m with variable
30 τ from 1 to 5 was performed, which implied that m is insensitive to τ.
31   With the potential values on D_2 or m, some criteria such as 10^A can be satisfied whereas
32 other criteria such as $10^{2+0.4m}$ cannot be satisfied. The latter criteria means that few hydrologic
33 records can be assessed for $m > 5$ attractors since as many as 10,000 points require 27 years of
34 daily records or around 900 years of monthly records. Thus, the three monthly series of
35 Danjiangkou, Cuntan and Manwan may be treated as chaotic with suggested variable $m = 4$.
36 Furthermore, the phase portraits of four streamflow series are portrayed in Fig. 8 where (τ, m) is
37 (3,3). Obviously, the state spaces in the 3-dimensional maps are clearly unfolded for Xiangjiaba,
38 Cuntan, and Manwan whereas no clear trajectory is revealed for Danjiangkou.

39 *(2) False Nearest Neighbors(FNN)*
40   The correlation integral method appears to be data intensive and certainly subjective.
41 For simplicity, the FNN method is commonly employed for the PSR of a hydrologic
42 streamflow series (*Wang et al., 2006b*). The FNN algorithm was originally developed for
43 determining the number of time-delay coordinates needed to recreate autonomous dynamics
44 directly from properties of the data itself (*Kennel et al., 1992; Abarbanel et al., 1993*). The
45 following discussion outlines the basic concepts of the FNN algorithm. Suppose the point

1 $\mathbf{Y}_i = \{x_i, x_{i+\tau}, x_{i+2\tau}, \cdots, x_{i+(m-1)\tau}\}$ has a neighbor $\mathbf{Y}_j = \{x_j, x_{j+\tau}, x_{j+2\tau}, \cdots, x_{j+(m-1)\tau}\}$, the criterion that

2 $\mathbf{Y}_j$ is viewed as a false neighbor of $\mathbf{Y}_i$ is:

3
$$\frac{\left|x_{i+m\tau} - x_{j+m\tau}\right|}{\left\|\mathbf{Y}_i\text{-}\mathbf{Y}_j\right\|} > R_{tol} \tag{6}$$

4 where $\|\ \|$ stands for the distance in a Euclidean sense, R_{tol} is some threshold with the common

5 range of 10 to 30 (*Wang et al., 2006b*). For all points i in the vector state space, Eq. (6) is

6 performed and then the percentage of points which have FNNs is calculated. The algorithm is

7 repeated for increasing m until the percentage of FNNs drops to zero, or some acceptable small

8 number such as 1%, where m is the target m (*Abarbanel et al., 1993*).

9   The sensitivity analysis of the percentage of FNNs (FNNP) on R_{tol} is respectively

10 demonstrated in Fig. 9, where τ is set as value of 3 and R_{tol} is from 10 to 30 with a step size of

11 5. Fig. 9 shows that FNNP is stable when $R_{tol} \geq 15$. Fig. 10 demonstrates FNNPs for the four

12 streamflow series with $R_{tol} = 15$. The identified m is 4 for Manwan and 5 for other three cases.

13 Two random series with the same data size as Cuntan and Danjiakou were respectively

14 generated by the random function. Their FNNPs with τ of 3 and R_{tol} of 15 are demonstrated in

15 Fig. 11 where a similar m was found. The FNN technique may not be able to distinguish

16 random process from deterministic system. Therefore, the phase space reconstructed by the

17 FNN may not reveal the true trajectory hidden in the dynamic system. In other words, the FNN

18 technique is not concerned with a dynamic system being deterministic or not. Thus, the

19 correlation integral method can be more reliable for unfolding a dynamical system. The

20 preliminary parameter pair (τ, m) in PSR is (3, 4) for Xiangjiaba, Cuntan, and Manwan. For the

21 purpose of the symmetry in models comparison, the parameter pair (τ, m) for Danjiangkou is (3,

22 5) from Fig. 10. Certainly, these parameters pair need to be checked further for their robustness

23 when they are used in KNN model.

24 ## 4. Implementation of prediction models

25 **4.1 Data preparation**

26   Streamflow series data were divided into three parts: model training, cross-validation

27 and validation. The last ten years' streamflow data, called validation set, were set aside for

28 validation. This validation set would not be used until all model development and training was

29 finished completely. Of the remaining data, the first two-thirds called training set was for

30 model training, and the other one-third called cross-validation set was for the purpose of

31 confirming and validating the initial analysis. The ANN model has difficulties to extrapolate

32 beyond the range of the data used for training. As a consequence, poor predictions can be

33 expected when the validation data contains values outside the range of those used for training.

34 It is also imperative that the training and validation sets are representative of the same

35 population. Statistical properties (mean, deviation, range) from them are compared in order to

36 measure the representative. Table 1 displays the statistical properties of these data sets.

37 Generally, the division of data is satisfied for ANN since the statistical parameters of the

38 training sets for all the streamflow series are closed to the cross-validation sets and the testing

39 sets.

1   **4.2 Evaluation of model performances**

2   The Pearson's correlation coefficient (r) or the coefficient of determination ($R^2 = r^2$),
3   have been identified as inappropriate measures in hydrologic model evaluation by *Legates and*
4   *McCabe* (1999). The coefficient of efficiency (CE) (*Nash and Sutcliffe, 1970*) is a good
5   alternative to r or R^2 as a "goodness-of-fit" or relative error measure in that it is sensitive to
6   differences in the observed and forecasted means and variances. *Legates and McCabe* (1999)
7   also suggested that a complete assessment of model performance should include at least one
8   absolute error measure (e.g., Root Mean Square Error) as necessary supplement to a relative
9   error measure. Besides, the Persistence Index (PI) (*Kitanidis And Bras, 1980*) is able to check
10 the prediction lag effect.
11   Thus, three measures of evaluating model performance are used in the present paper
12 comprising Root Mean Square Error (RMSE), the Nash-Sutcliffe Coefficient of Efficiency (CE)
13 (*Nash and Sutcliffe, 1970*), and the PI (*Kitanidis And Bras, 1980*). They are respectively

14 formulated as: $\text{RMSE} = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N}(T_i - \hat{T}_i)^2}$), $\text{CE} = 1 - \sum_{i=1}^{N}(T_i - \hat{T}_i)^2 \Big/ \sum_{i=1}^{N}(T_i - \bar{T})^2$ and

15 $\text{PI} = 1 - \sum_{i=1}^{N}(T_i - \hat{T}_i)^2 \Big/ \sum_{i=1}^{N}(T_i - T_{i-L})^2$. In these equations, N =number of observations, $\hat{T}_i$ =predicted

16 streamflow, T_i =observed streamflow, $\bar{T}$ =average observed streamflow, and T_{i-L} is the
17 streamflow estimated from a so-call persistence model (or called naïve model) that basically
18 takes the last streamflow observation (at time i minus the lead time L) as a prediction. CE
19 and PI values of 1 indicate perfect fits.

20 **4.3 Configurations of models and parameters optimization**

21   Cuntan streamflow series was used as a representative case due to the similarity among
22 Xiangjiaba, Cuntan, and Manwan. Two cases, Cuntan and Danjiangkou, are therefore analyzed
23 in the rest of this article.
24   Autoregressive-moving-average (ARMA) models are mathematical models of the
25 persistence, or autocorrelation, in a time series. ARMA models can contribute to understanding
26 the physical hydrological system by disclosing some information about the physical process
27 that builds persistence into the series. Models of ARMA and ANN were used for the purpose of
28 comparison with models based on PSR. The order of the ARMA model is included in
29 parentheses as ARMA(p,q), where *p* is the autoregressive order and *q* the moving-average order.
30 Parameters (*p, q*) in ARMA were estimated by trial and error with each of them varied over the
31 range of 0 to 12, excluding the unfeasible case where both of them are simultaneously equal to
32 0, in view of the predominant periodicity of 12 months (Figs. 3 and 4). The best pair of (*p, q*) is
33 associated with the minimal value of Akaike Information Criterion (AIC), which is a
34 commonly adopted statistical measures of goodness-of-fit of an ARMA(*p,q*) model. Goodness
35 of fit might be expected to be measured by some function of the variance of the model residuals,
36 i.e., the fit improves as the residuals become smaller.
37   ANNs mimic the functioning of the human brain and nervous systems in a simplified
38 computational form. They are constituted by highly interconnected simple elements (termed
39 artificial neurons or sometimes termed nodes) which receive information, elaborate them
40 through mathematical functions and pass them to other artificial neurons (Alvisi et al., 2006).
41 They acquire knowledge through a learning process that involves finding an optimal set of
42 weights for the connections and threshold values for the neurons. ANNs can be categorized as
43 single layer, bilayer and multilayer according to the number of layers, and as feed-forward,

1   recurrent, and self-organizing according to the direction of information flow and processing
2   (ASCE, 2000a). With respect to ANN model, the type of ANN and its architecture must be
3   determined. The ANN type used in this study was the static multilayer feed-forward network.
4   The static ANN is able to capture the dynamics of a system in the network model by using
5   delay time inputs. The architecture design of the ANN comprises the number of hidden layers
6   and the number of neurons in input layer, hidden layers and output layer. ANNs with one
7   hidden layer are commonly used in hydrologic modeling (*Dawson and Wilby, 2001; de Vos*
8   *and Rientjes, 2005*) since these networks are considered to provide enough complexity to
9   accurately simulate the dynamic and nonlinear-properties of the hydrologic process. Therefore,
10 a three-layer static ANN was finally employed.
11   One output neuron was selected whilst the input neurons were determined by following
12 the popular approach of examining the dependence between the input and output time series.
13 The model inputs are originally considered to take the previous three sequential data because
14 the ACF first attains zero and AMI achieves the first local minima at the lag time of 3
15 simultaneously (Fig. 3). A trial and error method was then performed to check the three inputs
16 by systematically increasing the number of inputs from the latest month to the past twelve
17 months. The test results showed that the model with the 12-month data as inputs was optimal
18 for Cuntan and Danjiangkou.
19   The ensuing task is to optimize the size of the hidden layer with the chosen three inputs
20 and one output. The optimal size of the hidden layer was found by systematically increasing the
21 number of hidden neurons from 1 to 20 until the network performance on the cross-validation
22 set was no longer improved significantly. The optimal number of hidden neurons may be
23 affected by the use of a training algorithm. The most popular supervised training algorithms are
24 gradient descent techniques (e.g. BackPropagation (BP) algorithm) and Newton or quasi-
25 Newton optimization techniques (e.g., Levenberg-Marquart (LM) algorithm, and Broyden-
26 Fltecher-Goldfarb-Shanno (BFGS) algorithm). They are called local optimization techniques.
27 Alternatives to these methods are global optimization techniques such as the Particle Swarm
28 Optimization (PSO), the Genetic algorithm (GA), and the Shuffled Complex Evolution (SCE-
29 UA). Generally, these local optimization methods have fast convergence but are susceptible to
30 local minima and unstable. The global optimization methods, on the other hand, are able to
31 overcome local minima and obtain more stable solutions but are slow convergence and tend to
32 find out a relative optimal solution. As an attempt to combine their merits of the two
33 optimization techniques, an integrated training algorithm combining LM with GA is adopted
34 (*Chau et al., 2005*). The basic idea behind this algorithm is that GA is first utilized in search of
35 a set of weights and biases for ANN. These chosen weights and biases are then used as initial
36 values for a further optimization via the LM algorithm. Consider Danjingkou as a pilot case,
37 Table 2 demonstrates the one-step-ahead forecast performance statistics with 30 time runs in
38 terms of RMSE, CE and PI along with training time of three training algorithms of LM, PSO
39 and integrated algorithm of LM and GA. As expected, the PSO has better stability of
40 performance at the expense of training time whereas the LM has faster convergence but
41 exhibits a weak stability of performance in terms of the performance values in the Range
42 column (last column of Table 2). The hybrid algorithm indeed combines their merits with both
43 fast convergence and good stability of performance. Thus, the LM-GA algorithm was used for
44 training in the following study. The PSO is also acceptable because the time-consuming in
45 training is not too long. It was found that the optimal size of the hidden layer was 5 for all three
46 algorithms. Finally, a 12-5-1 configuration of the ANN model with the LM-GA algorithm was
47 designed for Danjiangkou. Similarly, a 12-13-1 configuration of ANN was applied to Cuntan.

1   The hyperbolic tangent functions were used as a transfer function in the hidden layer and output
2   layer. All input and output data were linearly scaled to the range of [-1, 1] because the output of
3   this transfer function is bounded to the range of [-1, 1].
4   The KNN method is based on local approximation which makes use of only nearby
5   observations of the point of estimate. It has its origins as a non-parametric statistical pattern
6   recognition procedure, aiming at distinguishing between different patterns according to chosen
7   criteria (Mack and Rosenblatt, 1979; Aha *et al.*, 1991). Amongst various non-parametric kernel
8   methods, the K-NN method seems more intuitive, but nevertheless possesses powerful
9   statistical properties (Karlsson and Yakowitz, 1987a, b). Yakowitz (1987) and Karlsson and
10   Yakowitz (1987a, b) did considerable work in extending the KNN method to time-series and
11   forecasting problems, obtaining satisfactory results and constructing a robust theoretical base
12   for the KNN method. Please refer to Karlsson and Yakowitz (1987a) for details.
13   Two PSR-based models, KNN and ANN (hereafter referred to as ANN-PSR), were
14   applied to the following study. ANN-PSR is the phase space reconstruction-based artificial
15   neural networks integrating the techniques of phase space reconstruction and ANN. Both the
16   above two models have the same inputs which were the m coordinates of PSR. The number of
17   k was used as $k = m+1$ (*Sugihara and Mary, 1990*), but $k > m+1$ (*Farmer and Sidorowich,*
18   *1987*) was applied to ensure the stability of the solution. In order to verify the robustness of
19   m and τ, a trial-and-error procedure (i.e. testing the optimal combinations of τ and m) still
20   have to be performed although the preliminary parameter values of them have been
21   recommended previously. With all thirty-five combinations of τ (1 to 5) and m (2 to 8), the
22   identified optimal parameter pair (τ, m) in terms of RMSE were (5, 7) for Cuntan and (5, 6) for
23   Danjiangkou. Correspondingly, the fixed $k = 8$ and $k = 7$ were respectively adopted in Cuntan
24   and Danjiangkou. The ANN-PSR also employed the LM-GA algorithm for training. The
25   optimal size of the hidden layer was determined via the same method mentioned above. Finally,
26   the architectures of the ANN-PSR are 5-9-1 for both Cuntan and Danjiangkou. A summary on
27   relevant model parameters for one-step-ahead prediction were displayed in Table 3.

5. Results and discussion

5.1. Main results

30   Table 3 shows the one-month-ahead forecast performance of various models in terms of
31   the RMSE, CE, and PI for Cuntan and Danjiangkou. The KNN outperformed other models in
32   terms of the three performance indexes. However, the advantage of KNN over ARMA is very
33   weak. The comparison between ANN and ARMA demonstrates that the ANN cannot perfectly
34   capture the autocorrelation relationship in each streamflow series. ANN-PSR model displays
35   worse performance than the KNN in each case, which suggests that the local approach may be
36   more suitable for one-month-ahead forecast. For the perspective of different cases, the
37   performance indexes from each model are very low for Danjiangkou. Actually, Danjiangkou
38   series was characterized by very complex oscillations as depicted in Fig. 1. Its main signal was
39   heavily contaminated by noises as an evidence of appearance of many high frequency signals
40   (Fig. 2).
41   Figs. 12 and 13 present the detail of the observed and predicted validation data series,
42   their errors (predicted-observed), and relative errors (hereafter referred to REs) of forecasts
43   from four prediction models for Cuntan. These models exhibit good overall match between the
44   observed and predicted data series. The plots of errors (middle graphs of Figs. 12 and 13)
45   illustrates that each model underestimates or overestimates quite a number of peaks of

streamflows although the grand mean of REs is around 20%. The plots of REs (bottom graphs of Figs. 12 and 13) suggest that the low streamflows are mostly well forecasted. This is due to the frequent occurrences of the low streamflow patterns which allow the trained model to have better generalization of these regions. Figs. 12 and 13 also demonstrate that the timing of the peaks for these models is quite good. The results are consistent with the high PI in Table 3.

Figs. 14 and 15 show the detail of the observed and predicted validation data series, their errors, and REs from four prediction models for Danjiangkou. The forecasts of the peaks of streamflows are mostly underestimated although these models can capture the whole trend of validation data series. The grand mean of REs are about 50% but the maximum RE is up to 300% as shown in the plots of REs. There is a lag effect in the timing of the peaks for these models by visual inspection, which is reflected by the low PI in Table 3.

Histograms of forecast errors of one-month-ahead prediction for Cuntan are presented in Fig. 16, and corresponding Auto-Correlation Functions (ACFs) of forecast errors are depicted in Fig. 17. The histograms are quasi-normal distributions and errors satisfy random processes in terms of ACFs. The results suggest that employed four models were indeed trained fully and forecasts from them are reliable. Similar results also can be found for Dangjiangkou.

5.2. Discussion

The parameter of k in KNN model poses a great impact on the performance of KNN. As mentioned previously, the choice of k should still ensure the stability of the solution although a preliminary value of k is based on $k = m+1$. The verification of stability of k in terms of RMSE is presented in Fig. 18. The new value of k was 9 for both Cuntan and Danjiangkou (recall that the original k was 8 and 7 respectively). The performance of KNN with new k is presented in Table 4. Comparison of performance between the new KNN and the original KNN (Table 3) suggests that taking $k = m+1$ tends to be reasonable because the new k is nearby $m+1$ and there is only a small improvement of performance for the KNN with the new k.

The issue of lagged predictions in the ANN model has been mentioned by some researchers (*Dawson and Wilby, 1999; Jian and Srinivasulu, 2004; de Vos and Rientjes, 2005; Muttil and Chau, 2006*). The work of de Vos and Rientjes (*2005*) in the context of rainfall-runoff modeling investigated sources of the prediction lag effect and suggested two types of methods to resolve the issue. One of identified sources is due to the use of previous observed streamflow data as ANN inputs. Thus, the ANN tends to give the most weight to the latest streamflow input if the model inputs are highly auto-correlative. As a consequence, ANN models underrate the information contained in other input signals such as precipitation. An effective solution is to decrease the weight of streamflow components as model inputs by using a moving average time series of original streamflows as streamflow input components of the rainfall-runoff ANN.

Fig. 19 investigates the existence of the forecast lag effect by Cross-Correlation Functions (CCFs) between the predicted and observed validation data series. As expected by visual inspection previously, there is no lag effect for Cuntan whereas there is obvious lag effect in all models except for the KNN for Danjiangkou. The moving average operation was used to generate new monthly streamflow time series to construct the model inputs of ANN. Based on trial and error method, the optimal memory lengths of three months and six months were respectively used in the moving average of the streamflow series of Cuntan and Danjiangkou. With the same architecture as the original ANN, the new moving average ANN (hereafter referred to MA-ANN) achieved a significant improvement in performance of

predictions. Table 4 presents the forecast results using the MA-ANN for Cuntan and Danjiangkou. The high PI (over 0.9) indicates the elimination of timing error. Fig. 20 depicts the detail of forecasts of MA-ANN for Cuntan and Danjiangkou. The observed data was fitted perfectly in particular for Cuntan. The errors and REs also significantly decreased. Fig. 21 compares the CCFs between inputs and outputs in the MA-ANN model and the ANN model. The results show that absolute CCFs in the first twelve lags from the former are mostly larger than those from the latter, in particular for Danjiangkou. Therefore, the improved performance from MA-ANN is due to the increase of the correlation between inputs and outputs by moving average operation on original streamflow time series. This conclusion was different from the effect of the moving average method employed in the work of de Vos and Rientjes (*2005*). The moving average operation is however beneficial to the performance of ANN for one-step-ahead streamflow prediction indeed.

6. Conclusions

The purpose of this study attempts to determine a relative optimal forecast model for monthly streamflow data. Two methods of the PSR, the correlation dimension and the FNN, were employed. The results show that three cases, Xiangjiaba, Cuntan, and Manwan, exhibit chaotic characteristics on the basis of the correlation integral technique. The parameter pair (τ, m) identified by the correlation integral and FNN is only preliminary for the KNN and ANN-PSR models, and they need to be further chosen by trial and error method. The final (τ, m) is (5,7) and (5,6) for Cuntan and Danjiangkou respectively. Comparison of the performance for different models indicates that the KNN exhibited no obvious superiority to traditional ARMA. With respect to the PSR-based prediction model, the KNN outperformed the ANN-PSR. This implies that the local approximation technique is better than the global approximation technique for one-month-ahead discharge prediction (at least in the current cases). Additionally, the common three-layer-feed-forward ANN did not expose its powerful ability of mapping any complex function because its prediction performance was worse than ARMA. The CCFs between the observed and predicted data series show that the forecast lag effect exists for Danjiangkou. The operation of moving average on original streamflow series can significantly improve the performance of ANN and eliminate the timing error by increasing the correlation between inputs and outputs of ANN. For instance, the PI for Danjiangkou is from 0.39 in original ANN to 0.93 in MA-ANN and its corresponding CE is from 0.47 to 0.93. Based on the findings of this paper, the MA-ANN can be proposed as the optimal model for one-month-ahead forecast.

Abbreviation

ACF	AutoCorrelation Function
AIC	Akaike Information Criterion
AMI	Average Mutual Information
ANN	Artificial Neural Networks
AR	Auto-Regressive
ARMA	Auto-Regressive Moving Average
ARIMA	Auto-Regressive Integrated Moving Average
BFGS	Broyden-Fltecher-Goldfarb-Shanno
BP	BackPropagation
CCF	Cross-Correlation Function

1	CE	Coefficient of Efficiency
2	CT	Cuntan
3	DJK	Danjiangkou
4	FNN	False Nearest Neighbors
5	FNNP	Percentage of FNN
6	GA	Genetic algorithm
7	KNN	K-Nearest-Neighbors
8	LM	Levenberg-Marquart
9	LM-GA	Levenberg-Marquart and Genetic algorithm
10	MA-ANN	Moving Average Artificial Neural Networks
11	MW	Manwan
12	NNM	Nearest Neighbor Method
13	PI	Persistence Index
14	RE	Relative Error
15	PSO	Particle Swarm Optimization
16	PSR	Phase Space Reconstruction
17	ANN-PSR	Phase Space Reconstruction-based Artificial Neural Networks
18	RMSE	Root Mean Square Error
19	SARIMA	Seasonal ARIMA
20	SEC-UA	Shuffled Complex Evolution
21	SVM	Support Vectors Machine
22	XJB	Xiangjiaba
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		

1 References

2 Abraham, N.B., Albano, A.M., Das, B., de Guzman, G., Yong, S., Gioggia, R.S., Puccioni, G.P., and Tredicce, J.R.
3 (1986), Calculating the dimension of attractors from small data. Phys. Lett. A 114, 217–221.

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

6 Aha, D., Kibler, D., and Albert, M. (1991), Instance-based learning algorithms. Machine Learning, 6, 37-66.

7 Alvisi, S., Mascellani, G., Franchini, M., Bárdossy, A. (2006), Water level forecasting through fuzzy logic and
8 artificial neural network approaches. Hydrology and Earth System Sciences, 10, 1-17.

9 ASCE. (2000a), Artificial neural networks in hydrology 1: Preliminary concepts. Journal of Hydrologic
10 Engineering, 5(2), 115-123.

11 Broomhead, D.S., and King, G.P., (1986), Extracting qualitative dynamics from experimental data, Physica D, 20,
12 217.

13 Carlson, R.F., MacCormick,A.J.A. and Watts D.G., (1970), Application of linear models to four annual
14 streamflow series, Water Resour. Res., 6(4), 1070-1078.

15 Casdagli, M., Eubank,S., Farmer,J.D., and Gibson, J. (1991), State space reconstruction in the presence of noise,
16 Physica D, 51, 52-98.

17 Chau, K.W., Wu, C.L., Li, Y.S., (2005). Comparison of several flood forecasting models in Yangtze River. J.
18 Hydrol. Eng. 10 (6), 485–491.

19 Dawson, C. W., and Wilby, R. L., (2001), Hydrological Modeling Using Artificial Neural Networks. Progress in
20 Physical Geography, 25(1), 80-108.
21 De Vos, N.J. and Rientjes, T.H.M., (2005), Constraints of artificial neural networks for rainfall -runoff modeling:
22 trade-offs in hydrological state representation and model evaluation. Hydrology and Earth System Sciences, 9,
23 111-126.

24 Elshorbagy, A., Simonovic,S.P., and Panu, U.S., (2002), Estimation of missing stream flow data using principles
25 of chaos theory, Journal of Hydrology, 255, 123–133.

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

28 Essex,C., and Nerenberg, M.A.H., (1991), Proc. R. Soc. London, Series A 435,287.

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

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

32 Ghilardi, P., and Rosso, R., (1990), Comment on chaos in rainfall. Water Resour. Res. 26 (8), 1837–1839.

33 Grassberger, P., and Procaccia, I., (1983), Measuring the strangeness of strange attractors, Physica 9D, 189-208.

34 Haltiner, J.P., and Salas, J.D., (1988), Short-term forecasting of snowmelt discharge using ARMAX models,
35 Water Resources Bulletin, 24(5), 1083-1089.

Hegger, R., Kantz, H. and Schreiber, T., (1999), Practical implementation of nonlinear time series methods: The TISEAN package. CHAOS, 9(2), 413-435.

Hong, S.Z., and Hong,S.M., (1994), An amendment to the fundamental limits on dimension calculations. Fractals 2 (1), 123–125.

Huang, W.R., Xu, B., and Hilton, A. (2004), Forecasting Flows in Apalachicola River Using Neural Networks, Hydrological Processes, 18, 2545-2564.

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

Jayawardena, A.W., and Gurung, A.B., (2000), Noise reduction and prediction of hydro-meteorological time series: dynamical systems approach vs. stochastic approach. Journal of Hydrology, 228, 242–264.

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.

Kantz, H., and Schreiber, T., (2004), Nonlinear time series analysis (2nd edition). Springer.

Karlsson, M., and Yakowitz, S. (1987a), Nearest-neighbor methods for nonparametric rainfall–runoff forecasting. Water Resources Research, 23 (7), 1300-1308.

Karlsson, M., and Yakowitz, S., (1987b), Rainfall–runoff forecasting methods, old and new. Stochastic Hydrology and Hydraulics, 1, 303-318.

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

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

Kothyari, U.C., and Singh, V.P., (1999), A multiple-input single-output model for flow forecasting, Journal of Hydrology, 220, 12–26.

Koutsoyiannis, D., and Pachakis, D., (1996). Deterministic chaos versus stochasticity in analysis and modeling of point rainfall series. J. Geophys. Res. 101 (D21), 26441–26451.

Kugiumtzis, D., (1996), State space reconstruction parameters in the analysis of chaotic time series - the role of the time window length, Physica D 95, 13-28.

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

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.

Liong, S.Y., and Sivapragasam, C., (2002), Flood stage forecasting with support vector machines, Journal of American Water Resour, 38(1),173 -186.

Mack, Y. P. and Rosenblatt, M. (1979), Multivariate k-nearest neighbor density estimates. Journal of Multivariate Analysis, 9, 1-15.

María, C.M., Wenceslao, G.M., Manuel, F.B., José,M.P.S., and Román, L.C., (2004), Modelling of the monthly and daily behaviour of the discharge of the Xallas river using Box–Jenkins and neural networks methods, J. Hydrol., 296,38–58.

Marques, C.A.F., Ferreira, J. Rocha, A. Castanheira, J. P. Gonçalves, 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.

Martinerie, J.M., Albano, A.M., Mees, A.I., and Rapp, P.E., (1992), Mutual information, strange attractors, and the optimal estimation of dimension, Phys. Rev. A, 45, 7058.

Mees, A.I., Rapp, P.E., and Jennings, L.S., (1987), Singular-value decomposition and embedding dimension, Phys. Rev. A, 36, 340.

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

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

Pasternack, G.B., (1999), Does the river run wild? Assessing chaos in hydrological systems. Adv. Water Resour. 23, 253–260.

Procaccia, I., (1988), Complex or just complicated? Nature, 333, 498–499.

Ruelle, D., (1990), Proc. R. Soc. London, Ser. A 427,241.

Salas, J.D., Tabios III, G.Q., and Bartolini, P., (1985), Approaches to multivariate modeling of water resources time series, Water Resources Bulletin, 21(4), 683-708.

Sauer, T., Yorke, J. A., and Casdagli, M., (1991), Embedology, J. Stat. Phys., 65, 579– 616.

Schertzer, D., Tchiguirinskaia, I., Lovejoy, S., Hubert, P., Bendjoudi, H., Larchevêque, M., (2002), Which chaos in the rainfall runoff process? A discussion on Evidence of chaos in the rainfall-runoff process by Sivakumar. Hydrol. Sci. J. 47 (1), 139–147.

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

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

Sivakumar, B., Liong, S.Y., and Liaw, C.Y., (1998), Evidence of chaotic behavior in Singapore rainfall, Journal of the American water resources association, 34(2), 301-310.

Sivakumar, B., Berndtsson, R., Olsson, J., Jinno, K., (2001), Evidence of chaos in rainfall-runoff process. Hydrological Sciences Journal, 46(1), 131-146.

Solomatine, D.P., Maskey, M., and Shrestha, D.L., (2008), Instance-based learning compared to other data-driven methods in hydrological forecasting, Hydrological Processes, 22(2), 275-287.

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.

Takens, F., (1981), Dynamical systems and Turbulence, Lecture note in Mathematics, Vol. 898, Springer, New York.

Theiler, J., (1986), Spurious dimension from correlation algorithms applied to limited time-series data. Phys. Rev. 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, J of Hydrology 239, 132–147.

Tsonis, A.A., (1992), Chaos: From Theory to Applications, Plenum Press, New York.

Wang,W., Van Gelder, P.H.A.J.M., Vrijling, J.K., Ma, J., (2006a), Testing for nonlinearity of streamflow processes at different timescales. Journal of Hydrology, 322, 247-268.

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

Yakowitz, S. (1987), Nearest neighbor method for time series analysis. Journal of Time Series Analysis, 8 (2), 235-247.

Yakowitz, S., and Karlsson, M., (1987), Nearest neighbor methods for time series with application to rainfall-discharge prediction, In: Stochastic Hydrology, (eds.) J.B. MacNeil, and G.D. Umphrey, pp. 149-160, Dordrecht ; Boston : D. Reidel ; Norwell, Mass.

Yu, P.S., and Tseng, T.Y. (1996), A model to forecast flow with uncertainty analysis, Hydrological Sciences-Journal, 41(3), 327-344.

Yu, X.Y., Liong, S.Y. and Babovic, V., (2004), EC-SVM approach for real-time hydrologic forecasting, Journal of Hydroinformatics, 6(3), 209-233.
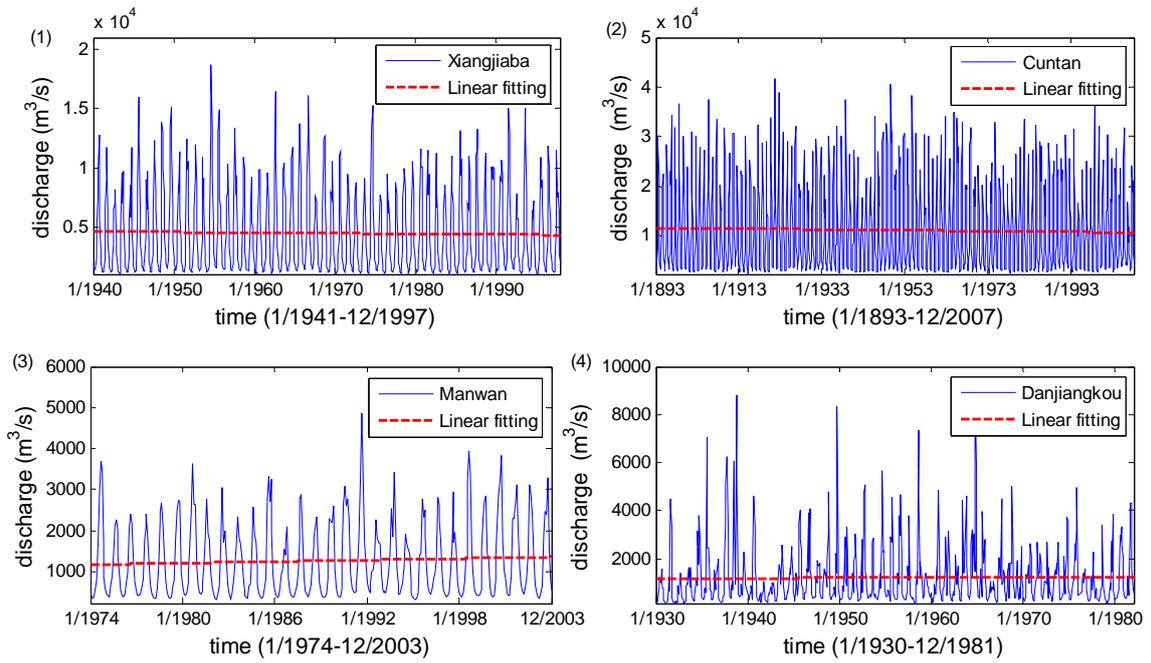
Figure 1. Monthly discharge series of (1) Xiangjiaba, (2) Cuntan, (3) Manwan, and (4) Danjiangkou
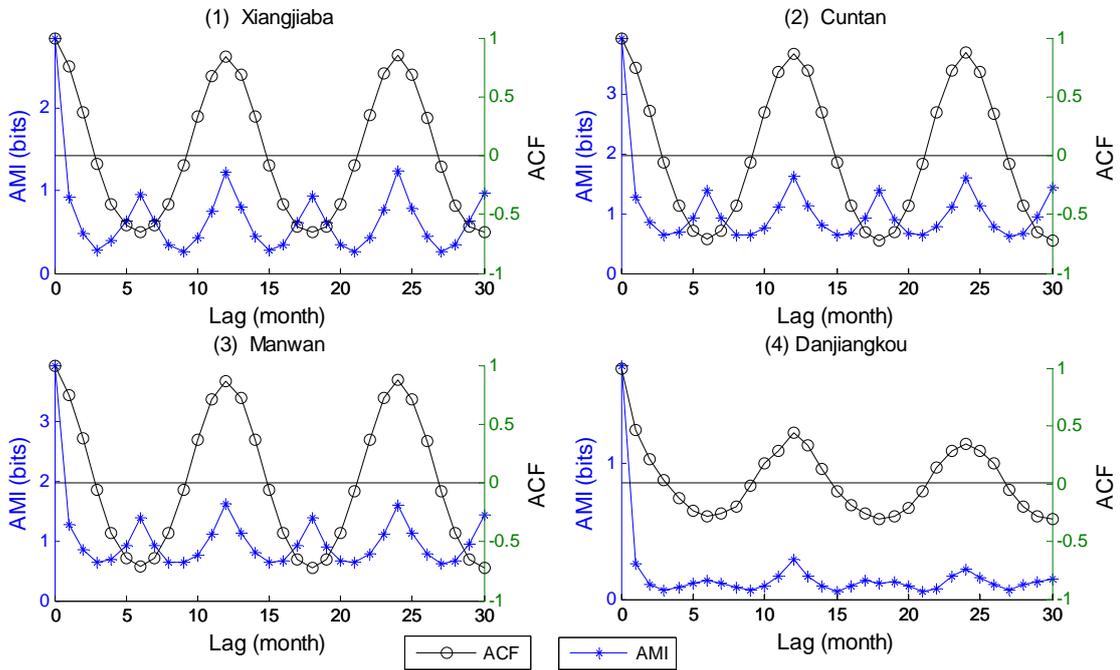
Figure 2. ACF and AMI of (1) Xiangjiaba, (2) Cuntan, (3) Manwan, and (4) Danjiangkou
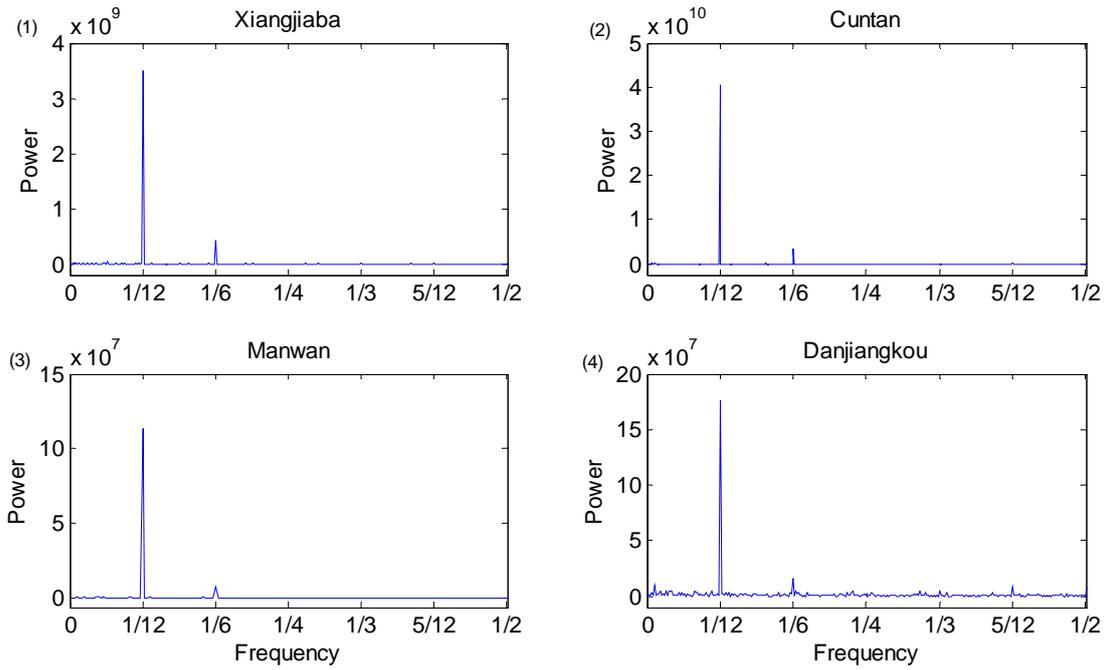
1
2
Figure 3. Periodogram of (1) Xiangjiaba, (2) Cuntan (3) Manwan, and (4) Danjiangkou
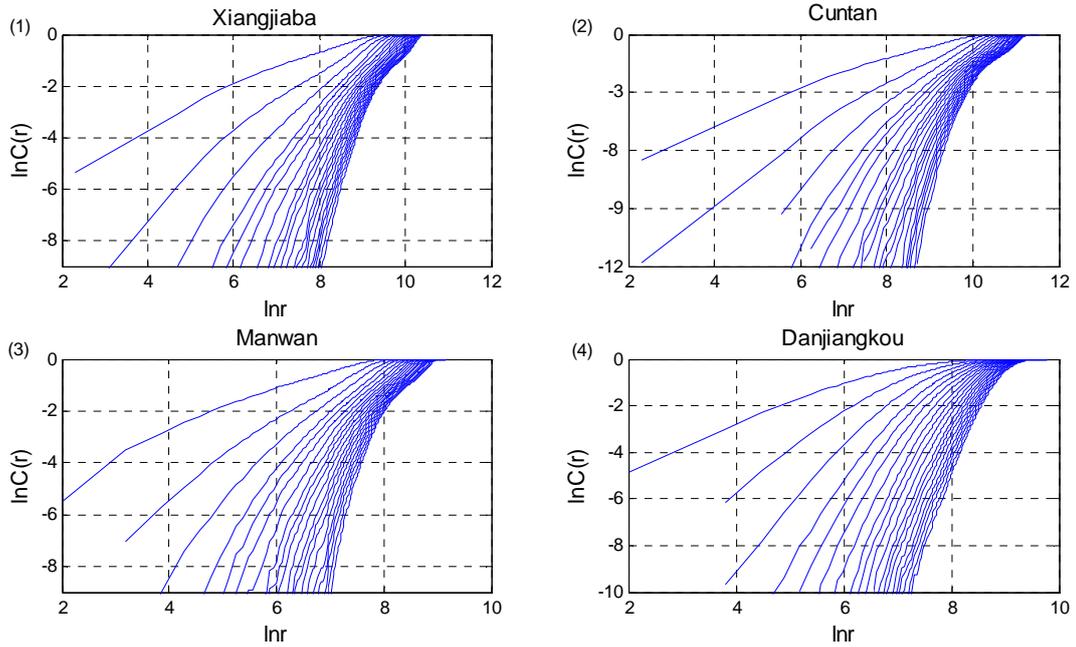
3
4   Figure 4. $\ln C(r)$ versus $\ln r$ plot for (1) Xiangjiaba, (2) Cuntan (3) Manwan, and (4) Danjiangkou
5

19
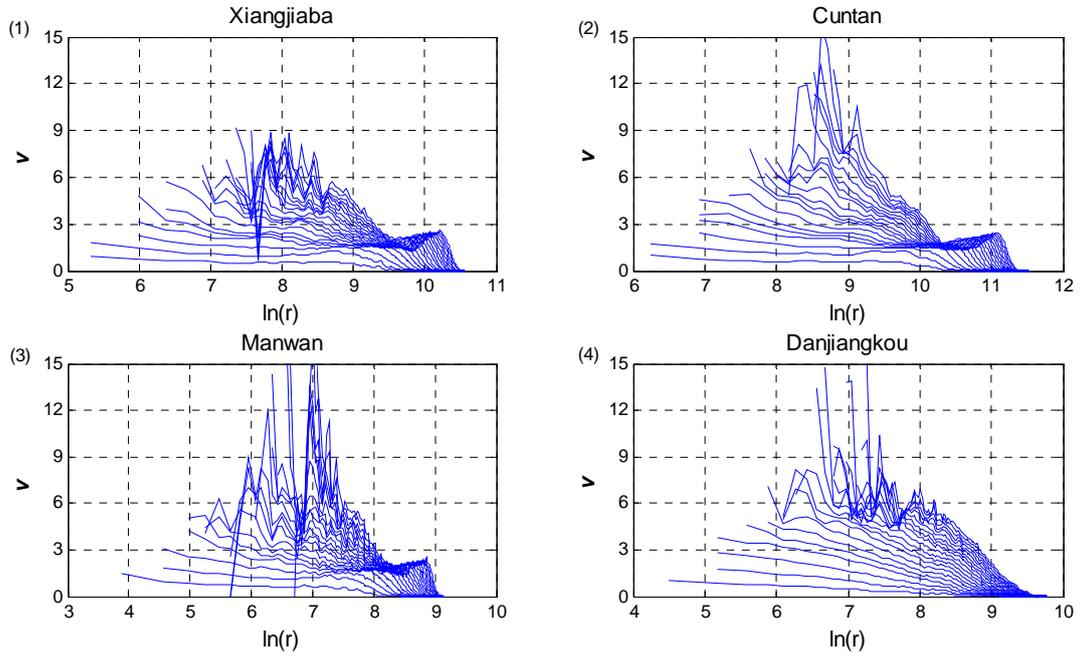
1
2   Figure 5. The estimation of correlation dimension (d_2) for (1) Xiangjiaba, (2) Cuntan (3) Manwan, and (4)
3   Danjiangkou. $\tau = 3$ and m is at the interval of [1,20] (increasing from bottom to top in each pane).
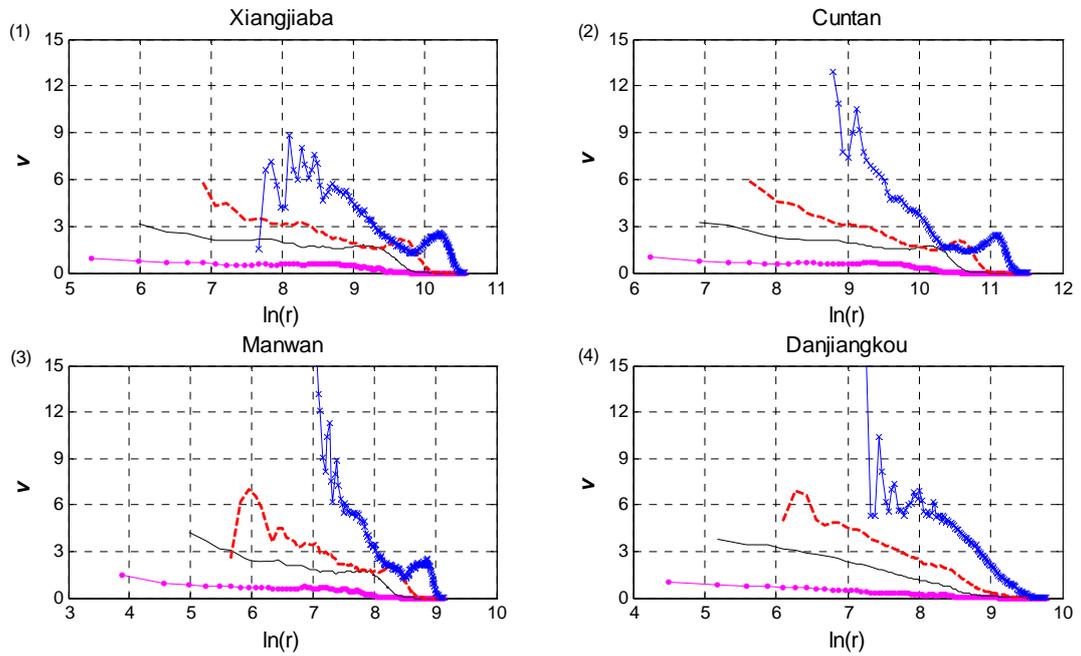
4
5   Figure 6. The estimation of correlation dimension (d_2) for (1) Xiangjiaba, (2) Cuntan (3) Manwan, and (4)
6   Danjiangkou. $\tau = 3$ and m is at 1(solid line with dots),4(solid line),8 (dashed line), and 20 (solid line with crosses).
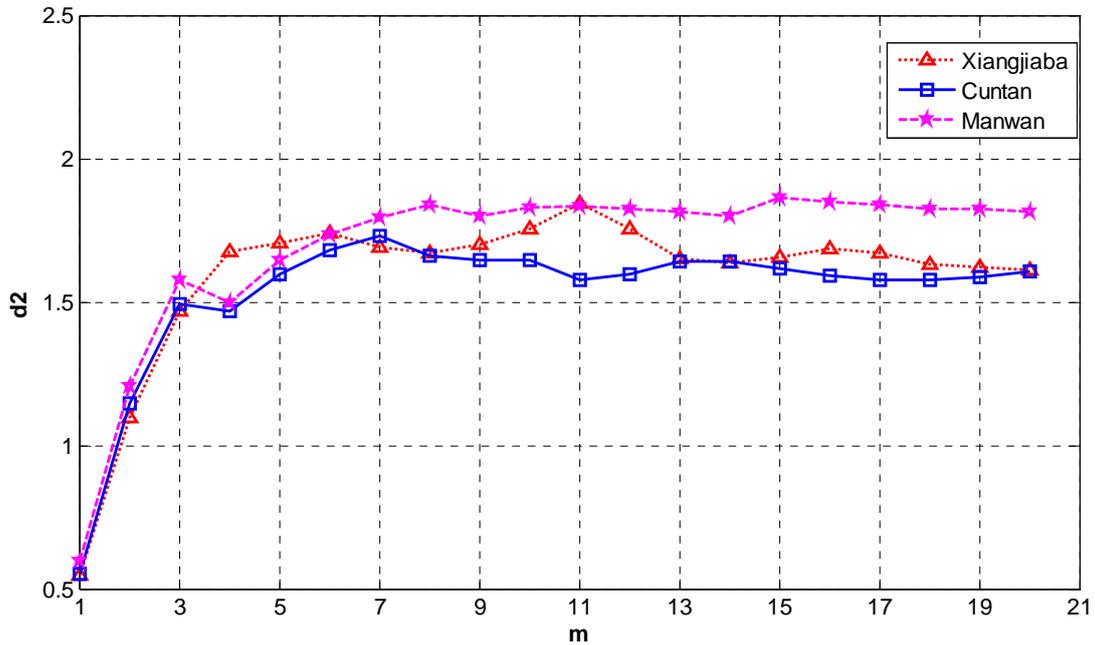
1
2   Figure 7. Relationship between d_2 and m for monthly discharges of Xiangjiaba, Cuntan, and Manwan
3
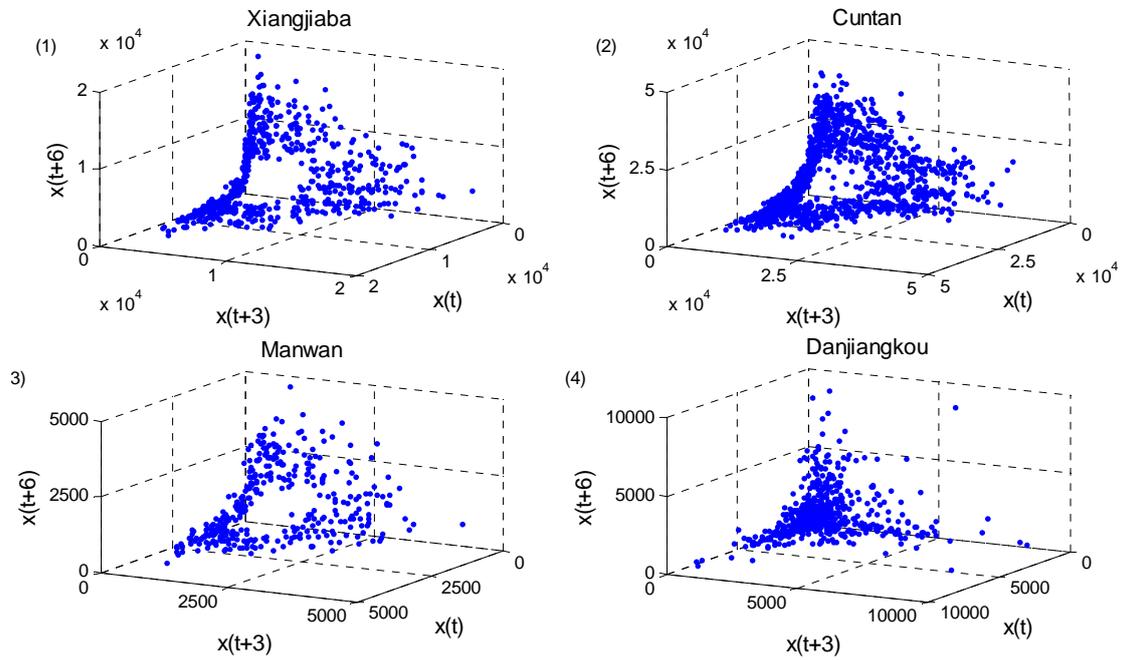
4
5   Figure 8. Phase portraits for (1) Xiangjiaba, (2) Cuntan (3) Manwan, and (4) Danjiangkou when $m=3$ and $\tau=3$.
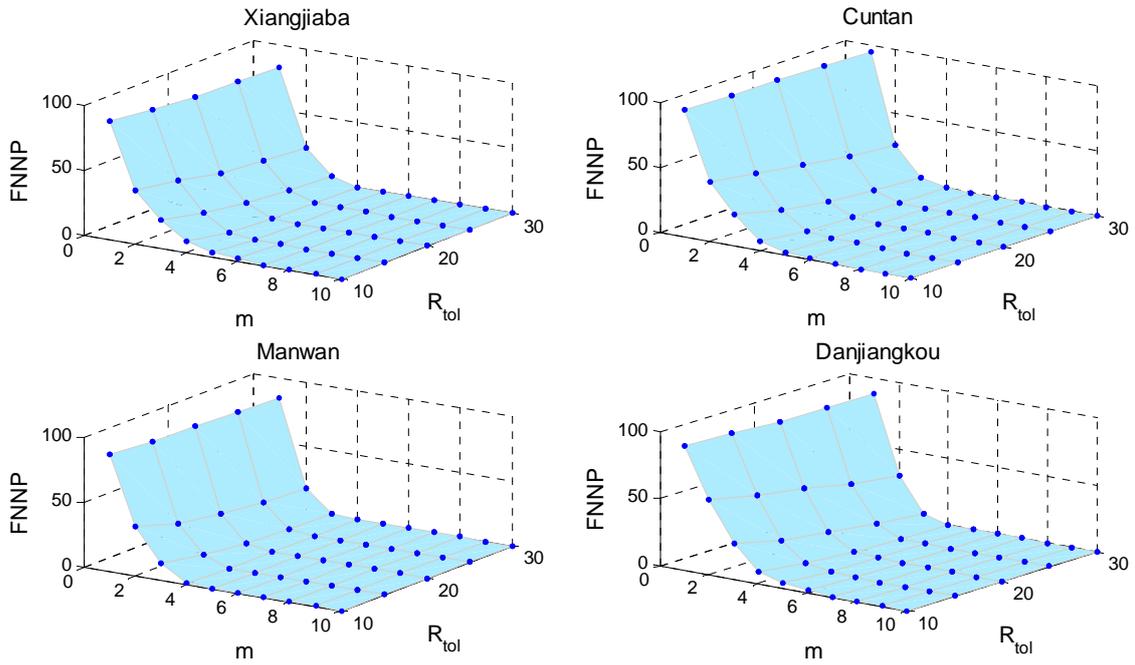
1
2   Figure 9. FNNP plots for (1) Xiangjiaba, (2) Cuntan (3) Manwan, and (4) Danjiangkou when $\tau = 3$ and R_{tol} is 10
3   to 30.
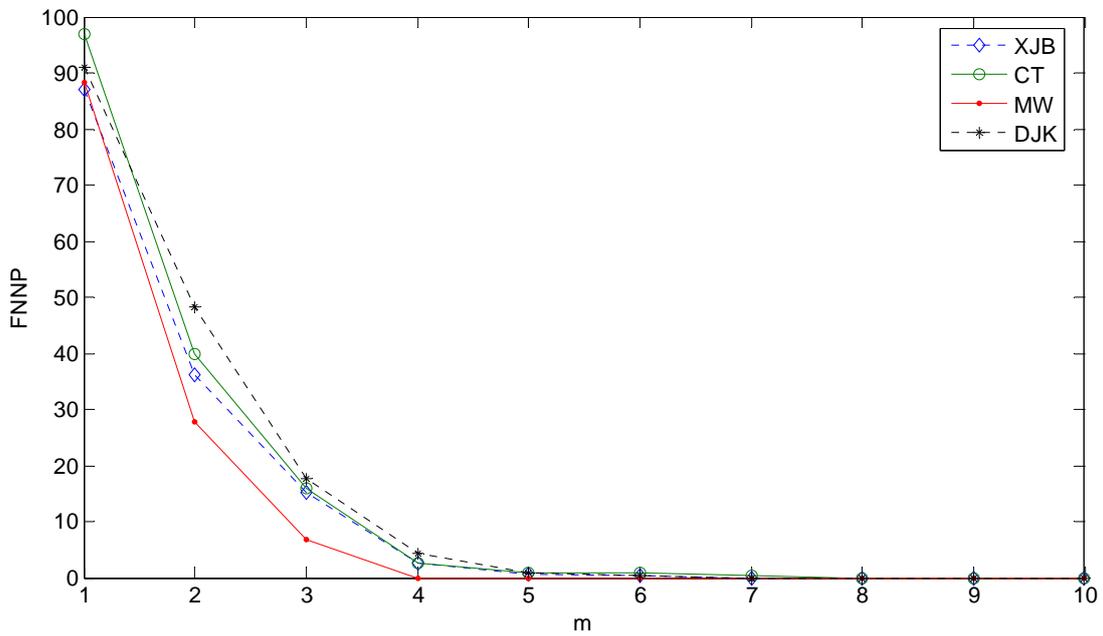
4
5   Figure 10. FNNP for Xiangjiaba, Cuntan, Manwan, and Danjiangkou when $\tau = 3$ and $R_{tol} = 15$.
6
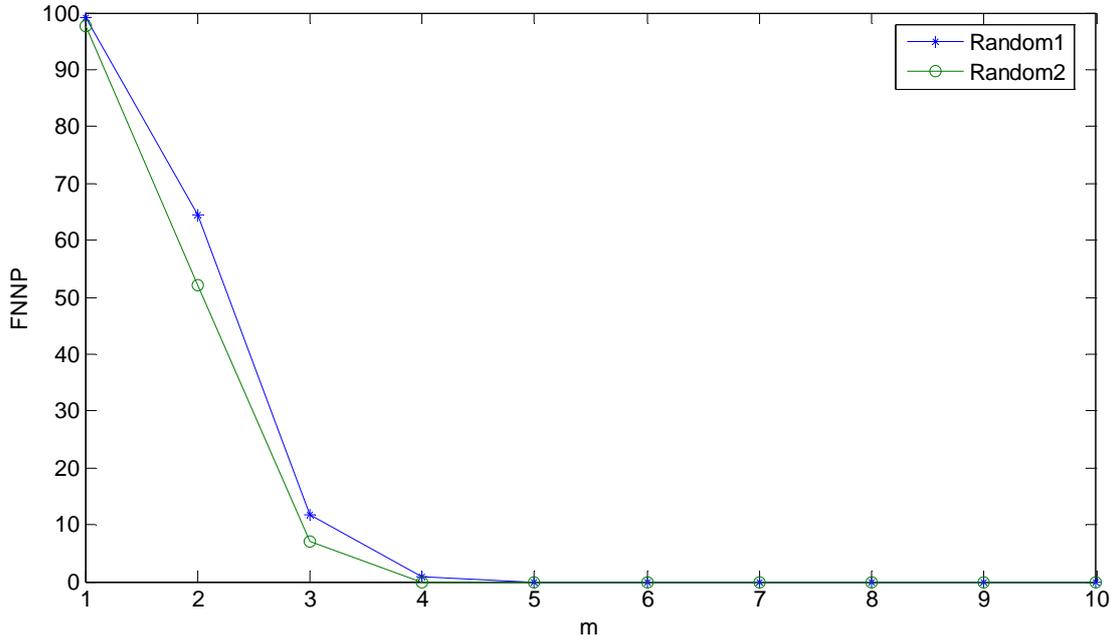
1
2   Figure 11. FNNP for Random process 1 with a data size of 1380 and Random process 2 with a data size of 696
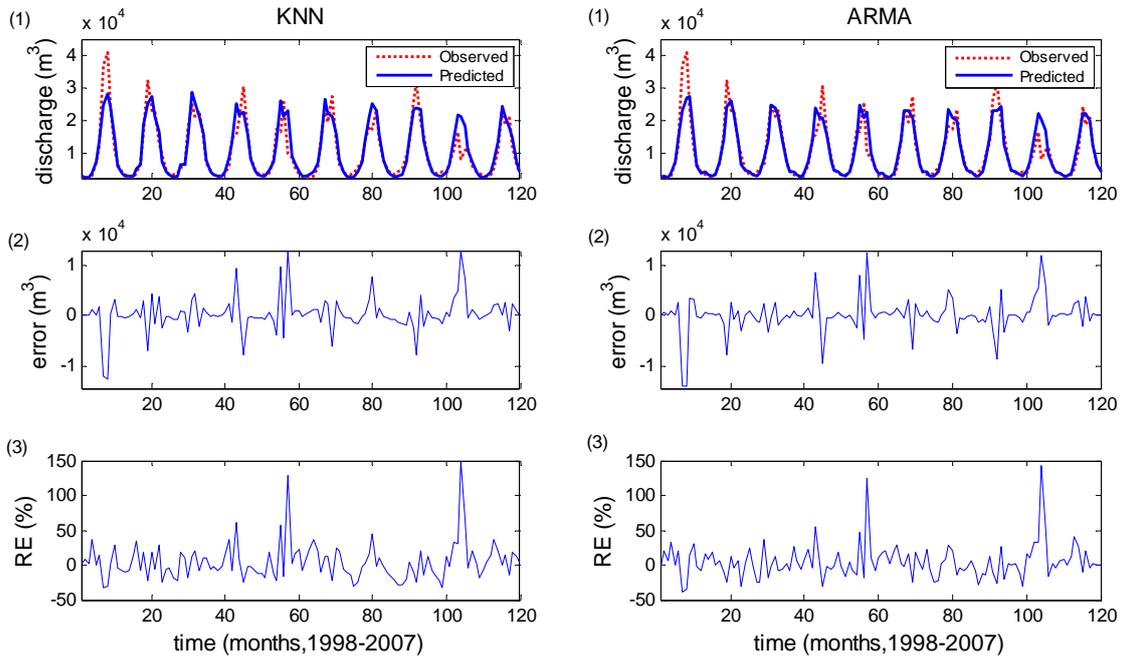

3
4   Figure 12. KNN (left column) and ARMA (right column) model results for one-day-ahead forecast on Cuntan:
5   (1) observed vs. predicted (2) error (observed – predicted) and (3) RE (representing Relative Error)

23

Figure 13. ANN (left column) and ANN-PSR (right column) model results for one-day-ahead forecast on Cuntan:
(1) observed vs. predicted (2) error (observed – predicted) and (3) RE (representing Relative Error)
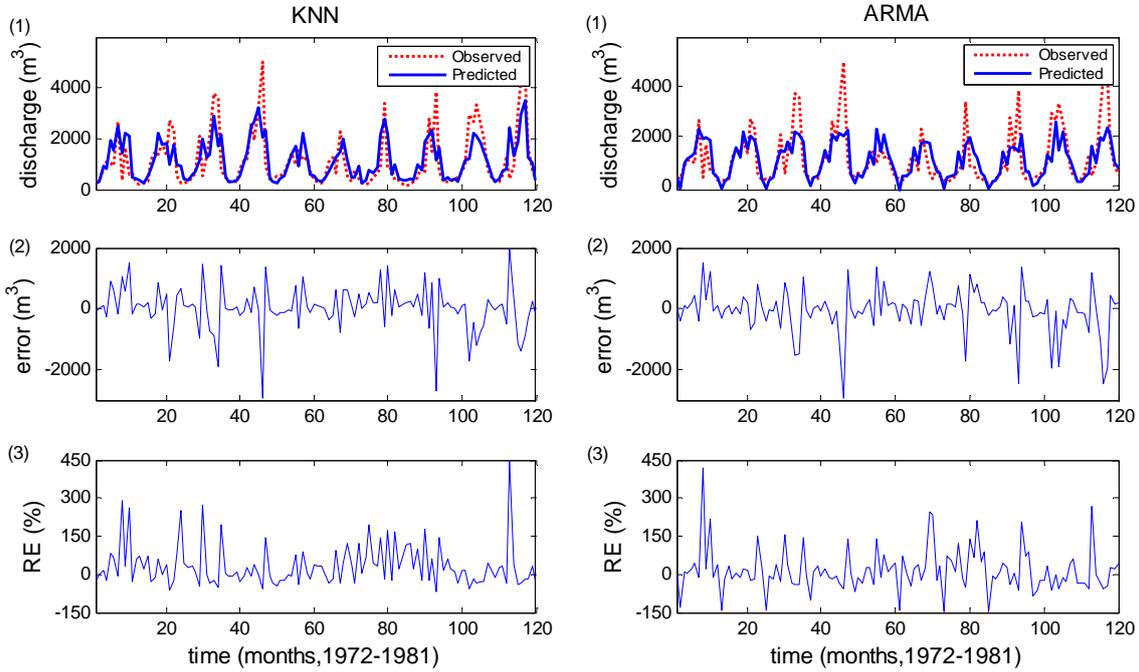

Figure 14. KNN (left lists) and ARMA (right lists) model results for one-day-ahead forecast on Danjiangkou:
(1) observed vs. predicted (2) error (observed – predicted) and (3) RE (representing Relative Error)
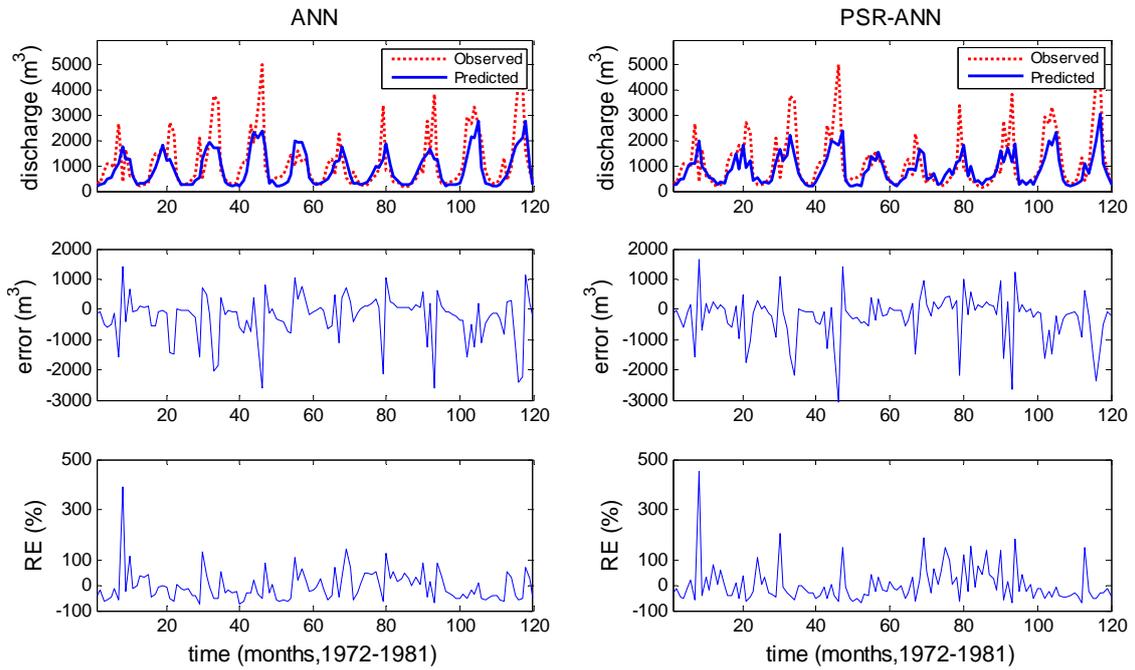
1
2 Figure 15. KNN (left lists) and ARMA (right lists) model results for one-day-ahead forecast on Danjiangkou:
3   (1) observed vs. predicted (2) error (observed – predicted) and (3) RE (representing Relative Error)
4
5
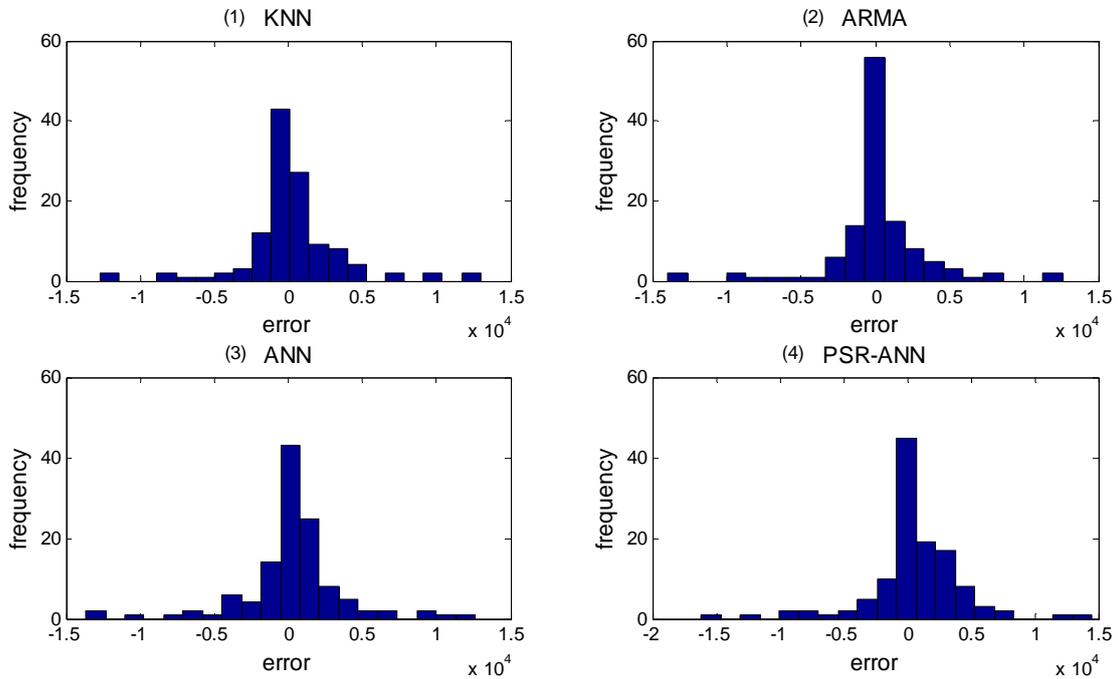
6
7 Figure 16. Histograms of errors of (1) K-NN, (2) ARMA, (3) ANN, and (4) ANN-PSR for Cuntan

25

1
2
3
4
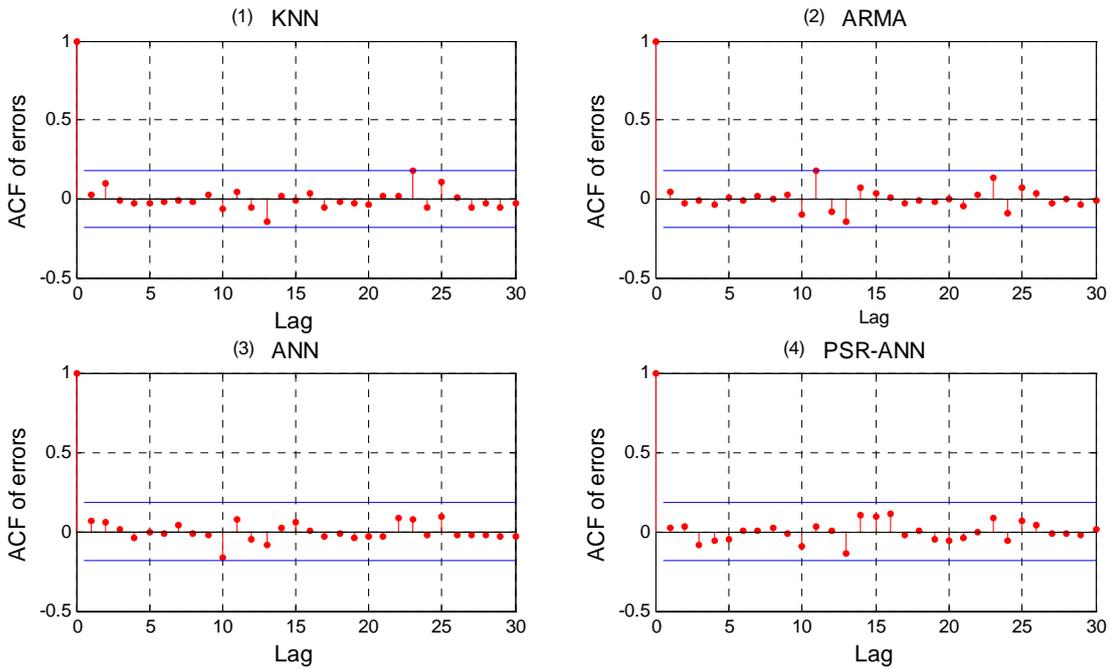
Figure 17. ACF of errors of (1) K-NN, (2) ARMA, (3) ANN, and (4) ANN-PSR for Cuntan


5
6
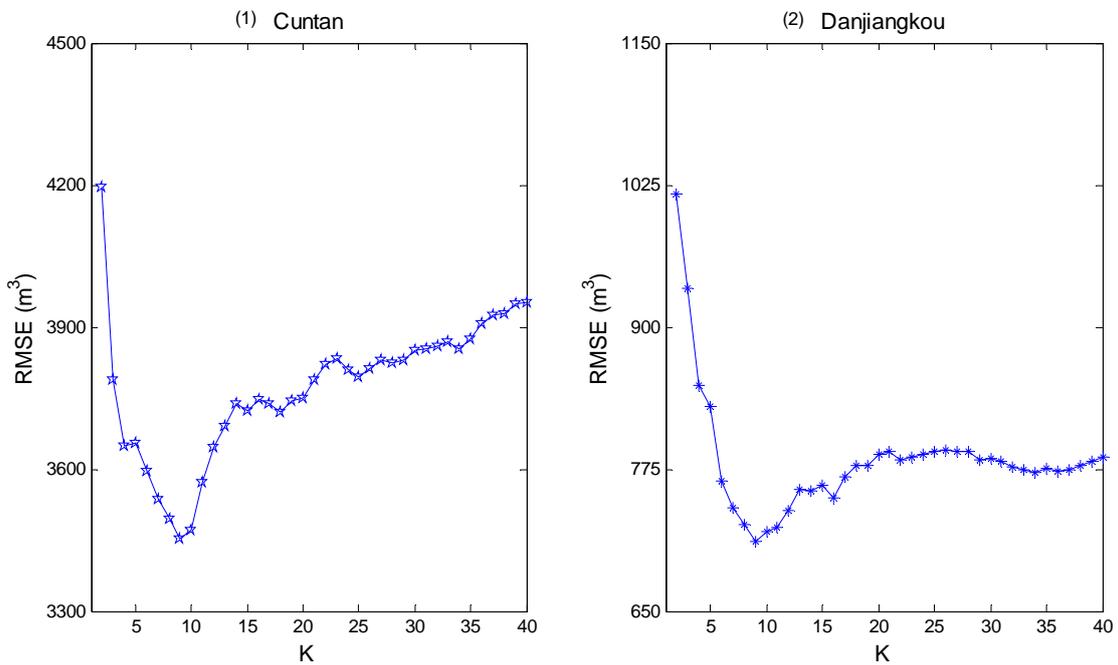
Figure 18. The check of stability of k in KNN method for (1) Cunan and (2) Danjiangkou
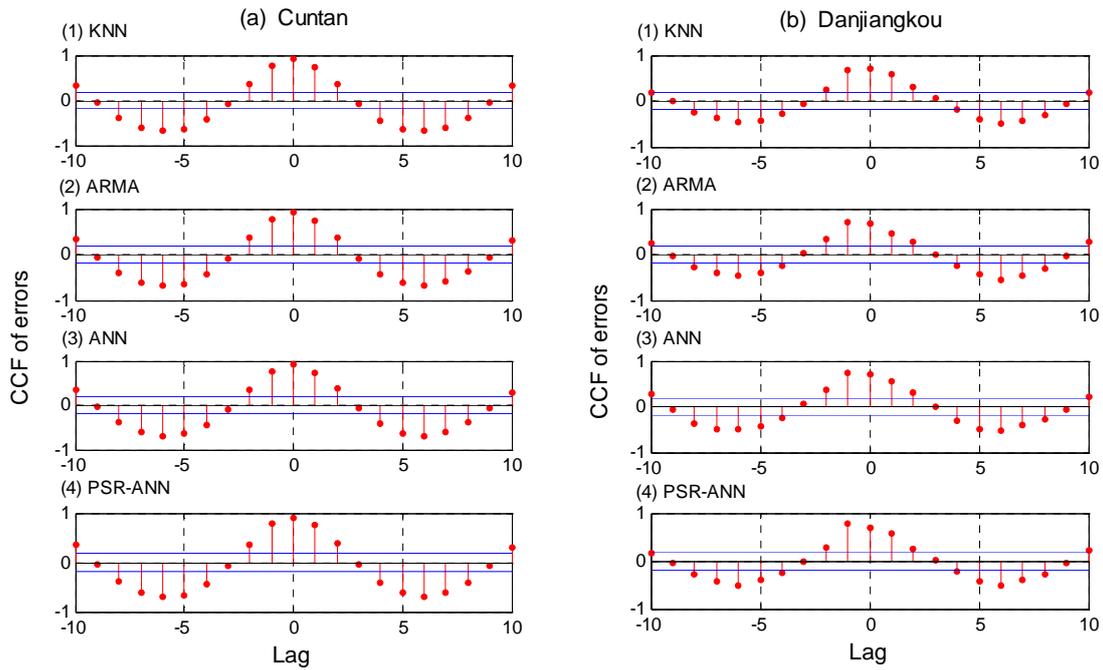
Figure 19. The check of the prediction lag effect via CCF between observed and predicted values. (a) for Cuntan and (b) for Danjiangkou ((1) K-NN, (2) ARMA, (3) ANN, and (4) ANN-PSR)
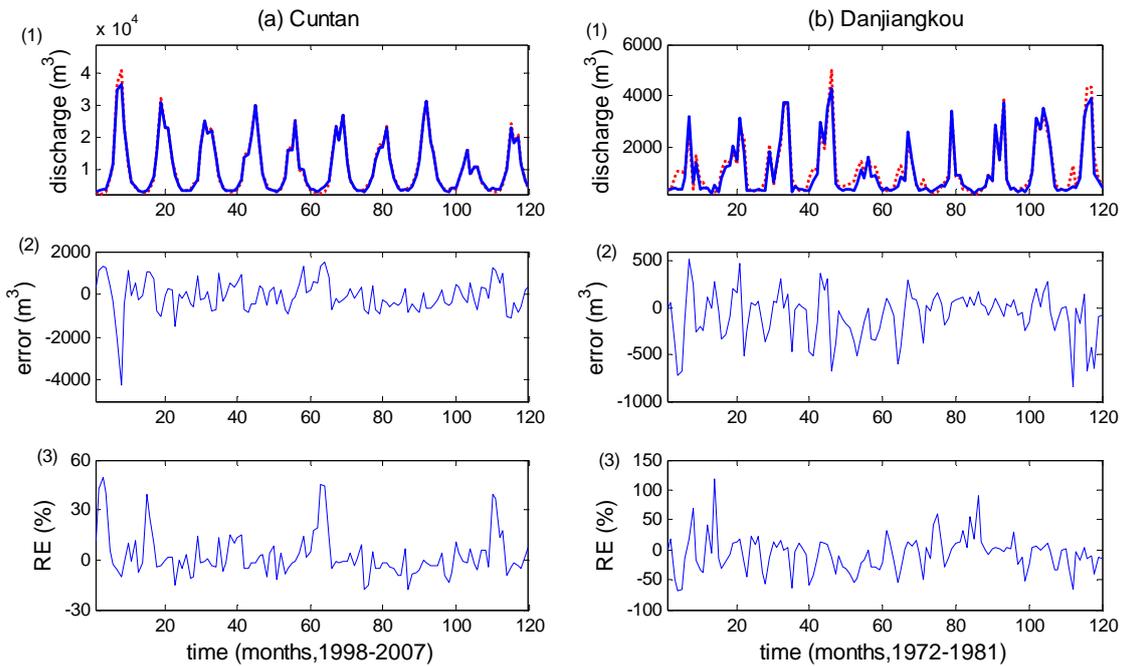

Figure 20. MA-ANN model results of one-day-ahead forecast for (a) Cuntan and (b) Danjiangkou. ((1) observed vs. predicted (2) error (observed – predicted) and (3) RE (representing Relative Error))

27
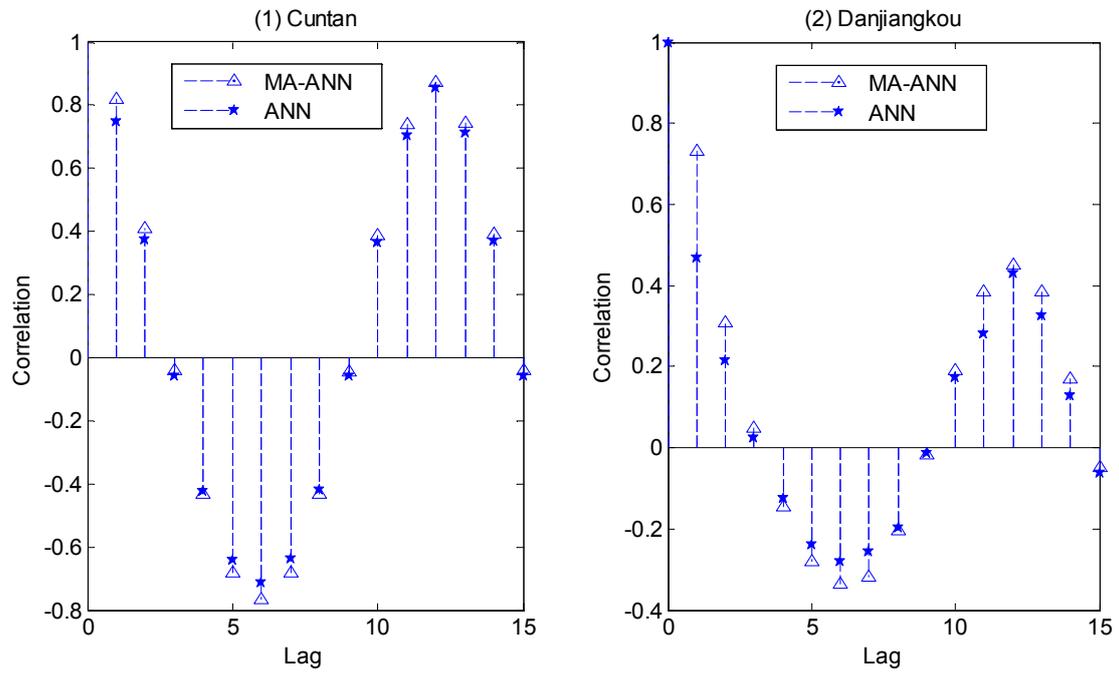
1
2   Figure 21. CCFs between inputs and outputs from the MA-ANN model and the ANN model for (1) Cuntan and (2)
3   Danjiangkou.
4

1
2　　　　　　Table 1. Statistical parameters for training, testing, and validation sets

Watershed and datasets	Statistical parameters					Watershed area and data period
	Mean	Standard deviation	Minimum	Maximum	Range	
Xiangjiaba						
Original data	4538.2	3671.5	1100	18700	17600	Area: 45.88×10^{4}km^2
Training set	4678.2	3832.4	1130	18700	17570	
Cross-validation set	4297.6	3411.2	1110	15300	14190	Data period: January 1940 to
Testing set	4475.2	3552.3	1100	15000	13900	December 1997
Cuntan						
Original data	11118	8649.7	2250	41700	39450	Area:
Training set	11342	8892.6	2250	41700	39450	
Cross-validation set	10766	8195.7	2420	34800	32380	Data period: January 1893 to
Testing set	10774	8488.7	2610	40800	38190	December 2007
Manwan						
Original data	1268.2	908.66	316	4860	4544	Area: 11.45×10^{4}km^2
Training set	1192.7	862.29	329	3710	3381	
Cross-validation set	1371.3	944.26	373	4860	4487	Data period: January 1974 to
Testing set	1300.1	943.19	316	3944	3628	December 2003
Danjiangkou						
Original data	1203.3	1285.1	139	9010	8871	Area: 9.5×10^{4}km^2
Training set	1203	1348.1	140	8800	8660	
Cross-validation set	1256.1	1310.7	139	9010	8871	Data period: January 1930 to
Testing set	1130	1053.5	159	4980	4821	December 1981

3

4　Table 2 Statistics of ANN model performance using various training algorithms with 30 time
5　　　　　　　　　　　　　　　　　　　　runs

Training algorithm and performance	Statistical parameters				
	Mean	Standard deviation	Minimum	Maximum	Range
LM					
RMSE	844.9	114.1	774.6	1429.5	654.9
CE	0.3	0.2	-0.9	0.5	1.3
PI	0.3	0.2	-1.0	0.4	1.4
Training time(sec.)	0.7	0.2	0.5	1.5	0.9
PSO					
RMSE	795.9	29.1	763.1	912.5	149.4
CE	0.4	0.0	0.2	0.5	0.2
PI	0.4	0.0	0.2	0.4	0.2
Training time(sec)	74.6	1.0	70.9	76.0	5.0
LM-GA					
RMSE	808.7	24.1	760.1	861.0	100.9

CE	0.4	0.0	0.3	0.5	0.1
PI	0.4	0.0	0.3	0.4	0.1
Training time(sec.)	19.3	0.7	18.3	21.0	2.7

Table 3. Comparison of model performance for one-step-ahead forecast on validation data and model parameters

Watershed	Model	RMSE	CE	PI	Model parameters		
					Architecture	(τ, m, k)	(p, q)
Cuntan							
	ARMA	3498.60	0.83	0.69			(11,11)
	ANN*	3557.10	0.82	0.68	(12,13,1)		
	K-NN	3495.50	0.83	0.69		(5,7,8)	
	ANN-PSR*	3878.60	0.79	0.61	(7,9,1)		
Danjiangkou							
	ARMA	763.81	0.47	0.44			(11,10)
	ANN*	794.98	0.43	0.39	(12,5,1)		
	K-NN	741.58	0.50	0.47		(5,6,7)	
	ANN-PSR*	814.74	0.40	0.36	(6,9,1)		

Note:* the index of performance is an average over the best 10 runs of total 30 runs.

Table 4. Improvement of model performance for one-step-ahead forecast from K-NN and MA-ANN models and model parameters

Watershed	Model	RMSE	CE	PI	Model parameters		
					Architecture	(τ, m, k)	Memory length for the moving average
Cuntan							
	MA-ANN	687.37	0.99	0.99	(12,13,1)		3
	K-NN	3453.90	0.83	0.69		(5,7,9)	
Danjiangkou							
	MA-ANN	268.71	0.93	0.93	(12,5,1)		6
	K-NN	710.73	0.54	0.51		(5,6,9)	