



# Monthly Streamflow Forecasting Using ELM-IPSO Based on Phase Space Reconstruction

Yan Jiang<sup>1</sup> · Xin Bao<sup>1,2</sup> · Shaonan Hao<sup>1</sup> · Hongtao Zhao<sup>1</sup> · Xuyong Li<sup>1</sup> · Xianing Wu<sup>3</sup>

Received: 24 May 2019 / Accepted: 21 July 2020 /

Published online: 4 August 2020

© Springer Nature B.V. 2020

## Abstract

We have developed a hybrid model that integrates chaos theory and an extreme learning machine with optimal parameters selected using an improved particle swarm optimization (ELM-IPSO) for monthly runoff analysis and prediction. Monthly streamflow data covering a period of 55 years from Daiying hydrological station in the Chaohe River basin in northern China were used for the study. The Lyapunov exponent, the correlation dimension method, and the nonlinear prediction method were used to characterize the streamflow data. With the time series of the reconstructed phase space matrix as input variables, an improved particle swarm optimization was used to improve the performance of the extreme learning machine. Finally, the optimal chaotic ensemble learning model for monthly streamflow prediction was obtained. The accuracy of the predictions of the streamflow series (linear correlation coefficient of about 0.89 and efficiency coefficient of about 0.78) indicate the validity of our approach for predicting streamflow dynamics. The developed method had a higher prediction accuracy compared with an auto-regression method, an artificial neural network, an extreme learning machine with genetic algorithm and with PSO algorithm, suggesting that ELM-IPSO is an efficient method for monthly streamflow prediction.

**Keywords** Streamflow prediction · Chaohe River basin · Chaotic dynamic characteristics · Phase space reconstruction · Extreme learning machine · Improved particle swarm optimization algorithm

---

✉ Yan Jiang  
yanjiang@rcees.ac.cn

<sup>1</sup> State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> PowerChina Resources Limited, Beijing 100044, China

## 1 Introduction

Flood control, drought relief, and the optimal utilization of water resources require accurate prediction of streamflow. However, the hydrological process is extremely complex and difficult to predict, especially in the medium and long term because of human impact, changing climatic conditions, and the geographical environment (Vicente-Guillén et al. 2012). A large number of researchers are devoted to understanding the dynamics of rainfall-runoff process (Bradford et al. 1991; Duan et al. 1992; Huang et al. 2014). In the past, the hydrological process was regarded as stochastic (Sivakumar et al. 2001). With the rapid development of nonlinear science, nonlinear time series analysis has brought a significant method revolution. The “science of chaos” has found applications in almost all the natural sciences, including hydrological sciences (Islam and Sivakumar 2002). Even simple deterministic systems can display complex or chaotic behavior. It is now believed that the nonlinear chaotic model can better describe the complex hydrological dynamic process (Sivakumar 2000), and chaos theory has become increasingly common in the study of the dynamics of hydrological process (Hu et al. 2013; Ouyang et al. 2016; Hong et al. 2016; Zhao et al. 2017).

Many researches have investigated chaotic behavior of hydrological processes (Mohammad 2016) by analyzing streamflow series using the runoff coefficient (Sivakumar et al. 2001), the exponent method (Xu et al. 2009), the correlation dimension method (Labat et al. 2016), and several independent methods, techniques and tools (Kedra 2013). Some studies have also used nonlinear chaotic methods to predict streamflow as univariate series (Porporato and Ridolfi 1997; Islam and Sivakumar 2002; Zhou et al. 2018) and as multivariate series incorporating information from other time series (Han et al. 2017), with chaos theory integrated using various approaches including local autoregressive polynomial methods (Bordignon and Lisi 2000), local approximations (Islam and Sivakumar 2002), genetic programming (Ghorbani et al. 2018), and artificial neural networks (ANN) (Khan et al. 2005; Dhanya 2010). Neural networks are particularly useful for forecasting because they deal well with the nonlinearity and instability of hydrological time series when the input vectors are designed using the phase space reconstruction method (Peng et al. 2017). There have been many achievements in the application of the ANN technique. However, Common ANNs are highly dependent on the iterative tuning of model parameters and the initial values of weights and biases, which easily lead to the instability of forecasting result. Therefore, employing different heuristic searching algorithms becomes popular in the training process.

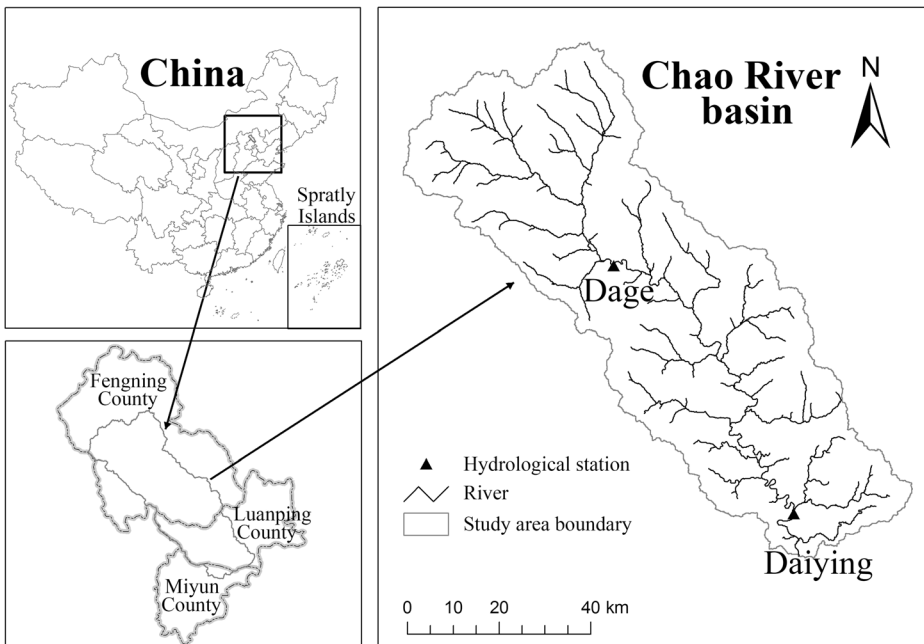
A new learning paradigm called an Extreme Learning Machine (ELM) has been proposed for training single hidden-layer feedforward neural networks (Huang et al. 2006). ELM is much faster and more adaptable than traditional ANN (Huang et al. 2015; Taormina and Chau 2015). In ELM, the biases of the hidden layer and the weights of the input and hidden layers are randomly generated, and the weights of the hidden and output layers can be determined directly using the Moore–Penrose generalized inverse method. An intelligent optimization algorithm is commonly used to optimize the biases and weights to reduce the influence of the parameters being randomly selected and improve the prediction performance of the ELM model. The particle swarm optimization approach (PSO) has many computational advantages over other optimization search methods (Jiang et al. 2010). However, the potential for premature convergence degrades

the performance of the algorithm and reduces the probability of finding global optima (Chu et al. 2010; Jiang et al. 2013). Using ideas drawn from population division and biological evolution, Jiang et al. (2015) proposed an improved particle swarm optimization (IPSO) to solve nonlinear optimization problems. In this paper, this method was applied for training an ELM to determine the optimal values of the biases and weights.

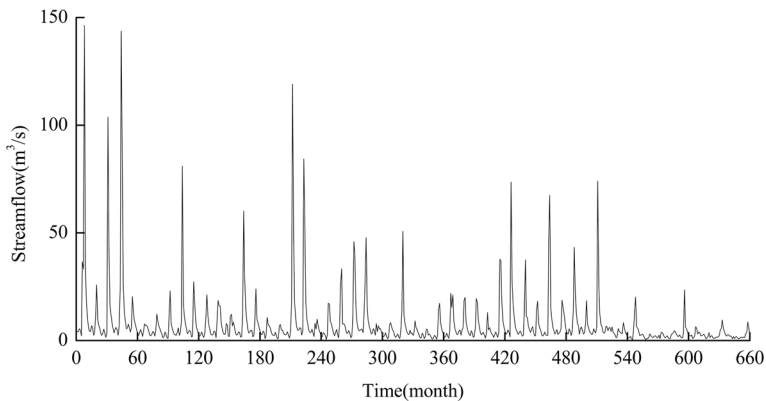
The objectives of the studies are as follows: (1) to analyze the chaotic behavior of the monthly streamflow series of the Chaohe River Basin using a variety of techniques. (2) to develop a hybrid model integrating chaos theory and extreme learning machines with optimal parameters selected by an improved particle swarm optimization (ELM-IPSO) to analyze and predict monthly streamflow.

## 2 Study Area and Data Used

The Chaohe River basin is located between  $40^{\circ}20' - 41^{\circ}27'N$  and  $116^{\circ}87' - 117^{\circ}34'E$ . The river originates in Fengning County, flows through Luanping County in Hebei Province, China, then runs down to Miyun County and empties into the Miyun Reservoir. The length of the river is about 170 km, and the average annual streamflow volume is  $18.04 \times 10^9 \text{ m}^3$ . Daiying hydrological station is a control station for the Chaohe River basin, and the catchment area upstream from the control section is  $4701 \text{ km}^2$  (Fig. 1). Monthly streamflow data of Daiying hydrological station, provided by the Beijing Water Authority, were used to analyze the chaos characteristics in the process of river flow. Figure 2 shows the variation in monthly streamflow for the period between January 1956 and December 2010.



**Fig. 1** Map of the Chaohe River Basin and the distribution of hydrological stations



**Fig. 2** Time series plot for monthly streamflow data

### 3 Methodology

Chaos theory was developed at the end of the nineteenth century. It deals with complex and unpredictable nonlinear systems (Dhanya and Kumar 2010). The essence of chaos is the sensitivity of the system to the change of initial conditions (Sivakumar 2004). Several studies have since applied ideas from chaos theory to understanding geophysical phenomena. The qualities that make a system chaotic are: (i) it is deterministic; (ii) it is sensitive to initial conditions; (iii) it is neither random nor disorderly.

#### 3.1 Phase Space Reconstruction

Phase space reconstruction is a useful tool for characterizing dynamical systems by a phase space diagram, which is essentially a coordinate system that has all the variables of the system as its basis. Each trajectory in the phase space diagram describes the evolution of the system, and each point represents the state of the system at a given time (Sivakumar 2000). All trajectories from different initial conditions in phase space will eventually converge to a subset, which is called the attractor of the system.

Phase space reconstruction was firstly proposed by Takens, who proved theoretically and by numerical simulation that state space reconstruction can preserve the geometric invariance of nonlinear dynamic systems (Takens 1981). Existing methods for phase space reconstruction include the method of time delays, the differential coordinate method, and the principal component analysis method, among which the method of time delays is the most popular. For a single variable time series  $x_1, x_2, \dots, x_n$ , its phase space reconstruction can be expressed as.

$$Y_i = (x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}), \quad i = 1, 2, \dots, n-(m-1)\tau, \quad (1)$$

Where  $m$  is called the embedding dimension,  $\tau$  is the delay time and  $n$  is the length of the time series. The calculation of the reconstruction parameters  $m$  and  $\tau$  is the key to using the delay coordinate method for phase space reconstruction.

Takens demonstrated that there exists an embedding dimension for  $m \geq 2d + 1$ , where  $d$  is the dimension of the dynamical system, for which regular trajectories (attractors) can be

constructed. According to Takens Theorem, the phase space can maintain the basic properties of the original state space. Therefore, phase space reconstruction is an effective tool to explore the characteristics of the dynamical system.

### 3.2 Identification of Chaotic Characteristics

Various techniques have been proposed for the identification of chaos including the Kolmogorov entropy method (Benettin et al. 1979), the correlation dimension method (Grassberger and Procaccia 1983), the Lyapunov exponent method (Wolf et al. 1985), the nonlinear prediction method (Farmer and Sidorowich 1987), the false nearest neighbor algorithm (Kennel et al. 1992), the method of redundancy (Paluš et al. 1995), and the surrogate data method (Schreiber and Schmitz 1996). Generally, we need to apply several methods to distinguish infallibly between a chaotic and stochastic system. In this paper, the correlation dimension method, the Lyapunov exponent method and the nonlinear prediction method were used to analyze the chaotic characteristics of streamflow series.

#### 3.2.1 Lyapunov Exponent

Lyapunov exponents are used to determine the chaotic characteristics of the system according to whether the phase trajectory has the features of diffusion motion. When the largest Lyapunov exponent is greater than 0, the system is chaotic.

The main methods for calculating the maximum Lyapunov exponent include Wolf’s algorithm (Wolf et al. 1985), the Jacobi matrices method (Sano and Sawada 1985), and the small data set method (Rosenstein et al. 1993). Based on the advantages of reliability, rapidity and accessible to an application, the small data set method was applied to calculate the maximum Lyapunov exponent in this paper.

Let  $Y_j$  and  $Y_{\hat{j}}$  be the reference point and the nearest neighbor of two trajectories in state space, then the distance between them is  $d_j(0) = \|Y_j - Y_{\hat{j}}\|$ .  $d_j(i) = \|Y_{j+i} - Y_{\hat{j}+i}\|$  will be the distance after  $i$  discrete-time steps. Hence, an exponential function  $d_j(i) \cong d_j(0)e^{\lambda(i \cdot \Delta t)}$  can describe the divergence form of trajectory with initial separation  $d_j(0)$ , where  $\Delta t$  is the sampling period of the time series and  $\lambda$  is the largest Lyapunov exponent. Therefore, it can be obtained  $\ln d_j(i) = \ln d_j(0) + \lambda(i \cdot \Delta t)$ , in which,  $\lambda$  can be easily calculated using a least-squares method (Rosenstein et al. 1993).

#### 3.2.2 Correlation Dimension Method

The main feature of chaos is the existence of strange attractors in phase space, which can be described by a correlation dimension with correlation integral. At present, the most widely used method to calculate the correlation dimension of a time series is the Grassberger-Procaccia algorithm (Grassberger and Procaccia 1983), which was therefore chosen for this work.

Suppose  $r$  is the radius of the sphere centered on  $Y_i$  or  $Y_j$ , then the correlation integral  $C(r)$  is given by:

$$C(r) = \lim_{n \rightarrow \infty} \frac{2}{n(n-1)} \sum_{i,j=1}^n \theta[r - \|Y_i - Y_j\|] \tag{2}$$

where  $\theta(\cdot)$  is the Heaviside function:

$$\theta(u) = \begin{cases} 0, & u \leq 0 \\ 1, & u > 0 \end{cases} \quad (3)$$

When  $r \rightarrow 0$ , the relationship between  $C(r)$  and  $r$  is as follows:  $\lim_{r \rightarrow 0} C(r) \propto r^D$ , where  $D$  is the correlation dimension and can describe the self-similar structure of a singular attractor. It can be calculated by:  $D = \log C_n(r) / \log r$ .

In the actual calculation,  $r$  is usually increased from a small value to a large one. For each  $r$ , the least-squares method is used to fit the plot of  $\log C(r)$  versus  $\log r$  and get the best line. The slope of the line is the correlation exponent. If the correlation exponent is saturated to a constant as the embedding dimension increases, then it is generally considered that the series is chaotic and the constant is the correlation dimension. If there is no saturation phenomenon, the system is entirely stochastic (Dhanya and Kumar 2010). Therefore, it can be distinguished chaotic sequences from stochastic sequences by whether the correlation dimension saturates or not.

### 3.2.3 Chaos Identification Method Based on Prediction Accuracy

The most common methods used for distinguishing dynamical chaos from stochastic noise in hydrological processes are the Lyapunov exponent method and the correlation dimension method. However, the value of the Lyapunov exponent is impacted by the choice of fitting region. The value of the correlation dimension is also affected by the embedding dimension. To avoid these problems, an approach is presented for identifying chaos based on the accuracy of nonlinear forecasts.

For a time series, the prediction accuracy can be measured by the correlation coefficient between the actual sequence and predicted sequence. The higher the correlation coefficient, the higher the prediction accuracy. The correlation coefficient is calculated as follows:

$$R = \frac{\sum_{t=1}^n \left[ \left( x_t - \bar{x} \right) \left( \hat{x}_t - \bar{\hat{x}} \right) \right]}{\sqrt{\sum_{t=1}^n \left( x_t - \bar{x} \right)^2} \sqrt{\sum_{t=1}^n \left( \hat{x}_t - \bar{\hat{x}} \right)^2}} \quad (4)$$

where  $x_t$  and  $\hat{x}_t$  are the observed value and predicted value, respectively,  $\bar{x}$  and  $\bar{\hat{x}}$  are the average values of  $x_t$  and  $\hat{x}_t$ , respectively, and  $n$  is the length of the time series. The parameter  $R$  indicates the strength of the linear relationship between the observed and simulated streamflow series.

Dynamic chaos and stochastic noise can be distinguished by comparing the predicted and actual trajectory (Sugihara and May 1990).

- (1) A fixed delay time is used to make a single-step prediction for a different embedding dimension. For a chaotic time series, the forecast accuracy will be at a maximum initially, after which the accuracy decreases with increasing embedding dimension. For stochastic time series, by contrast, the forecast accuracy does not change with the embedding dimension.
- (2) A fixed embedding dimension is used to make a multi-step prediction for a different delay time. For a chaotic time series, the forecast accuracy decreases with increasing

prediction-time interval, whereas for stochastic noise, the forecast accuracy has nothing to do with the prediction interval.

### 3.3 Chaotic Time Series Prediction

An  $m$ -dimensional vector  $X$  can be embedded into  $m$ -dimensional phase space using an  $m$ -dimensional map  $f_T$ , which can be expressed as:

$$Y = f_T(X) \tag{5}$$

where  $Y$  is also  $m$ -dimensional vector.

The input variables  $X$  and output variables  $Y$  can be described as:

$$X = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \vdots & \vdots & \ddots & \vdots \\ x_N & x_{N+\tau} & \cdots & x_{N+(m-1)\tau} \end{bmatrix}, Y = \begin{bmatrix} x_{2+(m-1)\tau} \\ x_{3+(m-1)\tau} \\ \vdots \\ x_n \end{bmatrix} \tag{6}$$

Where  $N = n - 1 - (m - 1)\tau$  is the number of sample points.

The phase space reconstruction method is usually used to find a proper formula  $f_T$  in Eq. (5). The local-region forecasting method based on the embedding theory of Takens is a simple and effective method for finding a map  $f_T$  (Sivakumar 2000). Neural networks are widely used for seeking the map  $f_T$  in many fields. The Extreme Learning Machines (ELM) is a kind of feedforward neural network, in which the input weights and hidden biases are randomly generated and don't need to be adjusted. Compared with traditional neural networks, it has excellent generalization performance and fast learning ability. In this study, the ELM method based on phase space reconstruction was used for predicting the monthly streamflow.

#### 3.3.1 Extreme Learning Machine

Set  $\{(x_t, y_t) | x \in R^n, y \in R^m, t = 1, 2, \dots, N\}$  as  $N$  training sets, where  $x_t = [x_{t1}, x_{t2}, \dots, x_{tm}]^T$  is the input sample and  $y_t = [y_{t1}, y_{t2}, \dots, y_{tm}]^T$  is the output sample. The ELM model with  $L$  hidden nodes can be expressed as:

$$f_L = \sum_{i=1}^L \beta_i g(w_i x_t + b_i) = o_t, \quad t = 1, 2, \dots, N \tag{7}$$

where  $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]$  and  $b_i$  are input weights and hidden biases respectively;  $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$  is the output weight between the hidden layer and the output layer;  $g$  is an activation function; and  $o_t = [o_{t1}, o_{t2}, \dots, o_{tm}]^T$  is the output value.

The training objective of the extreme learning machine network is to seek the optimum,  $w_i$ ,  $\beta_i$  and  $b_i$  such that  $\sum_{t=1}^N \|o_t - y_t\| = 0$ . Then

$$f_L = \sum_{i=1}^L \beta_i g(w_i x_t + b_i) = y_t, \quad t = 1, 2, \dots, N \tag{8}$$

The above formula can be simplified to  $H\beta = Y$ , where.

$$\begin{aligned}
 H &= \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} g(w_1x_1 + b_1) & \cdots & g(w_Lx_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1x_N + b_1) & \cdots & g(w_Lx_N + b_L) \end{bmatrix}_{N \times L}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times m}, Y \\
 &= \begin{bmatrix} Y_1^T \\ \vdots \\ Y_N^T \end{bmatrix}_{N \times m} \quad (9)
 \end{aligned}$$

Given the hidden node parameters  $(\hat{w}_i, \hat{b}_i)$  randomly, the output matrix of the hidden layer can be computed. Then, the smallest-norm least-squares solution,  $\hat{\beta}$ , is thus  $\hat{\beta} = H^T Y$ , where  $H^T$  is the Moore–Penrose generalized inverse of  $H$ .

### 3.3.2 Parameter Calibration

It is necessary to adopt some effective methods to optimize parameters  $(\hat{w}_i, \hat{b}_i)$  on the ELM model. The Genetic Algorithm (GA) (Wang 1997) and particle swarm optimization (PSO) (Jiang et al. 2010) are the foremost methods to improve the prediction performance. Both GA and PSO are parallel intelligent optimization algorithms. But PSO approach has rapid convergence over traditional GA. However, similar to GA, the possibility of premature convergence reduces its usefulness for global searches (Wang et al. 2012). To address this drawback, Jiang et al. (2015) improved the traditional PSO by introducing the idea of population hybrid evolution, named IPSO, to avoid premature convergence. In this paper, IPSO is used to enhance the learning performance of the extreme learning machine model. And a hybrid model has been proposed that integrates chaos theory and an extreme learning machine with optimal parameters selected by improved particle swarm optimization (ELM-IPSO) for monthly streamflow analysis and prediction.

The monthly streamflow forecasting using ELM-IPSO based on phase-space reconstruction is described as follows and shown in Fig. 3.

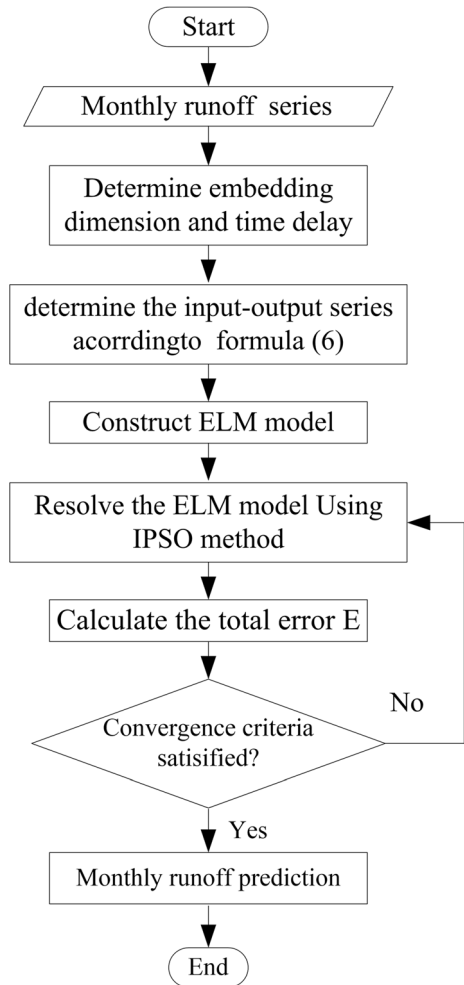
- Step 1: The input-output series  $(x_1, y_1), \dots, (x_N, y_N)$  for phase space reconstruction are determined using Eq. (6).
- Step 2: The extreme learning machine model is constructed with  $g$  chosen to be a sigmoid function,  $g(x) = 1 / (1 + e^{-x})$ .
- Step 3: The IPSO method is used to solve the ELM model.
- Step 4: The total error is calculated:  $E = \frac{1}{2} \sum_{j=1}^M \sum_{i=1}^N (y_{j,i} - \hat{y}_{j,i})^2$ .
- Step 5: If  $E$  is less than  $\varepsilon$  or the maximum number of generations is satisfied, the network training is complete.
- Step 6: The trained ELM model is used to predict streamflow.

### 3.4 Assessment Criteria

The prediction accuracy is evaluated by mean absolute error (MAE), root mean square error (RMSE), water balance relative error (RE), and Nash-Sutcliffe efficiency coefficient (NSE).



Fig. 3 Flow chart of the model construction procedure



They are defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Q_{obs,i} - Q_{sim,i}|, \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}, \tag{11}$$

$$RE = \frac{|\bar{Q}_{obs} - \bar{Q}_{sim}|}{\bar{Q}_{obs}} \times 100\%, \tag{12}$$

$$NSE = 1 - \frac{\sum_{i=1}^n (Q_{obs,i} - Q_{sim,i})^2}{\sum_{i=1}^n (Q_{obs,i} - \bar{Q}_{obs})^2}, \quad (13)$$

where  $Q_{obs,i}$  and  $Q_{sim,i}$  are the observed and predicted streamflow, respectively,  $\bar{Q}_{obs}$  and  $\bar{Q}_{sim}$  are the average values, and  $n$  is the length of the streamflow series.

The MAE is a measure of how close the predictions are to the observations. The RMSE is a way to quantify the difference between the predicted and observed values. The RE is the systematic relative error. The closer the values of MAE, MRSE and RE are to zero, the better the simulation effect. The NSE also measures the coincidence between the observed and simulated sequences. The value of the NSE is always expected to be close to unity for a right prediction.

## 4 Results and Discussion

The monthly streamflow data from 1956 to 2000 were analyzed for the existence of chaos and to determine the initial embedding dimension and delay time to reconstruct the phase space. The data from 2001 to 2010 were used for prediction.

### 4.1 Identification of Chaotic Characteristics

#### 4.1.1 Determination of Delay Time

The delay time  $\tau$  was calculated using the mutual information method for phase space reconstruction. Figure 4 shows the mutual information for various lag times. Because the first minimum value reached by the mutual information function is at lag time 6, it was selected as the delay time. However, the mutual information function is not a necessary or sufficient tool to describe whether a process is stochastic or chaotic. Therefore, it is necessary to use other methods to determine further whether the streamflow sequence is chaotic or not.

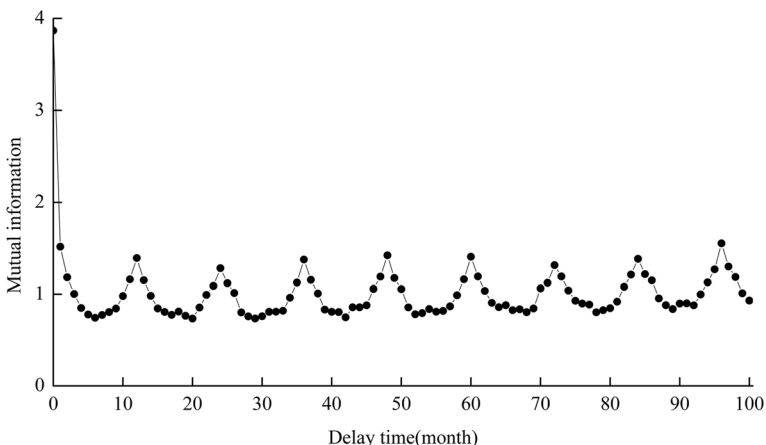


Fig. 4 Mutual information with delay time for monthly streamflow data

### 4.1.2 Correlation Dimension Method

The correlation integral  $C(r)$  was calculated using the Grassberger–Procaccia algorithm. Figure 5a shows a plot of  $C(r)$  versus  $r$  on a logarithmic scale for embedding dimensions,  $m$ , from 1 to 20. The slope of the plot determines the correlation exponent. Figure 5b shows the relationship between the correlation exponent values and the embedding dimension values. It can be noticed that the correlation exponent increases with embedding dimension and the slope of the plot tends to be saturated for embedding dimension  $m \geq 9$ . The saturation value of the correlation exponent is about 4.135. Therefore, it can be indicated that the streamflow series exhibits low-dimension chaotic behavior, and that streamflow prediction may be feasible using chaotic prediction methods.

### 4.1.3 Lyapunov Exponent

The largest Lyapunov exponent was calculated using a method for small data sets. Figure 6 shows a plot of  $y(i)$  versus  $i$ , where  $i$  is the discrete time step and  $y(i)$  is the average logarithmic distance of all neighbors after  $i$  discrete time steps. In Fig. 6, there is an approximately straight line before  $i = 5$  and the slope of the dotted line is equal to the theoretical value of the largest Lyapunov exponent. The positive largest Lyapunov exponent confirms that trajectories diverge exponentially and hence that the monthly streamflow is chaotic.

## 4.2 Phase Space Reconstruction Parameter Optimization

The phase space was reconstructed with delay times from 1 to 10 and embedding dimensions from 1 to 10. For different delay times, the ELM-IPSO method was used to get the corresponding optimal embedding dimension and prediction accuracy. Each combination of delay time and optimal embedding dimension (Table 1) results in different prediction accuracy. The delay time and embedding dimension corresponding to the maximal prediction accuracy were selected as the adjusted reconstructed parameters. The best prediction accuracy,  $\rho = 0.871$ , is achieved for embedding dimension 5 and delay time 1 (Table 1). Therefore,  $\tau = 1$  and  $m = 5$  were chosen as the adjusted optimal space phase reconstruction parameters for real prediction.

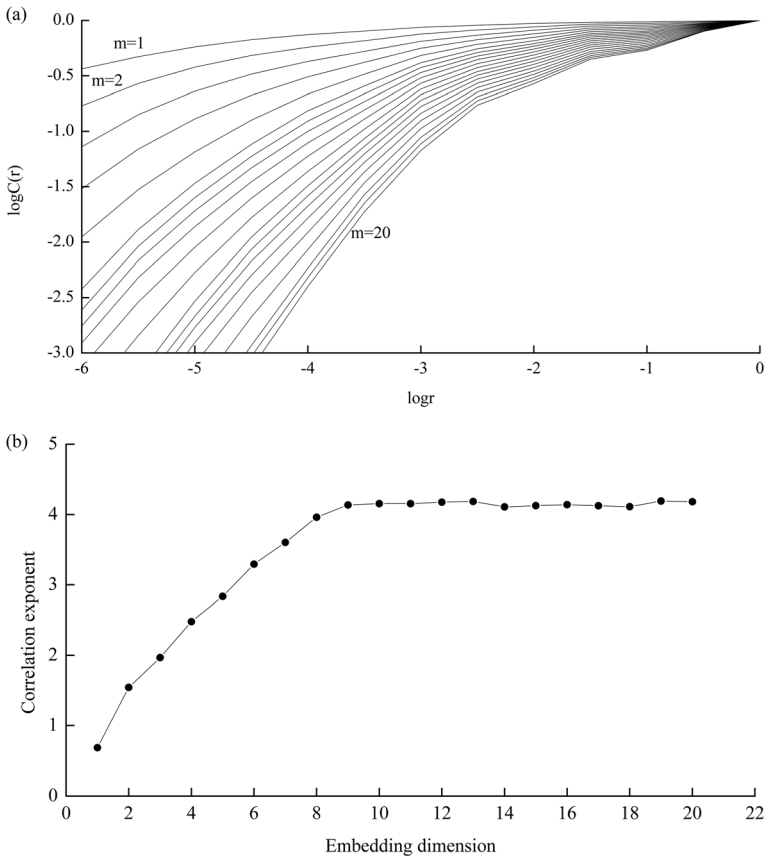
Table 1 Optimal embedding dimensions and maximum correlation coefficients for different delay times.

## 4.3 Prediction Accuracy

### 4.3.1 Parameter Settings

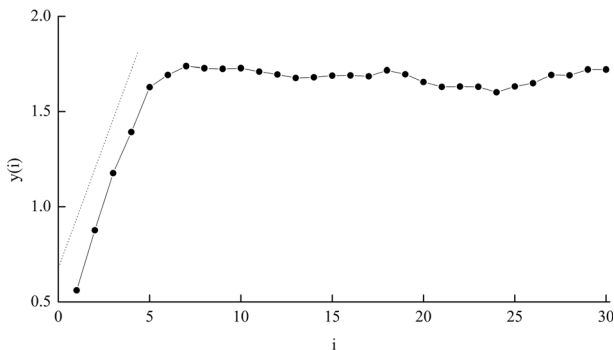
To test the performance of ELM-IPSO for monthly streamflow prediction, it was compared with an auto-regression method (AR), a three-layer feedforward artificial neural network (ANN), an extreme learning machine with genetic algorithm (ELM-GA) and with PSO algorithm (ELM-PSO). For all benchmark models, phase space was reconstructed firstly to design input vectors, and the training and validation data sets of all models are the same. The difference between the three methods is the choice of forecasting technique: ELM-IPSO, AR, ANN, ELM-GA or ELM-PSO. This allows us to evaluate which model is the most accurate.

In the ANN network, tansig function and logsig function are chosen as activation function of hidden layer and output layer, respectively. And Traingdx function is selected to train



**Fig. 5** **a**  $\log C(r)$  vs  $\log r$  for monthly streamflow data and **b** relationship between correlation exponent and embedding dimension

network. The number of hidden layer nodes is determined as 13 according to a trial and error method. And a maximum number of iteration, acceptable error and learning efficiency are set 5000, 0.01 and 0.1, respectively. In the ELM model, the sigmoid function is chosen as the transfer function and the number of hidden layer nodes is determined as 20 according to a trial



**Fig. 6**  $y(i)$  vs  $i$  using method for small data sets for monthly streamflow data

**Table 1** Different delay time and its corresponding optimal embedding dimension and the maximum correlation coefficient

$\tau$	1	2	3	4	5	6	7	8	9	10
$m$	5	8	6	4	2	9	9	8	8	3
$\rho$	0.871	0.747	0.757	0.811	0.749	0.756	0.770	0.792	0.756	0.779

and error method. The relevant experimental parameters for GA, PSO and IPSO algorithms are shown in Table 2. When the maximum iteration is reached, the algorithms are terminated. To avoid the influence of randomness, all algorithms need 10 trials to get the optimal solution.

### 4.3.2 Results and Analysis

Performance measures for the various prediction methods are shown in Table 3. These results indicate the following: 1) Although RE reached 23.01% for AR in the forecasting period, for the other assessment criteria is in the allowable range (no more than 20%) in the training and forecasting periods. 2) Both in the training and forecasting periods, for ELM-IPSO, the MAE, RMSE and RE are the lowest, and NSE and R are the highest. By these performance measures, the ELM-IPSO method is the most effective for streamflow prediction. 3) Compared with AR, ANN and ELM methods, it can be seen from the forecasting results that the AR method and ANN model cannot give a satisfactory performance. Therefore, the extreme learning machine method may be more suitable for streamflow forecasting. 4) Additionally, by comparing the results for ELM-GA and ELM-IPSO, it can be seen that relative to ELM-GA, for ELM-IPSO, MAE and RMSE decreased by 7.16% and 1.67% respectively in the training period, and by 3.57% and 2.01% in the forecasting period. NSE and R were improved by 4.17% and 2.44 respectively in the training period, and by 2.63% and 1.13% in the forecasting period. Similarly, by comparing ELM-PSO and ELM-IPSO, it can be seen that relative to ELM-GA, the ELM-IPSO decreases of 4.04% and 0.91% in MAE and RMSE respectively in the training period, however increased by 7.41% and 2.06 in the forecasting period; and was improved by 1.41% and 1.16% in NSE and R respectively in the training stage, and by both

**Table 2** The parameters of GA, PSO and IPSO

Parameter	Description	Value
$S$	The population size for all three algorithms	40
$c_1$	Crossover probability	0.8
$c_2$	Mutation probability	0.05
$d_1$	Constant factor for IPSO	2.8
$d_2$	Constant factor for IPSO	1.2
$\omega_{max}$	Initial inertia weigh factor for PSO and IPSO	0.9
$\omega_{min}$	Final inertia weigh factor for PSO and IPSO	0.4
$\beta$	Shrink factor for IPSO	0.001
$a_1$	Acceleration coefficient only for PSO	2.0
$a_2$	Acceleration coefficient only for PSO	2.0
$a_1$	Acceleration coefficient only for IPSO	2.05
$a_2$	Acceleration coefficient only for IPSO	2.05
$a_3$	Migration coefficient for IPSO	0.8
$M$	The number of sub-swarms for IPSO	4
$N$	The number of points in each sub-swarm for IPSO	10
$T$	The maximal iteration of each sub-swarm for IPSO	30
$MaxGen$	The total iterations for all four algorithms	3000

**Table 3** Forecasting performance of the different methods

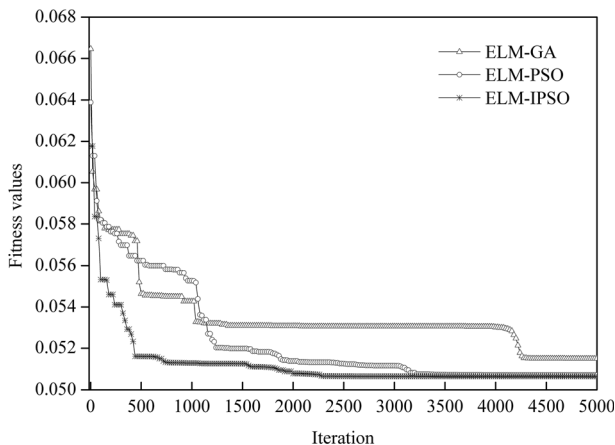
Criteria	Traning					Validation				
	MAE (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	RE (%)	NSE	R	MAE (m <sup>3</sup> /s)	RMSE (m <sup>3</sup> /s)	RE (%)	NSE	R
AR	6.75	7.98	-17.09	0.54	0.53	5.59	3.11	23.01	0.51	0.58
ANN	5.33	6.86	14.35	0.65	0.78	4.32	2.90	19.16	0.64	0.87
ELM-GA	4.61	7.79	18.69	0.72	0.85	1.12	1.49	15.97	0.76	0.88
ELM-PSO	4.46	7.73	15.37	0.74	0.86	1.08	1.43	15.84	0.78	0.89
ELM-IPSO	4.28	7.66	9.82	0.75	0.87	1.16	1.46	-11.04	0.78	0.89

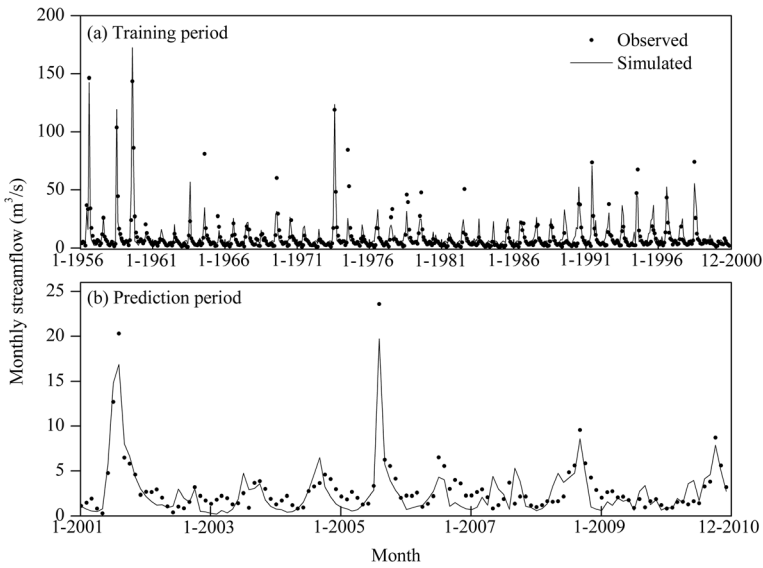
zero in the forecasting period. These results show that the monthly streamflow forecasting accuracy can be improved by using the IPSO algorithm to train ELM. Figure 7 shows the iterative processes of the fitness values of GA, PSO and IPSO for solving the ELM model. Compared with ELM-GA and ELM-PSO, ELM-IPSO has faster convergence speed and can find the optimal solution quickly.

Figure 8 shows the streamflow simulation results of the training and real prediction periods. There is excellent agreement between the observed streamflow and the forecasting streamflow. Figure 9 shows a scatter plot to evaluate model capabilities for simulating the dynamics of streamflow, in which a linear regression equation was used to analyze the correlation between the simulated and measured streamflow. It can be seen that the determination coefficients ( $R^2$ ) of 0.7587 and 0.7926 at the 0.01 significance level for the training and prediction periods respectively, which indicates the prediction of ELM-IPSO method has a good correlation with the observed data. These results further indicate the effectiveness of the ELM-IPSO method.

## 5 Conclusions

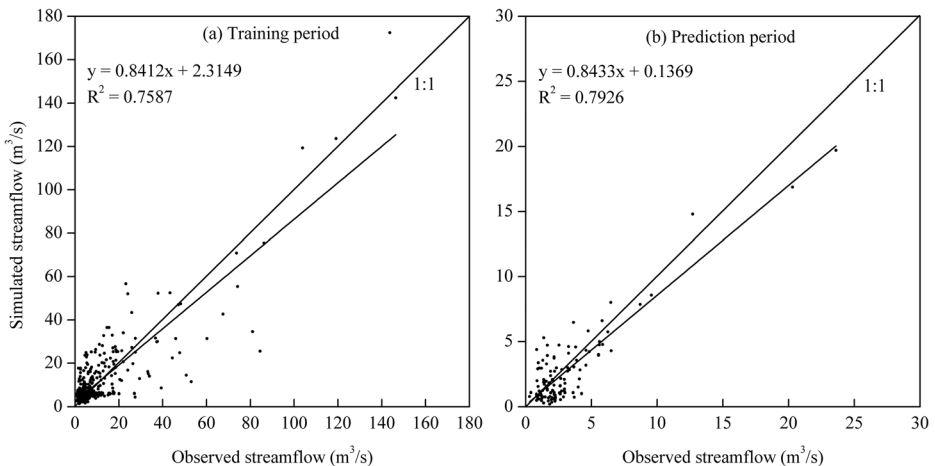
The purpose of this study is to analyze the chaotic properties of streamflow series using various techniques and propose a hybrid model integrating chaos theory and extreme learning machines to predict streamflow. Monthly streamflow data from Daiying hydrological station

**Fig. 7** The relationship between the fitness value and the number of iterations



**Fig. 8** Comparison of observed and predicted monthly streamflow during training and prediction periods

in the Chaohe River basin in northern China were used for the study. The behavior of streamflow dynamics is investigated by calculating their correlation dimensions using the Grassberger–Procaccia algorithm and the maximal Lyapunov exponents using methods for small data sets and nonlinear prediction. Then, based on phase space reconstruction, an extreme learning machine with parameters selected using an improved particle swarm optimization (ELM-IPSO) is developed to improve the streamflow prediction. Monthly streamflow data from 1956 to 2000 were used to determine the initial embedding dimension and delay time to reconstruct the phase space. The data from 2001 to 2010 were used for prediction. The accuracy of the streamflow prediction (linear correlation coefficient of about 0.89 and efficiency coefficient of about 0.78) indicate the validity of the proposed ELM-IPSO method for



**Fig. 9** Scatter plots of the training and prediction periods

predicting streamflow. Compared with AR, ANN, ELM-GA and ELM-PSO methods, ELM-IPSO has the lowest MAE, RMSE, and RE value, and the highest NSE and R value, during both the training and prediction stages. These results demonstrate that ELM-IPSO is an effective technique in improving the forecasting accuracy of monthly streamflow.

**Acknowledgments** This research was supported by the National Key Research and Development Plan (2016YFD0201206).

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare that they have no conflict of interest.

## References

- Benettin G, Froeschle C, Scheidecker JP (1979) Kolmogorov entropy of a dynamical system with an increasing number of degrees of freedom. *Phys Rev A* 19:2454–2460
- Bordignon S, Lisi F (2000) Nonlinear analysis and prediction of river flow time series. *Environmetrics* 11:463–477
- Bradford PW, Mark SS, Thor HM (1991) Searching for chaotic dynamics in snowmelt runoff. *Water Resour Res* 27(6):1005–1010
- Dhanya CT, Kumar DN (2010) Nonlinear ensemble prediction of chaotic daily rainfall. *Adv Water Resour* 33:327–347
- Duan QY, Sorooshian S, Gupta VK (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour Res* 28(4):1015–1031
- Farmer DJ, Sidorowich JJ (1987) Predicting chaotic time series. *Phys Rev Lett* 59:845–848
- Ghorbani MA, Khatibi R, Mehr AD (2018) Chaos-based multigene genetic programming: a new hybrid strategy for river flow forecasting. *J Hydrol* 562:455–467
- Grassberger P, Procaccia I (1983) Characterization of strange attractors. *Phys Rev Lett* 50(5):346–349
- Han M, Zhang RQ, Xu ML (2017) Multivariate chaotic time series prediction based on ELM-PLSR and hybrid variable selection algorithm. *Neural Process Lett* 46(2):705–717
- Hong M, Wang D, Wang Y, Zeng X, Ge S, Yan H, Singh VP (2016) Mid-and Longterm runoff predictions by an improved phase-space reconstruction model. *Environ Res* 148:560–573
- Hu Z, Zhang C, Luo G, Teng Z, Jia C (2013) Characterizing Crossscale chaotic behaviors of the runoff time series in an Inland River of Central Asia. *Quat Int* 311(9):132–139
- Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70:489–501
- Huang SZ, Chang JX, Huang Q, Chen YT (2014) Monthly streamflow prediction using modified EMD-based support vector machine. *J Hydrol* 511:764–775
- Huang G, Huang GB, Song SJ, You KY (2015) Trends in extreme learning machines: a review. *Neural Netw* 61:32–48
- Islam MN, Sivakumar B (2002) Characterization and prediction of runoff dynamics: a nonlinear dynamical view. *Adv Water Resour* 25:179–190
- Jiang Y, Liu CM, Huang CC, Wu XN (2010) Improved particle swarm algorithm for hydrological parameter optimization. *Appl Math Comput* 217:3207–3215
- Jiang Y, Li XY, Huang CC (2013) Automatic calibration a hydrological model using a master-slave swarms shuffling evolution algorithm based on self-adaptive particle swarm optimization. *Expert Syst Appl* 40(2):752–757
- Jiang Y, Liu CM, Li XY, Liu LF, Wang HR (2015) Rainfall-runoff modeling, parameter estimation and sensitivity analysis in a semiarid catchment. *Environ Model Softw* 67:72–88
- Kedra M (2013) Deterministic chaotic dynamics of Raba River flow (polish Carpathian Mountains). *J Hydrol* 509:474–503
- Kennel MB, Brown R, Abarbanel HD (1992) Determining embedding dimension for phase space reconstruction using a geometric method. *Phys Rev A* 45:3403–3411
- Khan S, Ganguly AR, Saigal S (2005) Detection and predictive modeling of chaos in finite hydrological time series. *Nonlinear Process Geophys* 12:41–53



- Labat D, Sivakumar B, Mangin A (2016) Evidence for deterministic chaos in long-term high-resolution karstic streamflow time series. *Stoch Env Res Risk A* 30:2189–2196
- Mohammad ZK (2016) Investigating vhaos and nonlinear forecasting in short term and mid-term river discharge. *Water Resour Manag* 30(5):1851–1865
- Ouyang Q, Lu W, Xin X, Zhang Y, Cheng W, Yu T (2016) Monthly rainfall forecasting using EEMD-SVR based on phase-space reconstruction. *Water Resour Manag* 30(7):2311–2325
- Paluš M, Pecan L, Pivka D (1995) Estimating predictability: redundancy and surrogate data method. *Neural Network World* 4:537–550
- Peng T, Zhou JZ, Zhang C, Fu WL (2017) Streamflow forecasting using empirical wavelet transform and artificial neural networks. *Water* 9(6):406. <https://doi.org/10.3390/w9060406>
- Porporato A, Ridolfi L (1997) Nonlinear analysis of river flow time sequences. *Water Resour Res* 33(6):1353–1367
- Rosenstein MT, Collins JJ, De Luca CJ (1993) A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D* 65:117–134
- Sano M, Sawada Y (1985) Measurement of the Lyapunov spectrum from a chaotic time series. *Phys Rev Lett* 55(10):1082–1085
- Schreiber T, Schmitz A (1996) Improved surrogate data for nonlinearity tests. *Phys Rev Lett* 77:635–638
- Sivakumar B (2000) Chaos theory in hydrology: important issues and interpretations. *J Hydrol* 227:1–20
- Sivakumar B (2004) Chaos theory in geophysics: past, present and future. *Chaos, Solitons Fractals* 19:441–462
- Sivakumar B, Berndtsson R, Olsson J, Jinno K (2001) Evidence of chaos in the rainfall-runoff process. *Hydrol Sci* 46(1):131–145
- Sugihara G, May R (1990) Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344(6268):734–741
- Takens F (1981) Detecting strange attractors in turbulence. *Lecture Notes in Mathematics* 898:366–381
- Taormina R, Chau KK (2015) Data-driven input variable selection for rainfall-runoff modeling using binary-coded particle swarm optimization and extreme learning machines. *J Hydrol* 529:1617–1632
- Vicente-Guillén J, Ayuga-Telléz E, Otero D, Chávez JL, Ayuga F, García AI (2012) Performance of a monthly Streamflow prediction model for Ungauged watersheds in Spain. *Water Resour Manag* 26:3767–3784
- Wang QJ (1997) Using genetic algorithms to optimize model parameters. *Environ Model Softw* 12:27–34
- Wang Y, Zhou JZ, Zhou C, Wang YQ, Qin H, Lu YL (2012) An improved selfadaptive PSO technique for short-term hydrothermal scheduling. *Expert Syst Appl* 39:2288–2295
- Wolf A, Swift J, Swinney HL, Vastano A (1985) Determining Lyapunov exponents from a time serie. *Physica D: Nonlinear Phenomena* 16(3):285–317
- Xu JH, Chen YN, Li WH, Ji MH, Dong S (2009) The complex nonlinear systems with fractal as well as chaotic dynamics of annual runoff processes in the three headwaters of the Tarim River. *J Geogr Sci* 19:25–35
- Zhao X, Chen X, Xu Y, Xi D, Zhang Y, Zheng X (2017) An EMD-based chaotic least squares support vector machine hybrid model for annual runoff forecasting. *Water* 9(3):153. <https://doi.org/10.3390/w9030153>
- Zhou JZ, Peng T, Zhang C, Sun N (2018) Data pre-analysis and ensemble of various artificial neural networks for monthly Streamflow forecasting. *Water* 10:628. <https://doi.org/10.3390/w10050628>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.