

# Assessment of several nonlinear methods in forecasting suspended sediment concentration in streams

Mohammad Zounemat-Kermani

## ABSTRACT

In this paper, the use of nonlinear nearest trajectory based on phase space reconstruction along with several data-driven methods, including two types of perceptron artificial neural networks with Levenberg–Marquardt (ANN-LM) and particle swarm optimization learning algorithms (ANN-PSO), adaptive neuro-fuzzy inference system (ANFIS) and gene expression programming for forecasting suspended sediment concentration (SSC) dynamics in streamflow is studied. The nonlinearity of the series was tested using the method of surrogate data at 0.01 significance level as well as correlation exponent method. The proper time delay is calculated using the average mutual information function. Obtained results of different models are compared using root mean square error (*RMSE*), Pearson's correlation coefficient (*PCC*) and Nash–Sutcliffe efficiency with logarithmic values ( $E_{ln}$ ). Of the applied nonlinear methods, ANFIS generates a slightly better fit under whole daily SSC values (the least amount of  $RMSE = 10.5$  mg/l), whereas ANN-PSO shows superiority based on the  $E_{ln}$  criterion (the highest amount of  $E_{ln} = 0.885$ ). According to the non-parametric Mann–Whitney test, all data-driven models represent the same forecasted results and are significantly superior to the nearest trajectory-based model at the 99% confidence level.

**Key words** | correlation exponent, modeling, particle swarm optimization, perceptron neural network, river flow data, univariate time series

Mohammad Zounemat-Kermani  
Department of Water Engineering,  
Shahid Bahonar University of Kerman,  
Kerman,  
Iran  
E-mail: [zounemat@uk.ac.ir](mailto:zounemat@uk.ac.ir)

## INTRODUCTION

Proper estimation of sediment concentration including bed load and suspended sediment is a critical task in hydrology and hydraulic engineering projects. Determining sediment concentration has implications for dam design, sediment erosion to waterways, and watershed management (Cigizoglu & Alp 2006). Suspended sediment concentration (SSC) is known as the ratio of the mass of sediment in a water-sediment mixture to the volume of the mixture. The assessment and prediction of the volume of transport SSC are vital in river and hydraulic engineering (Kisi 2008).

There are a variety of nonlinear data-driven simulating and predicting/forecasting techniques for hydrological time series data sets including methods based on the concept of trajectories of system dynamics such as chaotic prediction methods (Yang *et al.* 2011) and data-driven models including

artificial neural networks (ANNs), neuro-fuzzy models, and gene expression programming (GEP) (Zounemat-Kermani & Teshnehlab 2008; Shiri & Kisi 2011; Khatibi *et al.* 2012). Several studies can be found using artificial intelligence-based models, such as ANNs and adaptive neuro-fuzzy inference systems (ANFISs) for SSC modeling with promising results (Rajaei *et al.* 2009, 2011; Melesse *et al.* 2011; Kakaei Lafdani *et al.* 2013; Zounemat-Kermani *et al.* 2016).

In recent years, the use of phase space reconstruction in hydrology has received considerable attention for analyzing and forecasting hydrological time series data. They generate predictions/forecasts by finding local portions of the time series that closely resemble a portion of the points immediately preceding the point to be predicted. Nonetheless, related studies for SSC prediction based on nearest trajectory

strategy and local prediction methods are scarce. [Sivakumar & Wallender \(2005\)](#) investigated the predictability of river flow and suspended sediment transport in the Mississippi River. Their approach was based on the reconstruction of a single-dimensional series within a multi-dimensional system state for representing the underlying dynamics as well as the use of a local approximation technique for the forecasting process. They claimed that prediction accuracy was encouraging for 1-day-ahead results. [Shang \*et al.\* \(2009\)](#) employed the system state embedding-based weight predictor algorithm to make a short-term forecasting of sediment transport phenomena. They reported that the predicted values were in reasonable agreement with observed values within 15 days, but they appeared much less accurate beyond the limit of 20 days. [Yang \*et al.\* \(2011\)](#) proposed a new chaos Bayesian optimal prediction method by choosing optimal parameters in the local linear prediction method and improving the prediction accuracy with Bayesian theory. The prediction results indicated that the applied method can choose the optimal parameters adaptively in the prediction process of monthly discharge. [Adenan & Noorani \(2014\)](#) applied the local approximation method (LPM) for prediction of river flow at three measurement stations in Peninsular Malaysia. Results showed that the predicted results were close to agreement for all stations with a high correlation coefficient. [Albostan & Önöz \(2015\)](#) investigated the chaotic behavior of the river basin. For that purpose, the daily discharge data of four gauge stations were examined by using nonlinear data analysis methods. The LPM was also applied to evaluate the performance of the model. They reported that the predicted values were in good agreement with the observations by having high values of the correlation coefficient.

The present study attempts to detect the dynamical behaviors of the SSC processes (in the San Joaquin River, California) at daily temporal scale time series using a phase space reconstruction theory based on the time delay embedding method. In addition, the ability of nearest trajectory strategy – based on phase space reconstruction – to provide daily forecasting of SSC as well as several data-driven models (ANN-LM, ANN-PSO, ANFIS and GEP) is investigated. The outline of the paper is as follows. The methodology including nonlinear methods is described immediately below. The study area (San Joaquin River) and statistical analyses over the data are then presented. This is followed by the application

of simulating models and their results. Next, the findings are discussed and remarks and directions for future studies given, and finally, conclusions are drawn.

## METHODOLOGY

In this section, the methods based on nearest trajectory strategy, including reconstruction of the system state and local prediction method (LPM) and method of surrogate as well as other data-driven methods are briefly described.

### Nearest trajectory strategy: system state reconstruction and local prediction method

[Takens \(1981\)](#) illustrated that the state of many dynamic systems can be accurately reconstructed by a finite window of time series. According to the time delay embedding theorem of [Takens \(1981\)](#), a time series can be formulated into time-lagged vectors to realize the underlying dynamics of a dynamical system. A dynamical system with  $d$ -dimensional space can be constructed from time series  $x_t = \{x_1, x_2, \dots, x_N\}$ . The components of each state vector  $Y_t$  are defined through the delay coordinates based on the proper time delay ( $\tau$ ) as below ([Sivakumar 2002](#); [Zounemat-Kermani & Kisi 2015](#)):

$$Y_t = \{x_t, x_{t-\tau}, x_{t-2\tau}, \dots, x_{t-(m-1)\tau}\} \quad (1)$$

where  $m$  is the embedding dimension and  $\tau$  is the embedding time delay. Time delay embedding techniques (i.e., Taken's theorem) are widely used as the input vector for both linear and nonlinear dynamic models. According to [Kugiumtzis \(1996\)](#), the most important consideration in choosing the embedding parameters,  $m$  and  $\tau$ , is the window length  $w = m\tau$ . The time delay can be estimated using some methods such as autocorrelation function (ACF) or average mutual information (AMI) function. The AMI, unlike the ACF, takes into account nonlinear correlations and was thus used in this study for estimating the time delay. By applying the mutual information approach,  $\tau$  is selected as the first minimum of the time delayed AMI function ([Karatassou & Santamouris 2010](#)).

All nonlinear local and global models have parameters that must be selected before the model can be constructed

or tested. In this study, a local zeroth-order model is constructed using the nearest points on the  $k$  nearest neighboring trajectory segments. This forecasting technique is based on the fact that nearby trajectories either converge or do not diverge fast enough for small sample steps in the system state. This model evaluates the mapping function among the points which are near the forecast points. The time series identifies the sub-phase space by selecting  $k$  nearest adjacent (neighbor) points within the  $d$  dimensional space (Sannasiraj *et al.* 2004).

The TISEAN (nonlinear time series analysis) package includes a robust and simple method. The neighbors can be chosen within a circle of constant radius from the forecast point. Specifically, in the system space, all neighbors of  $x_t$  are sought, upon which the forecasting is simply observed at time  $t + 1$  over the neighborhood  $u_t$  (Hegger *et al.* 1999; Sannasiraj *et al.* 2004).

$$y_{t+1} = \frac{1}{|u_t|} \sum_{y_j \in u_t} y_{j+1} \quad (2)$$

The radius of the neighborhoods  $r$  should be large enough in order to cover the noise extent, but still smaller than a typical curvature radius. For each point  $y_t$  in the time series, the nearest neighbor  $y_j$  for each point is sought in an  $m$ -dimensional space (Hegger *et al.* 1999). In this study, the initial distance (radius) was chosen as  $1/1,000$ .

## Method of surrogate data

Surrogate data are random data sets which have the same autocorrelation structure as the dynamical real time series.

In other words, surrogates have the same spectral properties and the same statistical distribution as the original data. The correspondence between surrogates and the original data, in terms of the relative discrepancy, either approaches to zero as the number of iteration increases or converges to a finite value acting as an uncertainty threshold (Millan *et al.* 2009). This method serves as a null hypothesis (e.g., the SSC time series was generated by a stochastic/linear process), which consists of a candidate stochastic/linear process (Theiler *et al.* 1992). The objective of the hypothesis is to reject the null hypothesis by comparing the original and random data sets for testing the nonlinearity. In case of the existence of significant difference, the assumption of stochasticity of the time series will be rejected. In general, two approaches can be applied to generate surrogate data, adding noise to the same spectral signal from the original time series or generating a random series by shuffling the sequence. The first technique is used in the present study for the generation of surrogate data (Ng *et al.* 2007).

## Perceptron neural networks (ANN-LM and ANN-PSO) and ANFIS models

An ANN is a large parallel distributed processing system inspired by the biological nervous system. The most commonly used ANN structure is a multi-layer perceptron architecture (Figure 1(a)). A typical three-layered ANN is composed of multiple elements called nodes and connection pathways linking them. As can be seen in Figure 1(a), a neuron ( $n$ ) receives input data ( $x_1, x_2, \dots, x_n$ ), processes it based on the synaptic weights and bias ( $w$  and  $b$ ), and transmits an output signal to other neurons in the next layer ( $o$ ). At the onset, initial weight values are randomly assigned, to

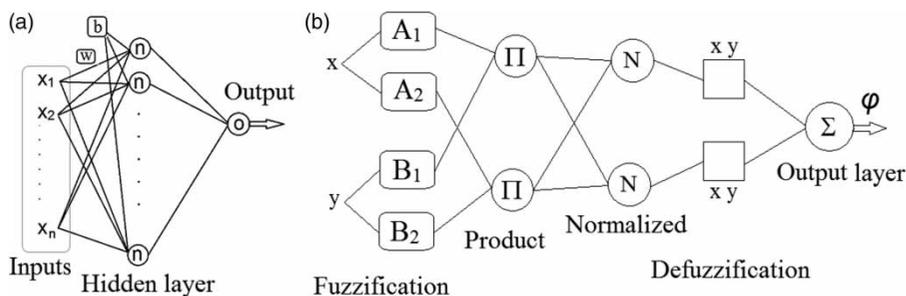


Figure 1 | Typical structure of (a) MLP artificial neural network and (b) adaptive neuro-fuzzy inference system.

be progressively corrected during a training process. Finally, the calculated outputs are compared with the measured outputs and errors are backpropagated (from right to left in Figure 1(a)) and the optimal weights are obtained by minimizing the errors through the training algorithm (Zounemat-Kermani 2012; Emiroglu & Kisi 2013; He *et al.* 2014). There are many various types of training algorithms which can be used for the ANN training process, such as back propagation (BP), genetic algorithm (GA), particle swarm optimization (PSO), and imperialist competitive algorithm (ICA). In this study, the conventional BP method based on the Levenberg–Marquardt (LM) algorithm and evolutionary computation PSO algorithm were implemented as training algorithms for updating the weights and biases. The LM method is a combination of two minimization methods of the gradient descent method and the Gauss–Newton method. PSO is a population-based optimization algorithm inspired by the social behavior of bird flocks and swarms of insects. The solving space is randomly initialized with a swarm of particles and afterwards the optima is searched within through updating generations. More detailed theoretical information about ANN-LM and ANN-PSO can be obtained from Zounemat-Kermani (2012) and Gaur *et al.* (2013). While multi-layer perceptron ANNs can have more than a single hidden layer, studies of ANN structure have shown that an ANN architecture with a single hidden layer is sufficient to approximate any complex nonlinear time series or function (Cybenko 1989). Consequently, a single hidden layer ANN was used in this study. The tangent sigmoid node activation function was implemented in the hidden layer, while a pure linear function was implemented in the output layer.

ANFIS is a universal approximator capable of approximating any real continuous function. The function of each node consists of fixed or adjustable parameters (Jang *et al.* 1997). Assuming a fuzzy inference system with two inputs ( $x$  and  $y$ ) and one output ( $\varphi$ ), then the rule base includes two fuzzy IF-THEN rules of the Takagi and Sugeno type (Zounemat-Kermani 2013):

$$\begin{aligned} &[\text{RULE } i: \text{IF } x \text{ is } A_i \text{ and } y \text{ is } B_i \text{ THEN } \varphi_i \\ &= p_i x + q_i y + r_i z + s_i]; i = 1, 2, \dots, n \end{aligned}$$

where  $A_i$  and  $B_i$  are membership functions,  $\varphi_i$  is the output

function of the fuzzy rule,  $n$  is the number of rules,  $p_i$ ,  $q_i$ ,  $r_i$ , and  $s_i$  are adjustable parameters. The ANFIS includes five layers, the fuzzification layer, product layer, normalized layer, de-fuzzification layer, and total output layer. The structure of an ANFIS is illustrated in Figure 1(b). More detailed information about ANFIS can be obtained from Jang (1993).

Gaussian-type functions were chosen as fuzzy set membership functions. More details of the applied ANN and ANFIS models are presented in Table 1.

### Gene expression programming

GEP is a genotype/phenotype GA. It can evolve programs of different sizes and shapes, encoded in linear fixed-length chromosomes. Similar to GAs, GEP uses populations of individuals, selects them according to fitness, and introduces genetic variation by using one or more genetic operators (Ferreira 2006). The GEP initializes a population consisting of random chromosomes; the fitness of each chromosome is evaluated according to a target value. Each chromosome is composed of multiple genes with each gene encoding a smaller sub-program. Every gene encodes a mathematical expression expressed as an expression tree. In GEP, the unconstrained operation of important genetic operators, such as mutation, transposition, and recombination are permitted as a result of structural and functional organization

**Table 1** | The training parameters of the ANN and ANFIS models

ANN		ANFIS	
Learning rate (LM)	0.01	Membership function	Gaussian
Momentum (LM)	0.85	And/Or method	Product/Maximum
Population size (PSO)	100	Aggregate/Implication method	Maximum/Product
Output layer's activation function	Pure linear	Defuzzification method	Wtaver <sup>a</sup>
Hidden layer's activation function	Tangent sigmoid		

<sup>a</sup>Note: Returns the weighted average of the  $x$ -values, with the  $y$ -values used as weights.

of the linear chromosomes (Hashmi *et al.* 2011; Shiri & Kisi 2011).

Having chosen appropriate input variables, the first step in the procedure for modeling SSC is the selection of a fitness function. In this research, the fitness function,  $f_i$ , was expressed as (Ferreira 2001):

$$f_i = \sum_{j=1}^n (R - |P_{(ij)} - T_j|) \quad (3)$$

where,  $R$  is the selection range,  $P_{(ij)}$  is the value predicted by the individual program  $i$  for fitness case  $j$ , and  $T_j$  is the target value for fitness case  $j$ . Chromosomes were created in a second step which consisted of choosing the set of terminals,  $T$ , and the set of functions,  $F$ . In this type of problem, the terminal set includes five lags of SSC. In the present study, 16 different mathematical functions (e.g., addition, subtraction, multiplication, etc.) were employed (Table 2). The third step was to define the chromosomal architecture. The fourth step was to select the linking function. In the final step, eight genetic operators were chosen. The parameters and operators used for each run are

summarized in Table 2. Full details regarding GEP modeling can be found in Ferreira (2001, 2006).

## STUDY AREA AND DATABASE

### Study area

The present study uses daily sediment data observed in the San Joaquin River near Vernalis station operated by the US Geological Survey in California from October 1st 1964 to September 30th 2012. The location of the stations is shown in Figure 2. The gage datum is 7.6 m above sea level and the drainage area at this site is 35,066 square kilometers.

### Data and statistical analysis

In this study the periods for training (October 1st 1964–September 30th 2007; about 90% of the total data) and testing sets (October 1st 2007–September 30th 2012; the remaining 10% of the total data) were selected based on the existing data of 48 water years. The statistical parameters

**Table 2** | Initial setting and parameters for the GEP model training

Function set	Symbol	Weight	Function set	Symbol	Weight
Addition	+	4	Logarithm( $x, y$ )	Log2	1
Subtraction	–	4	X to the power2	X2	1
Multiplication	×	4	X to the power3	X3	1
Division	/	2	Cubic root	$\sqrt[3]{}$	1
Square root	$\sqrt{}$	1	Gaussian	Gau	1
Exponential	Exp	1	Sine	sin	1
Natural log	Ln	1	Cosines	cos	1
Power	Pow	1	Arctangent	Atan	1
<b>General</b>			<b>Genetic operators</b>		
Chromosomes	60		Mutation rate	0.05	
Genes	5		Inversion rate	0.1	
Head size	10		IS transposition rate	0.1	
Linking function	Addition		RIS transposition rate	0.1	
Fitness function	RMSE		1-point recombination rate	0.2	
			2-point recombination rate	0.3	
			Gene recombination rate	0.1	
			Gene transposition rate	0.1	



**Figure 2** | Site location of Vernalis station in California (USGS Station No. 11303500; latitude  $37^{\circ}40'34''$ , longitude  $121^{\circ}15'55''$  NAD27).

of the entire data as well as the training data and testing data sets are presented in [Table 3](#).

It can be seen from [Table 3](#) that  $X_{\text{mean}}$ ,  $S_x$ ,  $C_v$ ,  $C_{s_x}$ , and  $X_{\text{max}}$  of SSC in the testing period are less than those

in the training period except for  $X_{\text{min}}$  of SSC, which is greater in the testing period. High positive skewness indicates their non-normal distribution and the existence of outliers. The variations of the daily SSC are shown in

**Table 3** | Statistical parameters for training, testing, and total SSC (mg/L) data in San Joaquin River near Vernalis station, CA<sup>a</sup>

Statistical parameters	Training set	Testing set	Total data
Number of data	16,021 (90%)	1,780 (10%)	17,801 (100%)
Xmean	79.96	51.00	76.94
Sx	49.83	30.13	48.96
Cv	0.62	0.59	0.64
Xmax	1,590	441	1,590
Xmin	6	12	6
Csx	5.07	4.74	5.00
r1	0.89	0.92	0.88
r2	0.75	0.80	0.76

<sup>a</sup>Note: Xmean is the average of the data; Xmax is the maximum value of the data; Xmin is the minimum value of the data; Sx is the standard deviation; Cv is the coefficient of variation; Csx is the skewness; and  $r_n$  is autocorrelation of SSC  $n$ -day lag.

Figure 3, in which the training (90%) and testing (10%) sets are depicted separately.

(Krause *et al.* 2005; Vafakhah 2013):

$$PCC = \frac{\sum_{i=1}^n (S_i^o - \bar{S}^o)(S_i^p - \bar{S}^p)}{\sqrt{\sum_{i=1}^n (S_i^o - \bar{S}^o)^2} \sqrt{\sum_{i=1}^n (S_i^p - \bar{S}^p)^2}} \quad (4)$$

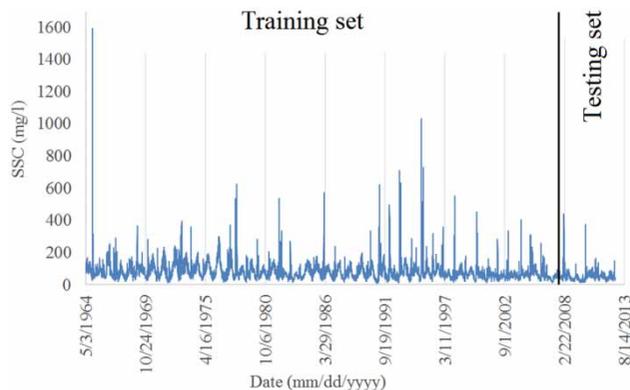
## APPLICATION AND RESULTS

### Evaluation of results

In this analysis, three different types of statistical criteria including the root mean squared error (*RMSE*), Pearson’s correlation coefficient (*PCC*), and  $E_{ln}$  (Nash–Sutcliffe efficiency with logarithmic values) were considered for evaluation of the performance of each model, as below

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (S_i^o - S_i^p)^2}{n}} \quad (5)$$

$$E_{ln} = 1 - \frac{\sum_{i=1}^n (\ln(S_i^o) - \ln(S_i^p))^2}{\sum_{i=1}^n (\ln(S_i^o) - \ln(\bar{S}^o))^2} \quad (6)$$



**Figure 3** | Daily time series variations of SSC in San Joaquin River near Vernalis station.

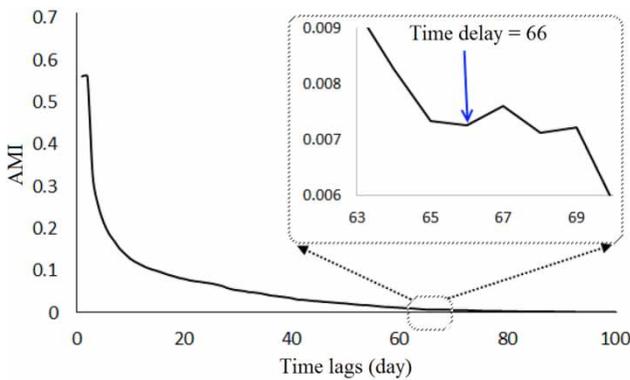
where  $S^o$  and  $S^p$  are the observed and the forecasted SSC, respectively,  $\bar{S}$  is the average SSC, and  $n$  is the number of data. The best value of *RMSE* is zero while absolute values for the *PCC* range from 0 to 1, with higher values (normally >0.9) indicating better model congruence. A value equal to zero indicates no correlation at all, whereas a value of 1 means that the dispersion of the prediction is exactly similar to that of the observation.

The efficiency criterion (*E*) proposed by Nash & Sutcliffe (1970) is defined as one minus the sum of the absolute squared differences between the predicted and observed values normalized by the variance of the observed values during the period under investigation. To reduce the problem of the squared differences and the resulting

sensitivity to extreme values, the Nash–Sutcliffe efficiency  $E$  is often calculated with logarithmic values of  $S^o$  and  $S^p$ . The  $E_{ln}$  gains the advantage of the logarithmic transformation of the SSC values. Thereafter, the peaks are flattened and the low SSCs are kept more or less at the same level. As a result, the influence of the low SSC values is increased in comparison to the SSC peaks (Krause *et al.* 2005).

In addition to the above-mentioned criteria, another criterion, namely, absolute criterion from residuals ( $Er$ ) based on the sum of residuals (the difference between observed and forecasted SSC) is used. Negative values of  $Er$  represent the under-estimation nature of the model, whereas positive values indicate over-estimation. A zero value indicates that the results are reluctant to being over/under-estimated which is considered as a good indication for the forecasting models:

$$Er = \frac{1}{n} \sum_{i=1}^n (S_i^p - S_i^o) \tag{7}$$



**Figure 4** | Selection of the proper delay time for the daily SSC time series using AMI function.

and one criterion for determining the over/under forecasting nature of each model.

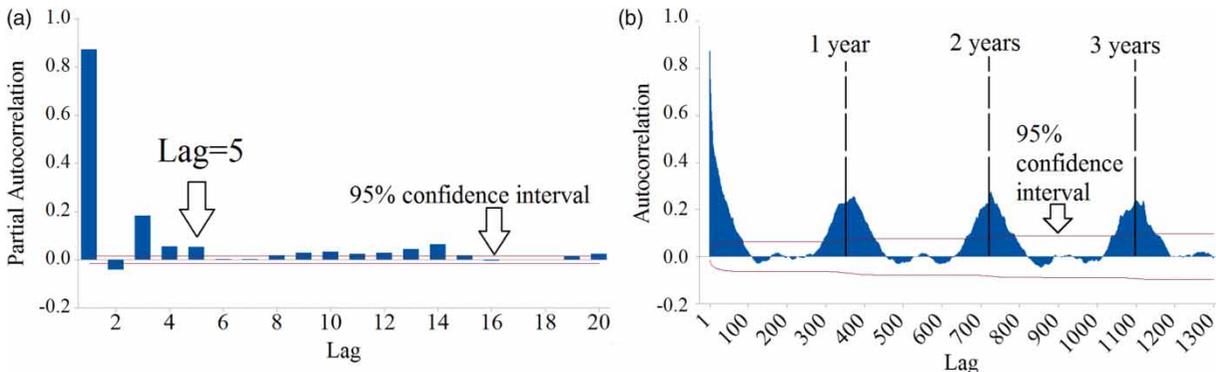
**Data preprocessing**

In this analysis, the delay time used for the system state reconstruction in the LPM model is chosen after examining the first minimum of the AMI function of the SSC data which is equal to is 66 days (Figure 4). A delay time of 66 days implies a threshold value which can be used as coordinates in a time delay vector.

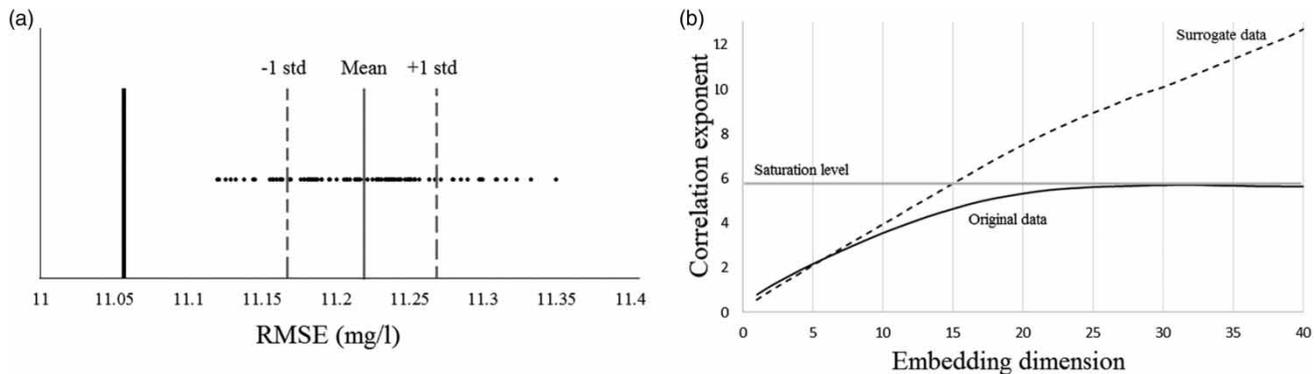
In addition, the partial autocorrelation function (PACF) is used for determination of proper input vector for the data-driven models. As can be seen from Figure 5(a), the input vector consists of five consequent lags for SSC data. In addition, for having a better insight into the nature of the SSC time series, the ACF has been used. Interpreting the results of ACF reveals that the time series does not have seasonality behavior, however, it shows that the SSC has a yearly nature (Figure 5(b)).

**Test of nonlinearity**

In this study, a statistical surrogate test for testing the null hypothesis (see the section ‘Method of surrogate data’) has been performed at the 1% significance level. For this purpose, a collection of 99 surrogates is generated for a one-sided hypothesis test for the testing set. The nonlinear *RMSE* prediction error is used as the test statistic. Figure 6(a) compares the *RMSE* prediction error of 99 surrogates and of the original data as well as intervals between the mean and  $\pm 1$  standard



**Figure 5** | (a) Partial autocorrelation function and (b) autocorrelation function of the SSC time series.



**Figure 6** | (a) Prediction error (RMSE) for original data and 99 surrogates and (b) variation of correlation exponent versus embedding dimension for original data and one of its surrogates.

deviation. The thick line indicates the *RMSE* of the original data. According to the obtained results (the prediction error of observed data is much less than that of surrogates), the null hypothesis is rejected, which means that the original time series SSC data do not come from a linear stochastic process at 1% significance level.

Another alternative way to detect possible nonlinearity in a time series is by applying amplitude-adjusted Fourier-transformed (AAFT) surrogate data set method (Theiler *et al.* 1992). In this respect, the behavior of the correlation exponent can be investigated for determining the presence of chaos in a time series, such that, for deterministic processes the value of the correlation exponent saturates after a certain value of embedding dimension (Zounemat-Kermani 2016). As illustrated in Figure 6(b), the relationship between correlation dimension and embedding dimension of the original SSC data and for its surrogate data, shows that in the case of surrogate data – unlike the original SSC series – the correlation exponents do not reach a saturation level. This issue also implies nonlinearity of the original SSC time series.

### Results of the nearest trajectory strategy model

In order to forecast the 1-day ahead of SSC using the nearest trajectory strategy (LPM) the time series analysis (TISEAN) package was used. Initially, the delay time was estimated by the AMI function using time lags of 1–100 days (as shown in Figure 4) to reconstruct the original system state ( $\tau = 66$ ). In this paper, the method used for the determination of the embedding dimension is based

on a calculation of the minimum *RMSE* of the forecasted results by increasing the embedding dimension to achieve the optimal dimension (Shang *et al.* 2009). Therefore, embedding dimensions 1 to 8 are used for the system state reconstruction.

Table 4 presents a summary of the forecasted results of the SSC reconstructed in embedding dimensions of 1 to 8. It can be seen that the proper embedding dimension for the daily forecasting of SSC is 2 (LPM(2) model). In order to check the validity of the calculated delay time by the AMI function, results of ten LPM(2) models with delay times of 10 to 100 have also been evaluated (Table 5). According to the obtained results, the best delay time is calculated as  $\tau = 60$  which is near to the 66 day delay time obtained from the AMI function. Results of the LPM(2) model in terms of curves of residuals and scatter plot are depicted in Figure 7.

**Table 4** | Forecasting results for daily SSC using the nearest trajectory strategy during the testing period

Model (embedding dimension)	RMSE (mg/L)	PCC
LPM(1)	11.332	0.924
LPM(2)	11.055	0.932
LPM(3)	11.582	0.928
LPM(4)	13.539	0.917
LPM(5)	14.843	0.903
LPM(6)	15.397	0.895
LPM(7)	16.063	0.886
LPM(8)	16.391	0.870

**Table 5** | Evaluation of LPM(2) model for predicting daily SSC considering various amounts of delay times

Delay time (day)	RMSE (mg/L)	PCC
10	11.558	0.927
20	11.562	0.928
30	11.124	0.930
40	11.065	0.931
50	11.067	0.931
60	11.049	0.933
70	11.098	0.931
80	11.469	0.922
90	11.598	0.918
100	11.587	0.919

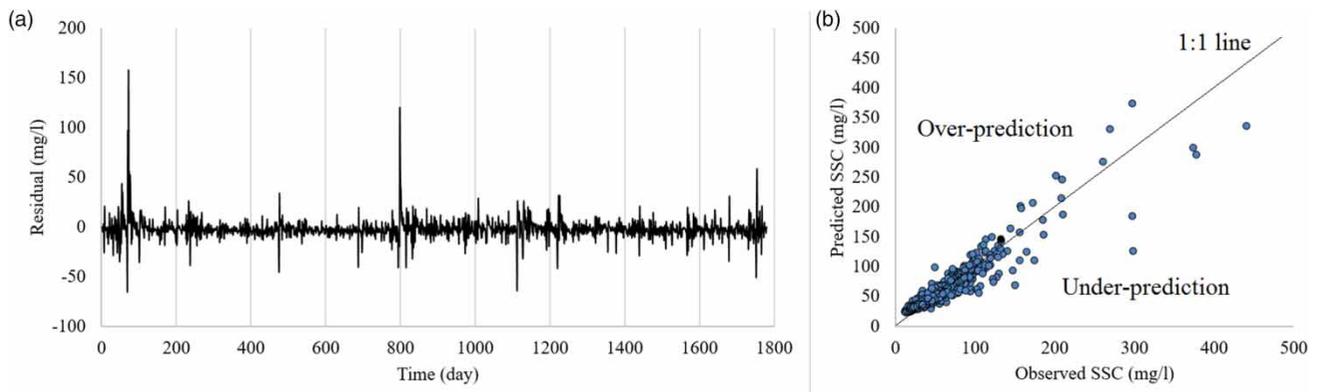
### Results of the data-driven methods

Five input combinations (see Table 6) and a single output (the 1-day ahead suspended sediment load ( $S_{t+1}$ )) were established for ANN-LM, ANN-PSO, ANFIS, and GEP models based on the PACF analysis (see Figure 5(a)).

In order to designate the appropriate number of neurons in the ANN's hidden layer, a trial and error procedure was used and the model with the lowest *MSE* for the testing set was considered as the representative ANN structure.

The *RMSE* and *PCC* statistics for the optimum ANNs, ANFIS, and GEP structures in the test period (Table 7) clearly show that the ANFIS(5-32-1) model is superior to the other models (lowest *RMSE* = 10.513 mg/L and highest *PCC* = 0.939) followed by GEP(5-55-9) model (*RMSE* = 10.731 mg/L and *PCC* = 0.936).

In this study, all the GEP models consisted of 16 functions (see Table 2 for further details) selected among all those available. This selection was based on the most comprehensive operators and their relevance to the nature of the problem. According to the PACF analysis, the terminal set contained one to five independent lagged historical variables similar to other forecasting models. Subsequently, each possible solution of the SSC prediction problem was randomly generated by combining some of the introduced functions and some terminals

**Figure 7** | Illustration of SSC predicted results in the testing period for LPM(2) model: (a) variations of residuals and (b) scatter plot of observed versus predicted values.**Table 6** | Model structures of ANNs, ANFIS, and GEP

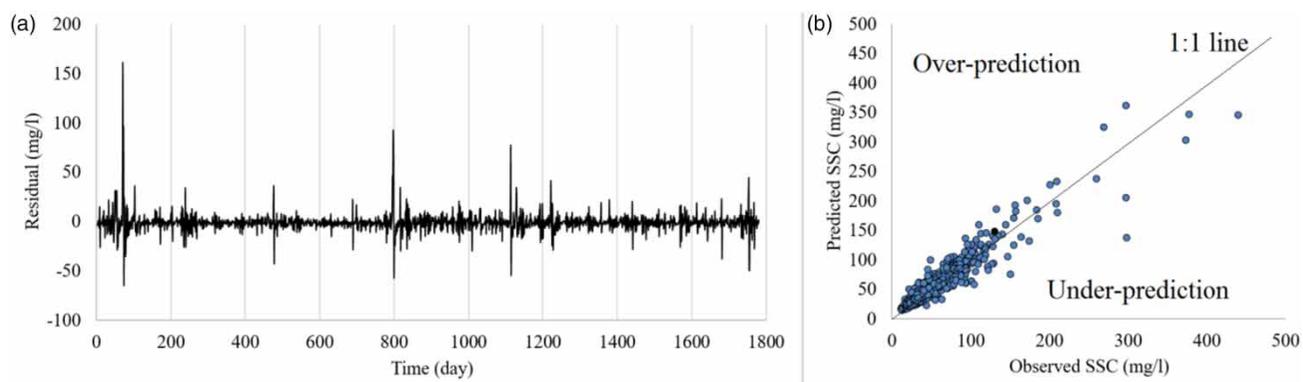
Input-Neurons-Output	Input-Rules number-Output	Gene No.-Chromosomes-Head size	Input(s)	Output
ANN(1-4-1)	ANFIS(1-2-1)	GEP(3-30-6)	$S_t$	$S_{t+1}$
ANN(2-5-1)	ANFIS(2-4-1)	GEP(4-40-7)	$S_t, S_{t-1}$	$S_{t+1}$
ANN(3-7-1)	ANFIS(3-8-1)	GEP(4-50-8)	$S_t, S_{t-1}, S_{t-2}$	$S_{t+1}$
ANN(4-8-1)	ANFIS(4-16-1)	GEP(5-55-9)	$S_t, S_{t-1}, S_{t-2}, S_{t-3}$	$S_{t+1}$
ANN(5-10-1)	ANFIS(5-32-1)	GEP(5-60-10)	$S_t, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}$	$S_{t+1}$

**Table 7** | Summary of the statistical parameters of the best applied nonlinear data-driven forecasting models for the test period of daily SSC

Model	Model inputs	RMSE (mg/L)	PCC
ANN-LM(4-8-1)	$S_t, S_{t-1}, S_{t-2}, S_{t-3}$	11.031	0.933
ANN-PSO(5-10-1)	$S_t, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}$	10.945	0.935
ANFIS(5-32-1)	$S_t, S_{t-1}, S_{t-2}, S_{t-3}, S_{t-4}$	10.513	0.939
GEP(5-55-9)	$S_t, S_{t-1}, S_{t-2}, S_{t-3}$	10.731	0.936

until the highest possible coefficient of determination was achieved. To arrive at the global optimal solution and to avoid getting trapped in a local minimum, all the genetic operators described in Table 2 were used, with the same appointed values employed for all the GEP models.

The ANFIS(5-32-1) model performance (as the best predictive model) is illustrated by means of time variation curves of SSC residuals and a scatter plot of observed and forecasted SSC (Figure 8).

**Figure 8** | Illustration of SSC predicted results in the testing period for ANFIS(5-32-1) model: (a) curves of residual variations and (b) scatter plot of observed versus predicted values.**Table 8** | Statistical comparison of forecasting models during the testing period

Model	RMSE (mg/L)	$E_r$ (mg/L)	$E_{ln}$	Enhancement in RMSE efficiency (%)	Enhancement in $E_{ln}$ efficiency (%)
ANFIS(5-32-1)	10.513	-1.04	0.883	4.9	1.1
GEP(5-55-9)	10.731	-1.02	0.882	2.9	1.0
ANN-PSO(5-10-1)	10.945	-1.05	0.885	0.3	1.4
ANN-LM(4-8-1)	11.031	-1.16	0.884	0.2	1.3
LPM(2)	11.055	-1.43	0.873	Base	Base

The negative values of  $E_r$  for each model indicates that all models under-forecasted SSC to a greater or lesser degree.

## DISCUSSION

A comparison of RMSE and PCC values (Tables 4 and 7) indicates that the ANFIS model outperformed all of the other SSC forecasting models. To achieve a better perspective of the overall results, the RMSE of the best forecasting models as well as their  $E_r$  values and Nash–Sutcliffe efficiency with logarithmic values ( $E_{ln}$ ) are tabulated in Table 8.

Based on the findings of the study, not much difference can be seen among the forecasted models, especially for the base SSC values (see Table 7, Enhancement in  $E_{ln}$  efficiency column). Thereafter, to have a better judgment about models' performance, two statistical tests were conducted (Table 9). First, the Anderson–Darling test was applied for checking each series normality. As all the series were found to be non-normal data ( $P$ -value < 0.005), the non-parametric Mann–Whitney test was conducted for checking the difference between the original and each model's forecasted results. According to the Mann–Whitney nonparametric test, that two samples come from the

**Table 9** | Results for normality test (Anderson–Darling) and nonparametric Mann–Whitney test

Model	Anderson–Darling normality test	Normality	Mann–Whitney test	Significantly different ( $\alpha = 0.01$ )	Significantly different ( $\alpha = 0.05$ )
Original	$P$ -value < 0.005	No	–	–	–
ANFIS	$P$ -value < 0.005	No	Significant at 0.166	No	No
GEP	$P$ -value < 0.005	No	Significant at 0.096	No	No
ANN-PSO	$P$ -value < 0.005	No	Significant at 0.037	No	Yes
ANN-LM	$P$ -value < 0.005	No	Significant at 0.031	No	Yes
LPM(2)	$P$ -value < 0.005	No	Significant at 0.002	Yes	Yes

same population, there is no significant difference between the original time series and data-driven models, whereas there is a significant difference for the local prediction model considering confidence level = 99%. In other words, it can be said that all the data-driven models performed satisfactorily with no statistical superiority, but they clearly outperformed the LPM model. However, considering confidence level equal to 95%, it can be seen from Table 8 that the ANFIS and GEP models are significantly superior to the other models.

The outcomes of this study imply that all the mentioned practical models are useful tools for forecasting 1-day ahead sediment concentration ( $PCC > 0.9$  and  $E_{ln} > 0.85$ ). However, according to the results of the Mann–Whitney test, it is recommended to apply nonlinear data-driven models rather than LPM. One of the main limitations of the implemented LPM method can be stated as its restriction to choosing the proper time delay and embedding dimensions (which are challenging issues) for constructing the phase space. Therefore, it is recommended to check the possibility of embedding the SSC dynamic system based on a differential phase space (rather than time delay phase space) and continuously examine the results of the forecasting model. In order to embed a chaotic time series data set into a differential phase space, unlike the time delay phase space theorem, the time scale parameter (rather than time lag) is used. Another difference between the two theorems is that during the differential phase space reconstruction, the noise in the chaotic time series has been reduced (Xu 2009).

## CONCLUSIONS

A preliminary attempt was made on the evaluation of nearest trajectory strategy using local prediction method (LPM) and

several data-driven models (ANN-PSO, ANN-LM, ANFIS and GEP) for the daily forecasting of SSC in Vernalis station located on the San Joaquin River, California. Forecasts based on the nearest trajectory strategy (LPM method) were made by reconstructing the SSC series in the system state using a delay time theorem. Results suggested that the dynamics of the underlying system could be embedded in a two-dimensional system. Although the LPM model provided reasonable results ( $RMSE = 11.06$  mg/L and  $PCC = 0.93$ ), it was inferior to data-driven models based on the Mann–Whitney statistical test (significantly different at the 99% confidence level). Future studies could involve applying other techniques such as conceptual methods in forecasting SSC in comparison with data-driven methods.

## REFERENCES

- Adenan, N. H. & Noorani, M. S. N. 2014 [Nonlinear prediction of river flow in different watershed acreage](#). *KSCE J. Civ. Eng.* **18** (7), 2268–2274.
- Albostan, A. & Önöz, B. 2015 [Implementation of chaotic analysis on river discharge time series](#). *Energ. Power Eng.* **7**, 81–92.
- Cigizoglu, H. K. & Alp, M. 2006 [Generalized regression neural network in modelling river sediment yield](#). *Adv. Eng. Softw.* **37**, 63–68.
- Cybenko, G. 1989 [Approximation by superposition of a sigmoidal function](#). *Math. Control Signals Syst.* **2**, 305–314.
- Emiroglu, M. E. & Kisi, O. 2013 [Prediction of discharge coefficient for trapezoidal labyrinth side weir using a neuro-fuzzy approach](#). *Water Resour. Manag.* **27** (5), 1473–1488.
- Ferreira, C. 2001 [Gene expression programming: a new adaptive algorithm for solving problems](#). *Complex Syst.* **13** (2), 87–129.
- Ferreira, C. 2006 [Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence](#). Springer, Berlin, Germany.
- Gaur, S., Ch, S., Grailot, D., Chahar, B. R. & Kumar, D. N. 2013 [Application of artificial neural networks and particle swarm optimization for the management of groundwater resources](#). *Water Resour. Manag.* **27** (3), 927–941.

- Hashmi, M. Z., Shamseldin, A. Y. & Melville, B. W. 2011 Statistical downscaling of watershed precipitation using gene expression programming (GEP). *Environ. Model. Softw.* **26**, 1639–1646.
- He, Z., Wen, X., Liu, H. & Du, J. 2014 A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J. Hydrol.* **509**, 379–386.
- Hegger, R., Kantz, H. & Schreiber, T. 1999 Practical implementation of nonlinear time series methods: the TISEAN package. *Chaos* **9**, 413–435.
- Jang, J. S. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics* **23** (3), 665–685.
- Jang, J. S. R., Sun, C.-T. & Mizutani, E. 1997 *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Upper Saddle River, NJ, USA.
- Kakaei Lafdani, E., Moghaddam Nia, A. & Ahmadi, A. 2013 Daily suspended sediment load prediction using artificial neural networks and support vector machines. *J. Hydrol.* **478**, 50–62.
- Karatasou, S. & Santamouris, M. 2010 Detection of low-dimensional chaos in buildings energy consumption time series. *Commun. Nonlinear Sci. Numer. Simul.* **15**, 1603–1612.
- Khatibi, R., Sivakumar, B., Ghorbani, M. A., Kisi, O., Koçak, K. & Zadeh, D. F. 2012 Investigating chaos in river stage and discharge time series. *J. Hydrol.* **414–415**, 108–117.
- Kisi, O. 2008 Constructing neural network sediment estimation models using a data-driven algorithm. *Math. Comput. Simul.* **79**, 94–103.
- Krause, P., Boyle, D. P. & Bäse, F. 2005 Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **5**, 89–97.
- 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.
- Melesse, A. M., Ahmad, S., McClain, M. E., Wang, X. & Lim, Y. H. 2011 Suspended sediment load prediction of river systems: an artificial neural network approach. *Agr. Water Manage.* **98** (5), 855–866.
- Millan, H., Kalauzi, A., Llerena, G., Sucoshanay, J. & Piedra, D. 2009 Meteorological complexity in the Amazonian area of Ecuador: an approach based on dynamical system theory. *Ecol. Complex.* **6**, 278–285.
- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **10** (3), 282–290.
- Ng, W. W., Panu, U. S. & Lennox, W. C. 2007 Chaos based analytical techniques for daily extreme hydrological observations. *J. Hydrol.* **342**, 17–41.
- Rajaei, T., Mirbagheri, S. A., Zounemat-Kermani, M. & Nourani, V. 2009 Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Sci. Total Environ.* **407** (17), 4916–4927.
- Rajaei, T., Nourani, V., Zounemat-Kermani, M. & Kisi, O. 2011 River suspended sediment load prediction: application of ANN and wavelet conjunction model. *J. Hydrol. Eng.* **16** (8), 613–627.
- Sannasiraj, S. A., Zhang, H., Babovic, V. & Chan, E. S. 2004 Enhancing tidal prediction accuracy in a deterministic model using chaos theory. *Adv. Water Resour.* **27**, 761–772.
- Shang, P., Na, X. & Kamae, S. 2009 Chaotic analysis of time series in the sediment transport phenomenon. *Chaos Soliton. Fract.* **41**, 368–379.
- Shiri, J. & Kisi, O. 2011 Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (South Western Iran). *J. Irrig. Drain. Eng.* **137** (7), 412–425.
- Sivakumar, B. 2002 A system state reconstruction approach to prediction of suspended sediment concentration in rivers. *J. Hydrol.* **258** (1–4), 149–162.
- Sivakumar, B. & Wallender, W. W. 2005 Predictability of river flow and suspended sediment transport in the Mississippi River basin: a non-linear deterministic approach. *Earth Surf. Proc. Land.* **30**, 665–677.
- Takens, F. 1981 Detecting strange attractors in turbulence. In: *Dynamical Systems and Turbulence, Lecture Notes in Mathematics 898* (D. A. Rand & L. S. Young, eds). Springer, Berlin, Germany, pp. 366–381.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B. & Farmer, J. D. 1992 Testing for nonlinearity in time series: the method of surrogate data. *Physica D* **58**, 77–86.
- Vafakhah, M. 2013 Comparison of cokriging and adaptive neuro-fuzzy inference system models for suspended sediment load forecasting. *Arab. J. Geosci.* **6** (8), 3003–3018.
- Xu, P. 2009 Differential phase space reconstructed for chaotic time series. *Appl. Math. Model.* **33** (2), 999–1013.
- Yang, X. H., Mei, Y., She, D. X. & Li, J. Q. 2011 Chaotic Bayesian optimal prediction method and its application in hydrological time series. *Comput. Math. Appl.* **61**, 1975–1978.
- Zounemat-Kermani, M. 2012 Hourly predictive Levenberg–Marquardt ANN and multi linear regression models for predicting of dew point temperature. *Meteorol. Atmos. Phys.* **117** (3–4), 181–192.
- Zounemat-Kermani, M. 2013 Hydrometeorological parameters in prediction of soil temperature by means of artificial neural network: case study in Wyoming. *J. Hydrol. Eng.* **18**, 707–718.
- Zounemat-Kermani, M. 2016 Investigating chaos and nonlinear forecasting in short-term and mid-term river discharge. *Water Resour. Manag.* **30** (5), 1851–1865.
- Zounemat-Kermani, M. & Kisi, O. 2015 Time series analysis on marine wind-wave characteristics using chaos theory. *Ocean Eng.* **100** (15), 46–53.
- Zounemat-Kermani, M. & Teshnehlab, M. 2008 Using adaptive neuro-fuzzy inference system for hydrological time series prediction. *Appl. Soft Comput.* **8** (2), 928–936.
- Zounemat-Kermani, M., Kisi, O., Adamowski, J. & Ramezani-Charmhineh, A. 2016 Evaluation of data driven models for river suspended sediment concentration modeling. *J. Hydrol.* **535**, 457–472.