ORIGINAL ARTICLE

Hybrid models for suspended sediment prediction: optimized random forest and multi-layer perceptron through genetic algorithm and stochastic gradient descent methods

Saeed Samadianfard[1] · Katayoun Kargar[2] · Sadra Shadkani[1] · Sajjad Hashemi[1] · Akram Abbaspour[1] · Mir Jafar Sadegh Safari[3]

Abstract

Owing to the nonlinear and non-stationary nature of the suspended sediment transport in rivers, suspended sediment concentration (SSC) modeling is a challenging task in environmental engineering. Investigation of SSC is of paramount importance in river morphology and hydraulic structures operation. To this end, for SSC modeling, first random forest (RF) and multi-layer perceptron (MLP) standalone models were developed, and then, they were optimized with genetic algorithm (GA) and stochastic gradient descent (SGD) to develop GA-MLP, GA-RF, SGD-MLP, and SGD-RF hybrid models. Variety of input scenarios are implemented for SSC prediction to find the best input combination. The streamflow and SSC data collected from two stations of Minnesota and San Joaquin rivers, respectively, located at South Dakota and California are utilized in the current study. Accuracies of the developed models are examined by means of three performance criteria of correlation coefficient (CC), scattered index (SI), and Willmott's index of agreement (WI). A significant promotion in accuracy of hybrid models has been seen in contrast to their standalone counterparts. As can be deduced from the results, GA-MLP-5 and GA-RF-5 models with CC of 0.950 and 0.944, SI of 0.290 and 0.308, and WI of 0.974 and 0.971, respectively, were found as best models for prediction of SSC at Minnesota river. The developed SGD-MLP-5 and SGD-RF-5 models with CC of 0.900 and 0.901, SI of 0.339 and 0.339, and WI of 0.945 and 0.946, respectively, gave accurate results at San Joaquin river. Through the application of SGD algorithm, the adaptive learning rate, epochs, rho, L1 and L2 were activated and presumed as 0.004, 10, 1, 0.000009 and 0, respectively. The ExpRectifier was considered as san activation operation due to its better efficiency in comparison with its alternatives for predicting SSC in SGD-MLP model. According to the results, the fifth scenario that incorporates SSC_{t-1}, SSC_{t-2}, Q_t, Q_{t-1}, and Q_{t-2} were found superior for SSC modeling in the studied rivers. The recommended hybrid algorithms based on GA and SGD optimization algorithms are proposed as practical tools for solving complex environmental problems.

Keywords Genetic algorithm · Stochastic gradient descent · Hybrid model · Prediction · Random forest · River · Suspended sediment

1 Introduction

Linked to the hydrological and environmental evolutionary modeling, there exists a significant progress in suspended sediment transport modeling in recent years. Understanding the sediment transport process and modeling such a complex phenomenon are of importance in water resources management [29, 30]. The suspended sediment concentration (SSC) in the river is a crucial problem in

environmental, hydraulic, and water resources engineering. Sediments change several features of river systems such as quality and health of water (transport pollutants), geography and navigability of river and channel [20, 44]. Sediments conveyed within the flow remain in suspension for a considerable length of river and time referred to the suspended load, and its prediction is a challenging task due to the effect of several hydrological and metrological parameters in a particular watershed [14, 47]. Conventionally, sediment rating curve method is widely applied for *SSC* computation. It shows the exponential relationship

---

Extended author information available on the last page of the article

🍁 Springer

between the river discharge and *SSC* through a regression analysis. As an exponential regression equation is over-fitted on entire data set, it may generate poor results on alternative data set. Therefore, a precise modeling approach is needed to solve such a complex problem [27]. Modeling of the sediment transport in rivers considering theoretical equations and mathematical models needs a wide range of data, but due to the lack of such an extensive data range, these models do not provide accurate estimations [47, 48].

Recently, numerous studies have been conducted for sediment transport modeling applying variety of machine learning methods. For example, Rajaee et al. [33] applied four methods of neuro-fuzzy (NF), artificial neural networks (ANNs) and multi-liner regression (MLR) to simulate the daily SSC in the USA. Based on the outcomes of this study, both ANN and NF models generated high performance in predicting SSC. Cobaner et al. [8] developed adaptive neuro-fuzzy model for *SSC* computation considering streamflow, rainfall, and suspended sediment data. The outperformance of adaptive neuro-fuzzy model is reported in contrast to the different types of ANN such as generalized regression neural network (GRNN), radial basis function neural network (RBNN) and multi-layer perceptron (MLP). Altunkaynak [2] and Zhang et al. [50] applied genetic algorithm (GA) and found out that its better accuracy in comparison with other approaches. Using neural differential evolution (NDE), NF and ANN and sediment rating curve (SRC) methods, Kisi [20] modeled daily *SSC* and found out that the NDE outperforms the other techniques in daily SSC estimation. Comparing the performance of linear genetic programming (LGP) to adaptive neuro-fuzzy inference system (ANFIS) and SRC methods for *SSC* estimation, Kisi and Guven [21] reported that the LGP model provides more accurate results than the other mentioned models. Singh and Chakrapani [43] examined the capability of feed-forward backpropagation (FFBP) ANN method for simulating the SSC considering rainfall, temperature, and discharge as model parameters. It was determined that increasing the number of input parameters improves the developed model's accuracy. The classification and regression tree (CART), RBNN, MLR, ANN, M5 model tree, and least square support vector regression (LSSVR) were used to estimate the suspended sediment at a basin of rive in India by Kumar et al. [23]. It was found that both ANN and LSSVR models generated accurate results.

Studies mentioned above applied standalone machine learning algorithms for SSC modeling. Due to the complexity of the problem and deficiencies of the standalone models in terms of adjusting the variety of model parameters, hybrid algorithms can be reliable approaches for modeling the SSC in rivers. As examples from the literature, because of the complication of the relationship of SSC and streamflow (Q), the wavelet-artificial neural network model (W-ANN) was utilized in predicting the SSC factor in the Kuye River by the Liu et al. [24]. They decomposed the daily time series of SSC and Q into sub-series at different scales as inputs for the model. It was pointed out that the W-ANN model had the best performance, which has higher forecasting precision than other models like SRC and ANN. ANFIS-FCM (fuzzy c-means clustering model) was suggested by Kisi and Zounemat-Kermani [22] to estimate the SSC in the USA. Based on the obtained results, the ANFIS-FCM gave better results than other models, including ANFIS, ANN, and SRC. Zounemat-Kermani et al. [52] utilized the data of hydrometric stations that placed in different part of the USA, such as Arkansas, Delaware and Idaho to assess four different models of support vector regression (SVR) and three different ANN methods in daily SSC prediction/approximation and compared to that of MLR and SRC methods. Statistical parameters indicated the superior performance of SVR and ANN models in comparison with the traditional methods. Malik et al. [26] reported the outperformance of co-active neuro-fuzzy inference system (C-ANFIS) technique in comparison with the MLR, multiple nonlinear regression (MNLR), MLP and SRC. Ghose and Samantaray [16] used ANN-FFBP and regression approaches together with their hybridized versions as GA-BPNN and GA-regression for the same purpose at various basins of the Suktel River to realize their sensitivity at regional scale. Liu et al. [25] modeled SSC time series for Kuye River watershed, China using hybrid ensemble empirical mode decomposition (EEMD-ANN), EEMD-MLR, ANN, and MLR methods. The performance of EEMD-ANN and EEMD-MLR models were improved by a factor of 52.9% and 41.0%, in contrast to the ANN and MLR methods.

Advanced hybrid machine leaning algorithms were implemented for river suspended sediment estimations as an evolutionary hydrological modeling. For instance, [40] appraised impression of a hybrid model merging support vector machine (SVM) with whale optimization algorithm (SVM-WOA) and compared to SVM-PSO and conventional SVM and RBFN models for estimating SSC at Sundargarh and Salebhata stations in Mahanadi River, India. The results showed that SVM-WOA accomplished superiorly in comparison with SVM-PSO, SVM and RBFN models for five various input scenarios. Roushangar et al. [36] modeled SSC and river discharge of two stations of Mississippi river and improved performance of the implemented models using wavelet transform (WT) and ensemble empirical mode decomposition (EEMD) approaches. Results indicated that data processing with WT was more suitable than the EEMD in enhancement of the models' performance. Data processing improved the

models' performance by a factor of up to 15%. It was found that using the merged kernel extreme learning machine (KELM) method, the previous stations data could be applied successfully for SSC and river discharge modeling when the stations' own data were not available. Dang [10] improved coupled models of discrete wavelet transform (DWT) with ANFIS, named DWT-ANFIS, and principal component analysis (PCA) with ANFIS, named PCA-ANFIS, for SSC simulation. The merged and single ANFIS models were trained and tested utilizing long-term daily SSC and river discharge which were measured on the Schuylkill and Iowa Rivers in the United States. The results revealed that the PCA-ANFIS accomplished better than the single ANFIS and the coupled DWT-ANFIS. Mehri et al. [28] utilized four intelligent methods of ANFIS-PSO, ANFIS-GA, ANFIS, and group method of data handling (GMDH) to estimate the sediment concentration distribution. Since both GA and PSO optimization approaches were utilized to improve the ANFIS model, the efficiency of these models was improved. The results showed that the performance of the ANFIS-PSO was better than ANFIS-GA, ANFIS, and GMDH models for estimation of suspended sediment distribution.

While the outperformance of RF and MLP has been already reported in the relevant literature, this study is designed to enhance the RF and MLP models' accuracy for SSC modeling through hybridization with GA and SGD to develop novel MLP-GA, RF-GA, MLP-SGD and RF-SGD models. The weakness in accurate estimation of suspended sediment using conventional regression formulas was already reported in the literature. To this end, in this study, an attempt was made to make an accurate estimation of the suspended sediment concentration using efficient methods such as MLP and RF and to improve the results using GA and SGD methods. Through the modeling, historical SSC and river discharge values of two stations were used for the modeling. Still, to the best of our knowledge, hybrid SGD-MLP and SGD-RF models have not been used for *SSC* estimation.

2 Materials and methods

2.1 Study area

Daily discharge (Q) and *SSC* data of the Minnesota and San Joaquin Rivers (Fig. 1) for the time period of 2000–2017 were acquired from the United States Geological Survey (USGS). The Minnesota River is one of the Mississippi Rivers tributaries having almost 534 km long, in the USA state of Minnesota. The station name is Minnesota with station number of 05,325,000, basin area of 35,065 km^2, latitude of 37° 40′34″ and longitude of 121° 15′55″. The studied river in Joaquin River is located at central California. The correspondent USGS station name is San Joaquin with station number of 11,303,500, basin area of 38,590 km^2, latitude of 44° 10′08'' and longitude of 94° 00′11″. Table 1 shows the statistical characteristics of SSC and Q parameters in both Minnesota and San Joaquin stations. It can be seen from Table 1 that SSC and Q parameters presented skewed distributions.

2.2 Multi-layer perceptron (MLP)

Multi-layer perceptron (MLP) neural network is considered as the most common neural networks for the nonlinear fitting with higher accuracy. In order to acquire that performance with high precision, this method selects an appropriate number of neurons and layers at its structure. The training process is used to find the suitable amount of weight for the connections between neurons. The back-propagation network is a common algorithm for the feed-forward neural network in which the output of each layer is considered as the input of the next one [13]. The MLP involves three layers of input, hidden, and output layers. In order to train the MLP neural networks, several algorithms are applied and the selection of each one can affect its accuracy and the learning rate of the network [6]. Figure 2 illustrates the flowchart of the MLP model. In this process, the number of the hidden layers and neurons of each hidden layer should be designated in a way to model give high performance. In structure of MLP, Levenberg–Marquardt (LM) algorithm is mostly utilized for calculation of output signals. Variety of hidden layers can be adopted for MLP structure design; however, one hidden layer provides satisfactory results in hydrological problems [38].

2.3 Random forest (RF)

The random forest (RF) is an ensemble learning method proposed by Breiman [5]. In order to acquire better generalization abilities, ensemble learning builds multiple base learners or combines several trees at its structure [31, 37]. Among various rule generation approaches, a RF method is an effective and practical approach. The RF uses the algorithms of the decision tree (classification and regression tree (CART)) as the base algorithm. It is notable that the RF is a more robust approach in comparison with other decision tree ensembles [9]. The RF method has ability in defining the appropriate predictor and re-scaling the data like other techniques is not required. The conventional regression tree has weak performance due to its tendency to over-fit on the training data set. The RF method uses randomness characteristics to overcome this problem [42]. In this method, each decision tree is made up of a bootstrap sample from the calibration dataset, which comprises about
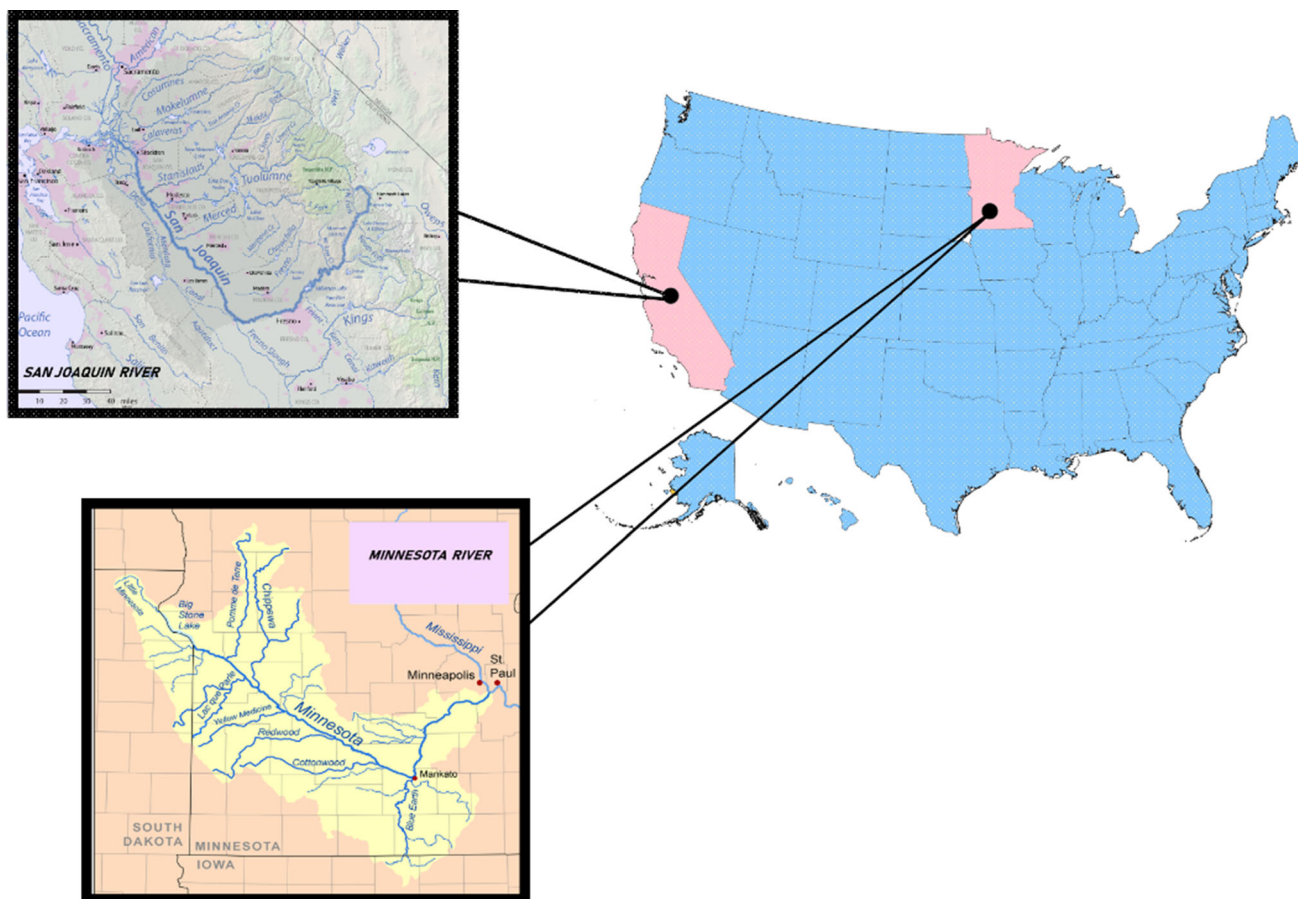
Fig. 1 Geographical location of the Minnesota and San Joaquin Rivers

Table 1 Statistical features of the applied data

Station	Variable	Mean	Minimum	Maximum	Standard deviation	Coefficient of variation	Skewness
Minnesota	SSC	171.150	5.000	2030.000	163.250	0.954	3.524
	Q	170.239	5.380	2361.625	239.801	1.408	3.597
San Joaquin	SSC	51.950	2.000	453.000	34.022	0.655	3.500
	Q	98.635	2.376	1138.337	147.681	1.497	3.496

two-thirds of the sample. The rest of the elements are considered as out-of-bag data. Variables are designated in random and based on the lowest Gini index and then, the best split is selected. According to the repetition at each bootstrap, the value for out-of-bag data is acquired. In this process, the repetition in each tree is continued until reaching the stop condition defined previously. In RF method, parameters are optimized by the usage of mean square error *(MSE)* and calibration dataset: confidence, number of trees, minimal leaf size, maximal depth, minimal size for split, subset ratio, and number of prepruning alternatives [12]. A RF method is effectively utilized for

the broad datasets analysis. Figure 3 illustrates the structure of the RF method.

2.4 Genetic algorithm (GA)

Holland [18] and Goldberg [17] developed the genetic algorithm (GA). GA is a powerful method for the exploratory development of large-scale hybrid optimization problems. GA encodes the problem as a set of strands that contain fine particles, and then, it changes the strands to stimulate the process of gradual evolution. Compared to the local search algorithms, in public search where there is only one acceptable solution, GAs consider a community
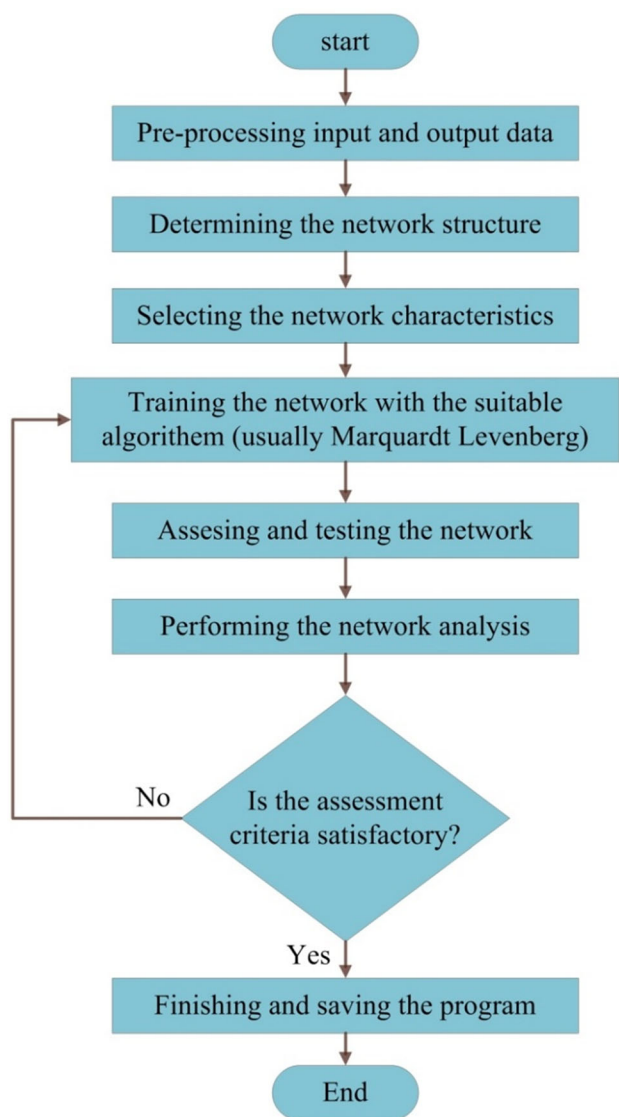
Fig. 2 Flowchart of the MLP model

of individuals. Working with a group of people makes it possible to study the main structures and characteristics of different people, which leads to the identification and discovery of more efficient solutions [15]. In this study, GA selects the most relevant disciplines and eliminates those disciplines that are less relevant to the study population. Each member of the population, which is an approximation of the final answer, is coded as strings of letters or mergers. These strands are called chromosomes. The most common mode is the display with the digits zero and one. Other modes of using three digits, real numbers and integers are also used. The values on the chromosomes alone do not have a specific meaning, but must be decoded and have meaning and result only as decision variables. It should be noted that the search process is performed on encrypted information unless it is from genes with real values. Once the chromosomes have been encoded, the efficiency or fit of each member of the population can be calculated. Fit is a relative scale that indicates the suitability of individuals to produce the next generation. In the nature, fit is equivalent to one's ability to survive. The objective function plays a decisive role in determining the fit of individuals. During reproduction, the fit of each individual is determined with the help of the primary information obtained from the objective function. These values are used in the selection process to lead it to select the right people. The higher the fit of the individual in relation to the population has a more chance to be elected. The lower the relative fit, the less likely it is to be selected for the next-generation production. The act of replication in a GA is used to exchange genetic information between a pair or more individuals. The simplest type of multiplication is the intersection of a point. After this step, a mutation operator is likely to be applied to the generated strands. In mutation, each individual alone can change according to the laws of
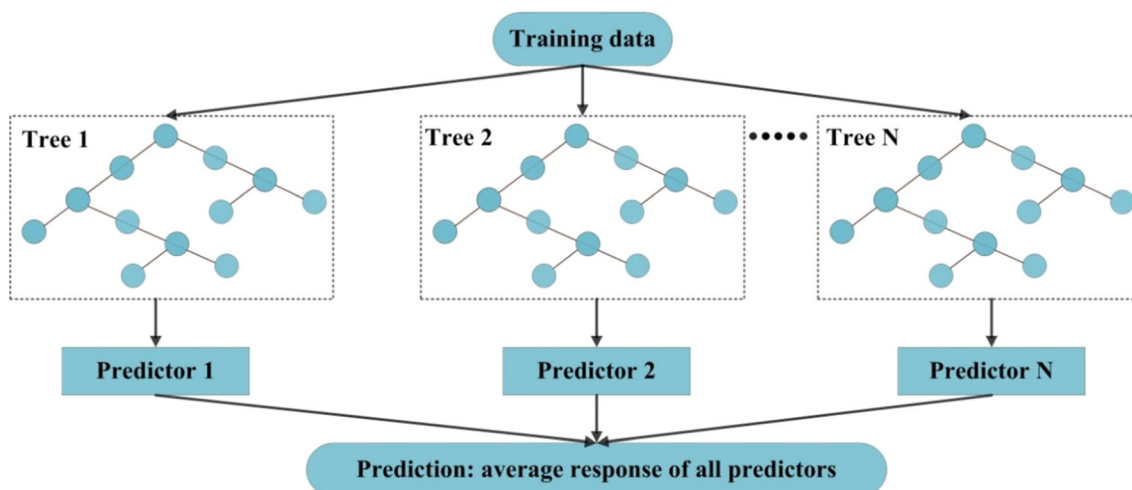

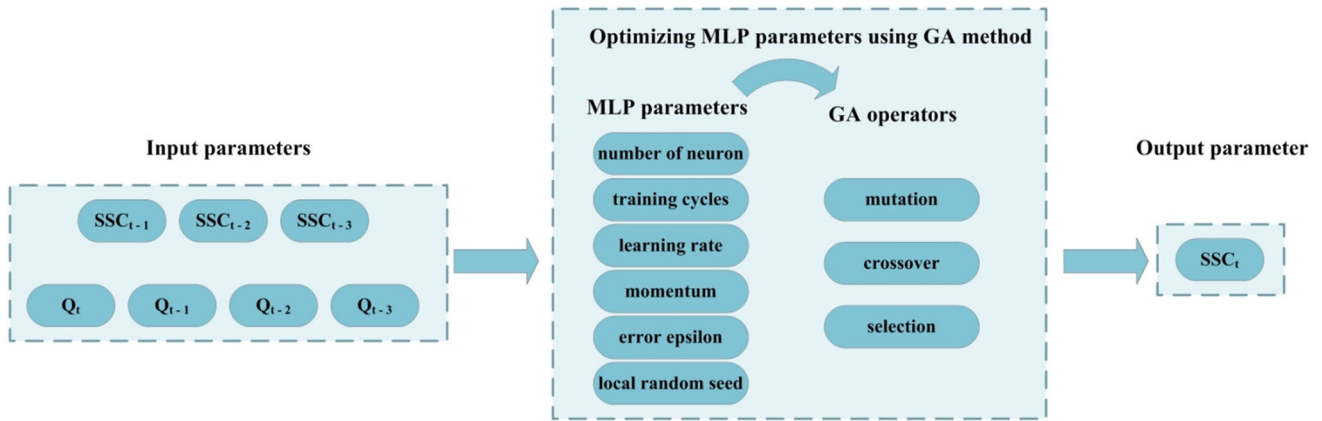
Fig. 3 Flowchart of the RF model

Fig. 4 Flowchart of the GA-MLP model

probability. Mutation means changing the value of one of the string cells from zero to one or from one to zero. After the amplification and mutation steps, the chromosomes are decoded and the value of the target function is calculated, then a fit is assigned. If necessary, the selection and reproduction steps, etc., are performed again. During this process, the average efficiency of the response population is expected to increase. The algorithm ends when a specific goal is met. For example, if a certain number of generations are created, the amount of merit of people reaches a certain amount, or a certain point is reached in the search space [4].

2.5 Stochastic gradient descent (SGD)

Stochastic gradient descent (SGD) is an iteration-based method for developing a derivative function called a target function, which is a stochastic approximation of the gradient descent (GD) method. In fact, the SGD is an algorithm to obtain the minimum value of a function in several iterative loops and the values for which the function takes its minimum value. The difference between a SGD and a standard GD is that, unlike the standard GD, which uses all training data to develop the target function, the SGD uses a randomly selected set of training data for optimization. This method has many applications in statistical and machine learning problems [34].

In the machine learning application, a problem usually appears in which, it is important to determine a function such as f of statistical data with one or more parameters and then, to define these parameters in such a way that the sum (or average) of the amounts of the function f for each piece of data statistically cause the minimum possible amount. It is assumed that there is a set of statistical data where the function f is determined only on the basis of the parameter θ, in which case by giving the i data from the data set to the function f a function of θ is obtained that called $\vartheta_i(\theta)$. Now, the problem is streamlined to find a θ that minimizes the following expression:
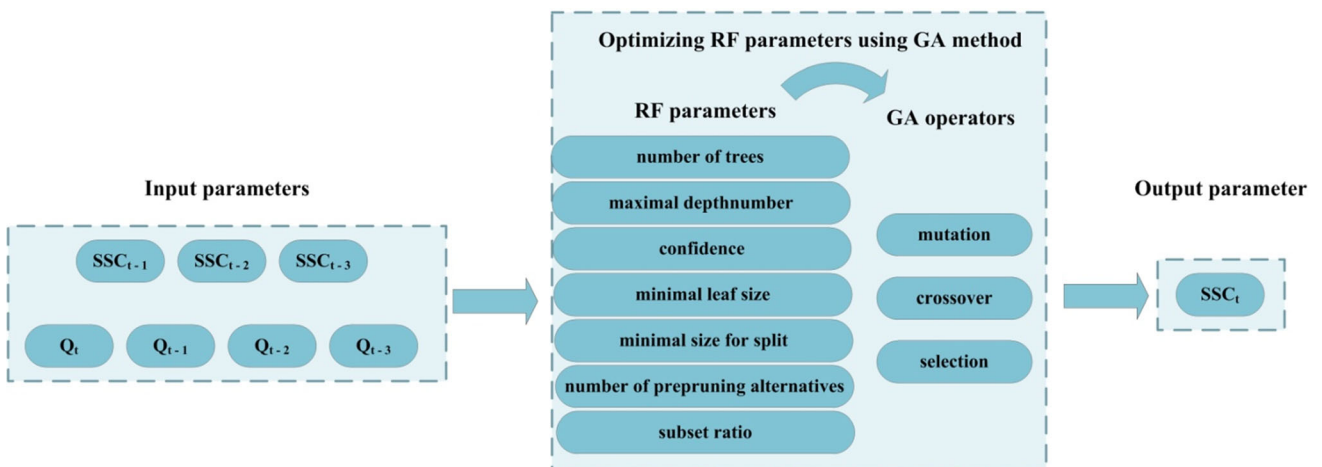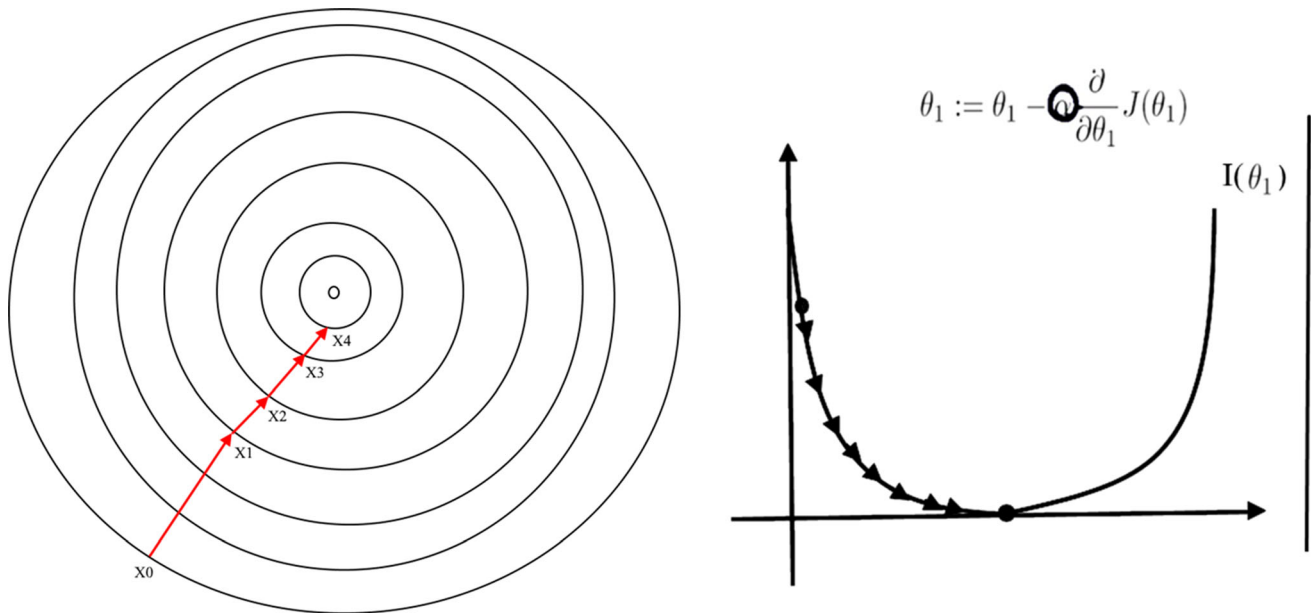

Fig. 5 Flowchart of the GA-RF model

Fig. 6 SGD operation for one-input and two-input functions

$$\vartheta(\boldsymbol{\theta}) = \left(\frac{1}{n}\right) \sum_{i=1}^{n} \vartheta_{\mathbf{i}}(\boldsymbol{\theta}) = \boldsymbol{E}|\vartheta_{\mathbf{i}}(\boldsymbol{\theta})| \qquad (1)$$

where $\vartheta(\boldsymbol{\theta})$ is a target function. In many cases, the target function becomes a simple function on which the application of the SGD method is not intricate and time consuming. In these cases, the standard GD is used, such as the family of exponential functions of a parameter used to appraise economic functions. However, since the standard or stochastic GD method requires the calculation of the objective function gradient in each loop, in some cases where the target function parameters are large or the training data set is very large, the calculation accomplished in each loop can be very time consuming and intricate. For this reason, a SGD is used, which in each loop accomplish this operation only for a section of the training data set that we have. In the SGD method, in each loop, the desired operation is not performed on only one member of the training data set that is randomly selected in each loop, and instead is performed on a subset of it where there are two reasons for this:

1. Dispersion reduces the amount acquired for the parameter in each loop and convergence is more stable.
2. Utilizing matrix operations that have a very fast execution.

2.6 GA-MLP

Determining the number of neurons in hidden layers, training cycles, learning rate, momentum, error epsilon,

and local random seed is one of the complicated modeling procedures in the MLP method. For this purpose, a hybrid algorithm including GA and MLP was utilized for modeling the *SSC* in rivers. This process started with the selection of a random initial population in which each individual consists of various numbers of hidden layer neurons. Then, the elite population with the best individuals is selected. The model is run repeatedly, and for each individual, the function is calculated and the obtained functions are stored. In the last stage, if the termination criteria had satisfaction results, the individual with the best function is saved. Otherwise, this process continues to find an appropriate population with a new function. The Levenberg–Marquardt algorithm is used mostly in the training stage of this process, but it has a random nature. By the usage of GA, the model is protected against this problem and chooses the best transfer function for the hidden and output layers. Figure 4 displays the flowchart presenting the GA-MLP method.

2.7 GA-RF

To promote the RF model's productivity and accuracy, the optimization of key parameters in the RF model structure is a necessary task [35]. To achieve higher performance, Zhou et al. [51] applied a novel approach where a few number of learners were utilized with high-quality. Based on [1], trees have different percentages in traditional RF's precision, for example, some of them can make incorrect predictions and underestimate the performance and efficiency of the model. Different strategies are used to

increase the model's accuracy, including a general climbing strategy, a greedy algorithm and so on; however, they have some deficiencies. For instance, using the greedy method to promote the performance of the RF model can give rise to becoming confined at local optima. Therefore, GA is implemented to solve that problem by choosing the best subset of features that is able to improve the RF model performance. Consequently, the RF model that is optimized by the GA has high accuracy in comparison with the traditional one. Figure 5 shows the flowchart illustrating the GA-RF method.

2.8 SGD implementation on MLP and RF

In the general implementation of a SGD, θ is the vector that includes all the parameters of the cost function. Firstly, θ is set to the ideal vector. Then, for each update of this vector, a member of the training data set is randomly selected, and at the $\boldsymbol{\alpha}$ rate, the vector of the cost function gradient at point θ is subtracted from θ:

$$\theta = \theta - \alpha \nabla_\theta \vartheta_i\left(\theta; x^{(i)}, y^{(i)}\right) \tag{2}$$

where ϑ is a cost function and $(x^{(i)}, y^{(i)})$ is a randomly selected member of the training data, and $\vartheta_i(\theta; x^{(i)}, y^{(i)})$ denotes the $\boldsymbol{i}$ sentence of the objective function. $\boldsymbol{\alpha}$ is the rate at which $\boldsymbol{\theta}$ is updated and has an experimental value that prolongs convergence if it is too small, and convergence may not happen if it is too large [45].

In another implementation, in each loop, a random member of the data set is not selected, but in each loop, the all data set is randomly rearranged once, and then the upgrade operation is accomplished in order of $\vartheta_1, \vartheta_2, \ldots, \vartheta_n$ where n shows the size of the training dataset. The following pseudocode indicates this implementation:

1. Give θ and $\boldsymbol{\alpha}$ the input value
2. Repeat until the minimum is reached
3. Randomly retrieve training data
4. Repeat for $\boldsymbol{i}$ from 1 to $\boldsymbol{n}$: $\theta = \theta - \alpha \nabla_\theta \vartheta_i(\theta; x^{(i)}, y^{(i)})\backslash$

Usually the update operation is not performed for ϑ from a single member of the training data set, but for a subset of this data called a small set. Figure 6 shows how GD works for the single-input and dual-input function.

SGD algorithm has advanced facilities including epochs, rho, L_1 or L_2 adjustment, momentum training, adaptive learning rate and rate annealing that enable high prediction precision in modeling by both MLP and RF models. In addition to optimizing MLP results using SGD, the network contains many hidden layers containing neurons with tanh (hyperbolic tangent function), rectifier (where x is the input value, select the maximum of $(0, x)$), max out (select the maximum input vector coordinates), and ExpRectifier (exponential rectifier function) activation operations.

The size of the weight updates is described by the user determined learning rate when adaptive learning rate is inactivated and is a function of the difference between the predicted value and the target value. That variance commonly named delta, which is only presented at the output layer. Backpropagation is applied to accurate estimation of the output at each hidden layer. The momentum is ramped up slowly since redundant momentum can result in oscillation. The rate annealing, momentum training, dropout rate annealing, and momentum training parameters activate if the adaptive learning rate is disabled.

2.9 Performance criteria

In this study, three performance indexes including correlation coefficient (CC), Willmott's index of agreement (WI), and scattered index (SI) are utilized in order to measure the model's accuracy. The mathematical description of *CC*, *SI*, and *WI* can be expressed, respectively, as follows:

Table 2 Implemented models and their input parameters

Input parameter	MLP based models	RF based models
SSC_{t-1}, Q_{t-1}	MLP-1, GA-MLP-1, SGD-MLP-1	RF-1, GA-RF-1, SGD-RF-1
SSC_{t-1}, SSC_{t-2}, Q_{t-1}, Q_{t-2}	MLP-2, GA-MLP-2, SGD-MLP-2	RF-2, GA-RF-2, SGD-RF-2
SSC_{t-1}, SSC_{t-2}, SSC_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3}	MLP-3, GA-MLP-3, SGD-MLP-3	RF-3, GA-RF-3, SGD-RF-3
SSC_{t-1}, Q_t, Q_{t-1}	MLP-4, GA-MLP-4, SGD-MLP-4	RF-4, GA-RF-4, SGD-RF-4
SSC_{t-1}, SSC_{t-2}, Q_t, Q_{t-1}, Q_{t-2}	MLP-5, GA-MLP-5, SGD-MLP-5	RF-5, GA-RF-5, SGD-RF-5
SSC_{t-1}, SSC_{t-2}, SSC_{t-3}, Q_t, Q_{t-1}, Q_{t-2}, Q_{t-3}	MLP-6, GA-MLP-6, SGD-MLP-6	RF-6, GA-RF-6, SGD-RF-6

Table 3 Parameters of the RF and GA-RF models (Minnesota station)

Parameter	RF	GA-RF-1	GA-RF-2	GA-RF-3	GA-RF-4	GA-RF-5	GA-RF-6
Number of trees	100	63	27	81	78	27	94
Maximal depth	10	5	27	65	36	27	86
Confidence	0.100	0.282	0.349	0.224	0.148	0.383	0.282
Minimal leaf size	2	41	55	98	1	55	55
Minimal split size	4	49	51	13	80	51	51
Number of repruning alternatives	3	0	18	68	55	77	18
Subset ratio	0.200	0.000	0.904	0.255	0.295	0.901	0.901

Table 4 Parameters of the RF and GA-RF models (for San Joaquin station)

Parameter	RF	GA-RF-1	GA-RF-2	GA-RF-3	GA-RF-4	GA-RF-5	GA-RF-6
Number of trees	100	1	81	27	27	63	81
Maximal depth	10	5	36	27	27	27	36
Confidence	0.100	0.3463	0.2131	0.3826	0.3826	0.2823	0.3463
Minimal leaf size	2	41	1	1	55	23	1
Minimal split size	4	1	80	62	51	49	20
Number of prepruning alternatives	3	3	55	77	0	77	3
Subset ratio	0.200	0.1423	0.2945	0.9009	0.9009	0.9685	0.2945

Table 5 Parameters of the MLP and GA-MLP models (for Minnesota station)

Parameter	MLP	GA-MLP-1	GA-MLP-2	GA-MLP-3	GA-MLP-4	GA-MLP-5	GA-MLP-6
Training cycles	200	77	77	77	77	77	77
Learning rate	0.0100	0.3991	0.3991	0.3991	0.1825	0.1825	0.3991
Momentum	0.9000	0.0774	0.0	0.0	0.0	0.0774	0.0
Error epsilon	0.0001	0.0	0.0	0.0	0.0	0.0	0.0
Local random seed	1992	77	1	77	77	77	56

Table 6 Parameters of the MLP and GA-MLP models (for San Joaquin station)

Parameter	MLP	GA-MLP-1	GA-MLP-2	GA-MLP-3	GA-MLP-4	GA-MLP-5	GA-MLP-6
Training cycles	200	4	39	71	63	77	77
Learning rate	0.0100	0.1423	0.1423	0.1423	0.1825	0.1825	0.1423
Momentum	0.9000	0.5456	0.0774	0.0774	0.5646	0.5646	0.5646
Error epsilon	0.0001	0.0	0.0	0.0	0.0	0.0	0.0
Local random seed	1992	29	18	77	23	23	56

$$CC = \frac{\left(\sum_{i=1}^{n} O_i P_i - \frac{1}{n}\sum_{i=1}^{n} O_i \sum_{i=1}^{n} P_i\right)}{\left(\sum_{i=1}^{n} O_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} O_i\right)^2\right)\left(\sum_{i=1}^{n} P_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} P_i\right)^2\right)} \tag{3}$$

$$SI = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}}{\overline{O}} \tag{4}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(O_i - P_i)^2}{\sum_{i=1}^{n}\left(|P_i - \overline{O}_i| + |O_i - \overline{O}_i|\right)^2}\right] \tag{5}$$

Table 7 General results of the computations for the RF, GA-RF and SGD-RF models (Minnesota station)

Model	Statistical parameters		
	CC	SI	WI
RF-1	0.919	0.369	0.957
GA-RF-1	0.931	0.341	0.963
SGD-RF-1	0.928	0.347	0.960
		0.34	
RF-2	0.933	0.337	0.965
GA-RF-2	0.936	0.330	0.967
SGD-RF-2	0.935	0.331	0.964
RF-3	0.935	0.332	0.966
GA-RF-3	0.939	0.322	0.968
SGD-RF-3	0.939	0.324	0.966
RF-4	0.937	0.327	0.966
GA-RF-4	0.939	0.319	0.968
SGD-RF-4	0.938	0.323	0.967
RF-5	0.938	0.325	0.968
GA-RF-5	0.944	0.308	0.971
SGD-RF-5	0.943	0.308	0.971
RF-6	0.937	0.325	0.967
GA-RF-6	0.944	0.308	0.971
SGD-RF-6	0.942	0.311	0.969

Table 9 General results of the computations for the RF, GA-RF, and SGD-RF models (San Joaquin station)

Model	Statistical parameters		
	CC	SI	WI
RF-1	0.848	0.427	0.917
GA-RF-1	0.859	0.400	0.920
SGD-RF-1	0.878	0.371	0.932
RF-2	0.872	0.393	0.931
GA-RF-2	0.878	0.384	0.934
SGD-RF-2	0.878	0.374	0.933
RF-3	0.867	0.400	0.927
GA-RF-3	0.879	0.375	0.933
SGD-RF-3	0.878	0.373	0.933
RF-4	0.884	0.366	0.938
GA-RF-4	0.892	0.353	0.942
SGD-RF-4	0.891	0.355	0.939
RF-5	0.884	0.371	0.937
GA-RF-5	0.886	0.367	0.939
SGD-RF-5	0.901	0.339	0.946
RF-6	0.876	0.380	0.932
GA-RF-6	0.880	0.377	0.934
SGD-RF-6	0.901	0.340	0.946

Table 8 General results of the computations for the MLP, GA-MLP and SGD-MLP models (Minnesota station)

Model	Statistical parameters		
	CC	SI	WI
MLP-1	0.930	0.460	0.935
GA-MLP-1	0.932	0.338	0.963
SGD-MLP-1	0.930	0.344	0.961
MLP-2	0.941	0.321	0.968
GA-MLP-2	0.945	0.307	0.971
SGD-MLP-2	0.941	0.323	0.964
MLP-3	0.942	0.325	0.967
GA-MLP-3	0.943	0.310	0.970
SGD-MLP-3	0.940	0.318	0.968
MLP-4	0.948	0.296	0.973
GA-MLP-4	0.949	0.294	0.973
SGD-MLP-4	0.933	0.334	0.964
MLP-5	0.951	0.311	0.971
GA-MLP-5	0.950	0.290	0.974
SGD-MLP-5	0.944	0.307	0.969
MLP-6	0.952	0.303	0.972
GA-MLP-6	0.952	0.291	0.975
SGD-MLP-6	0.948	0.299	0.973

in which P_i is predicted and O_i is the observed ith value and n is the number of data. CC is a statistical tool to determine the type and degree of relationship of one quantitative variable with another quantitative variable. CC is one of the criteria used to determine the correlation between two variables. The CC indicates the intensity of the relationship as well as the type of relationship (direct or inverse). This coefficient is between 1 and -1, and if there is no relationship between the two variables, it is equal to zero. If the range of data used is large, the amount of root mean square error (RMSE) in the modeling error evaluation section will also be high, which does not indicate that the model is inaccurate. For this purpose, in this study, the SI index was used, which is the result of dividing the RMSE index by the average test data. The closer the SI value to zero shows that the model is more accurate. WI is also one of the standardized indicators for calculating the model prediction error, the value of which is between zero and one. WI = 1 indicates the highest agreement and WI = 0 indicates no agreement. This index is also highly sensitive to limit values due to the use of difference squares [49]. Furthermore, Taylor diagrams were implemented for additional investigation of utilized hybrid GA-MLP, GA-RF, SGD-MLP, and SGD-RF models performances in SSC estimation [46].

Table 10 General results of the computations for the MLP, GA-MLP and SGD-MLP models (San Joaquin station)

Model	Statistical parameters		
	CC	SI	WI
MLP-1	0.881	0.366	0.936
GA-MLP-1	0.881	0.364	0.934
SGD-MLP-1	0.881	0.363	0.931
MLP-2	0.882	0.443	0.909
GA-MLP-2	0.885	0.360	0.938
SGD-MLP-2	0.877	0.420	0.911
MLP-3	0.882	0.399	0.928
GA-MLP-3	0.882	0.366	0.937
SGD-MLP-3	0.880	0.371	0.936
MLP-4	0.896	0.360	0.941
GA-MLP-4	0.899	0.340	0.947
SGD-MLP-4	0.901	0.335	0.941
MLP-5	0.893	0.375	0.938
GA-MLP-5	0.897	0.343	0.944
SGD-MLP-5	0.900	0.339	0.945
MLP-6	0.892	0.362	0.941
GA-MLP-6	0.893	0.347	0.940
SGD-MLP-6	0.897	0.343	0.943

3 Result and discussion

In the current study, MLP, RF, GA-MLP, GA-RF, SGD-MLP, and SGD-RF models with different input combinations were utilized for estimating the SSC in two stations and their results are examined in terms of accurate estimation. Moreover, due to the fact that there is not any direct way for dividing the entire data to train and test data sets in data driven methods, different proportions were utilized in the literature, e.g., Choubin [7] implemented a total of 63% of their data for training, whereas Qasem et al. [32] and Kargar et al. [19] used 67% of data, Dodangeh et al. [11], Asadi et al. [3], Shabani et al. [41] and Samadianfard et al. [39] utilized 70%, and Zounemat-Kermani et al. [53] exploited 80% of whole data for model development. Therefore, for the model development is this study, data were split into training (70%) and testing (30%). Accordingly, the time period of 2000–2012 was used to train the models and the 2013–2017 data were implemented as the test data set.

Table 2 provides the input parameters for each model where the SSC and Q parameters are shown in the current time (SSC_t and Q_t) and also with the previous daily lag times. It can be seen from Table 2 that six different scenarios were considered for SSC estimation utilizing different input combinations of SSC_{t-1}, SSC_{t-2}, SSC_{t-3}, Q_t, Q_{t-}

$_1$, Q_{t-2}, Q_{t-3} parameters. It is worthy to note that the selected scenarios are considered based on auto correlation of SSC and Q variables. For predicting SSC through MLP and RF and optimizing with SGD algorithm at Minnesota and San Joaquin rivers, the adaptive learning rate due to better performance was activated and presumed equal to 0.004. Moreover, the amounts of epochs, rho, L1 and L2 were presumed as 10, 1, 0.000009 and 0, respectively. In addition, the ExpRectifier was considered as activation operation because of its better efficiency in contrast to other activation operations for predicting SSC in SGD-MLP model. Additionally, Tables 3 and 4 show the default and optimized parameters used in RF and GA-RF models development in estimating the SSC for two different stations, including Minnesota and San Joaquin, respectively. Similarly, the related parameters of the MLP and GA-MLP models are displayed in Tables 5 and 6 for the mentioned stations.

Table 7 gives results of the RF, GA-RF, and SGD-RF models in Minnesota station. Diverse combinations are considered for these models, and the accurate estimations with high performance were obtained from RF-5 with CC of 0.938, SI of 0.325, and WI of 0.968 among standalone RF models and, GA-RF-5 and SGD-RF-4 with CC of 0.944 and 0.943, SI of 0.308, and WI of 0.971 among the hybrid RF ones. It is noticeable that the GA-RF-5 has a high accuracy in comparison with the RF-5 model and GA improved the model by reducing 5.2% of the SI parameter. In the GA-RF-5, number of trees is 27, maximal depth of 27, the confidence of 0.383, minimal leaf size of 55, minimal size for split of 51, number of prepruning alternatives of 77 and subset ratio of 0.9009 that are shown in Table 3. Evident is that the GA plays a vital role as an optimizer in SSC estimation. In Minnesota station, based on the results in Table 8, the MLP-4 has CC of 0.948, SI of 0.296, and WI of 0.973, and the GA-MLP-5 model has CC of 0.950, SI of 0.290, and WI of 0.974. These two models provide more accurate results in contrast to other models. Due to the presence of the GA, the GA-MLP-5 has more accurate predictions and GA enhanced the model's precision. As it is shown in Table 8, GA decreases the SI parameter by a factor of 2%. The GA-MLP-5 model has 77 training cycle, 0.1825 learning rate with the momentum of 0.0774, 0 error epsilon, and with a local random speed of 77 (Table 5). Also, it should be noted from Table 8 that although SGD had positive effects on increasing the prediction accuracy of the standalone MLP model, but in comparison with GA, it showed lower capability in reducing the prediction errors. By and large, in the Minnesota station, the GA-MLP-5 has more accurate performance in comparison with other optimized models. Also, among the SGD-MLP and SGD-RF models, obtained results indicated that SGD-MLP-6 with CC of 0.948, SI of
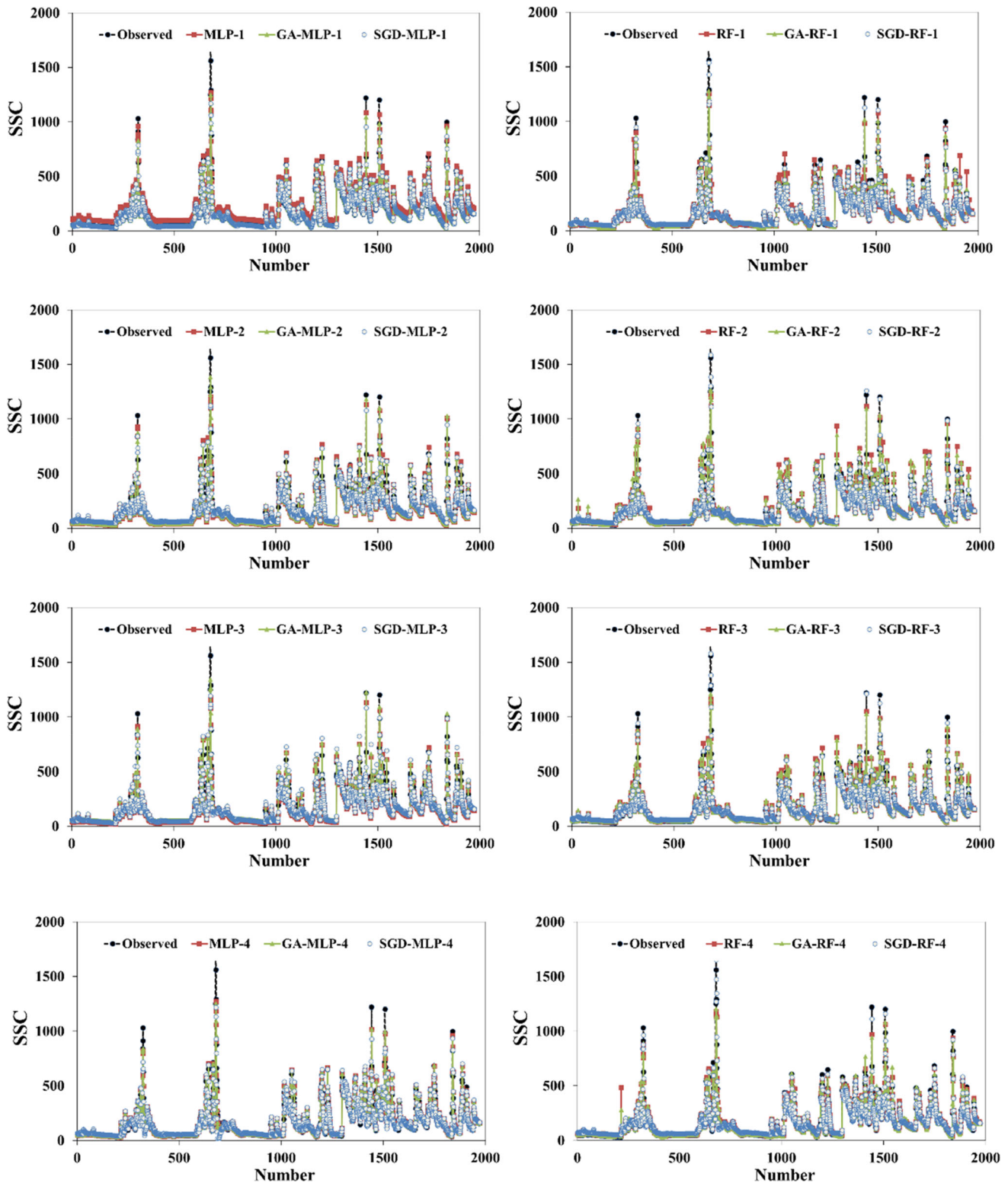
Fig. 7 Observed and estimated values with MLP, GA-MLP, SGD-MLP, RF, GA-RF, SGD-RF models at Minnesota station

0.299, and WI of 0.973 and SGD-RF-5 with CC of 0.943, SI of 0.308, and WI of 0.971 presented more accurate performances.

In Table 9, the results of the RF, GA-RF, and SGD-RF models in San Joaquin station are displayed. The RF-4 with CC of 0.884, SI of 0.366, and WI of 0.938 is considered as
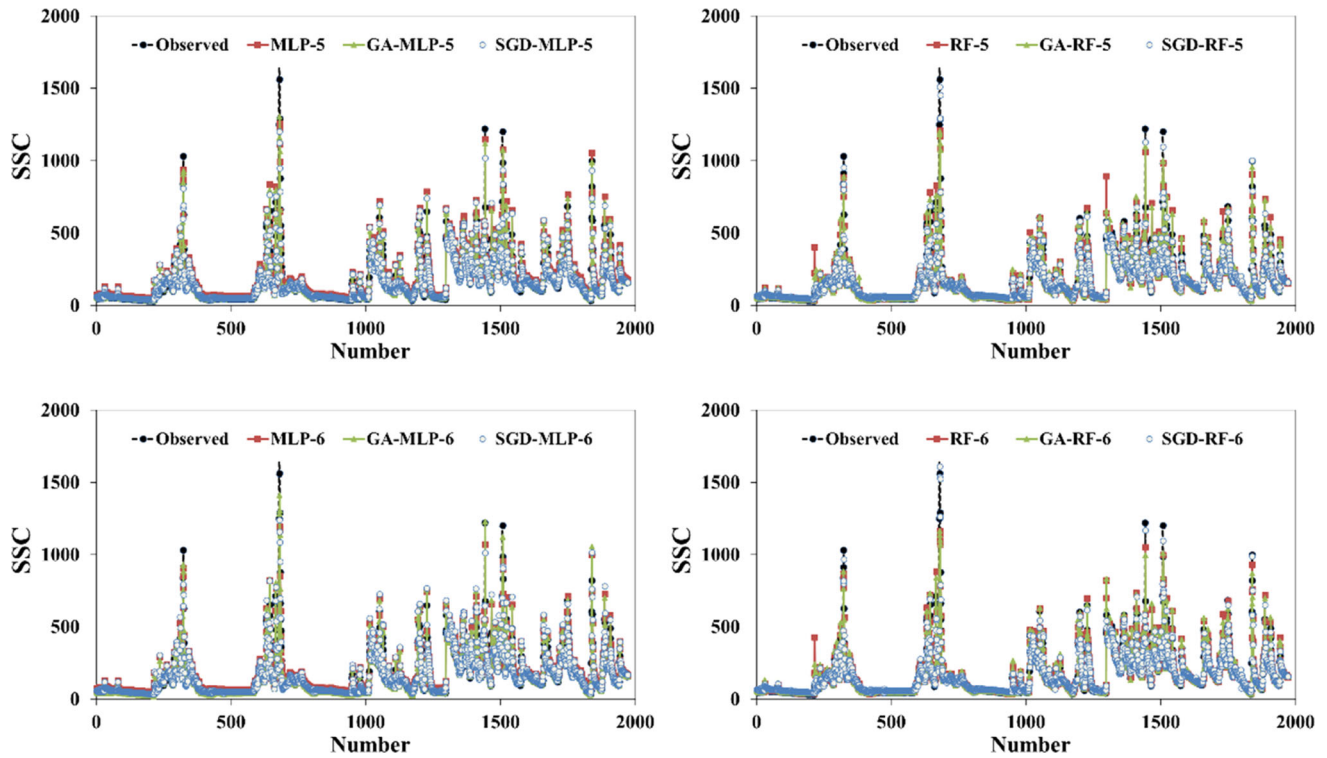
Fig. 7 continued

the best one among various combinations of RF, and the GA-RF-4 with CC of 0.892, SI of 0.353, and WI of 0.942 has the highest performance among other GA-RF models. GA-RF-4 has 27 trees with a maximal depth of 27, the confidence of 0.3826, minimal leaf size of 55, minimal size for split of 51, number of prepruning alternatives of 0 and subset ratio of 0.9009 (Table 4). According to Table 9, the GA and SGD improved the precision of all standalone models. In the San Joaquin station, SGD-MLP-4 and SGD-RF-5 have the best performance with CC of 0.901, SI of 0.335 and 0.339, WI of 0.941 and 0.941, respectively (Table 10). In this station, SGD has reduced the *SI* parameter by factors of 6.9% and 8.6% in contrast to the RF-5 and MLP-4 models, respectively.

As it is seen from Figs. 7 and 8, GA-RF-5 and GA-MLP-6 models provided more accurate results for SSC estimation in comparison with other models in the Minnesota station. Furthermore, in the San Joaquin station, SGD-RF-5 and SGD-MLP-4 models illustrated better performances. In both Minnesota and San Joaquin stations, hybrid optimized models gave accurate results in predicting the *SSC* values, whereas other models state poor outcomes. Based on the scatter plots presented at Figs. 9 and

10, in two studied stations, the most accurate MLP, RF, GA-MLP, GA-RF, SGD-MLP, and SGD-RF models are shown. Overall, the combination of input parameters has no significant effects on their outcomes. For instance, in Minnesota station, the GA-MLP-5 with SSC_{t-1}, SSC_{t-2}, Q_t, Q_{t-1}, Q_{t-2} as input parameters and in San Joaquin station, SGD-MLP-4 model with inputs of SSC_{t-1}, Q_t, Q_{t-1} were considered as the best models.

Additionally, Taylor diagrams are utilized in order to scrutinize standard deviation and correlation values between observed and estimated *SSC*. The RF-5, GA-RF-5, SGD-RF-5, MLP-4, GA-MLP-5 and SGD-MLP-6 models for the Minnesota station and RF-4, GA-RF-4, SGD-RF-5, MLP-4, GA-MLP-4 and SGD-MLP-4 models for San Joaquin station are displayed in Fig. 11. The length of the space from the green point (a reference point) to each point described as centered root mean square error (RMSE). Consequently, the minimum interval between the green point and the correspondent point shows the most precise model [46]. Pursuant to Fig. 11, in the Minnesota station, the red point (GA-MLP-5) is the nearest point to the reference point, and also, the light blue point (SGD-MLP-4) has the least distance from the green point at the San
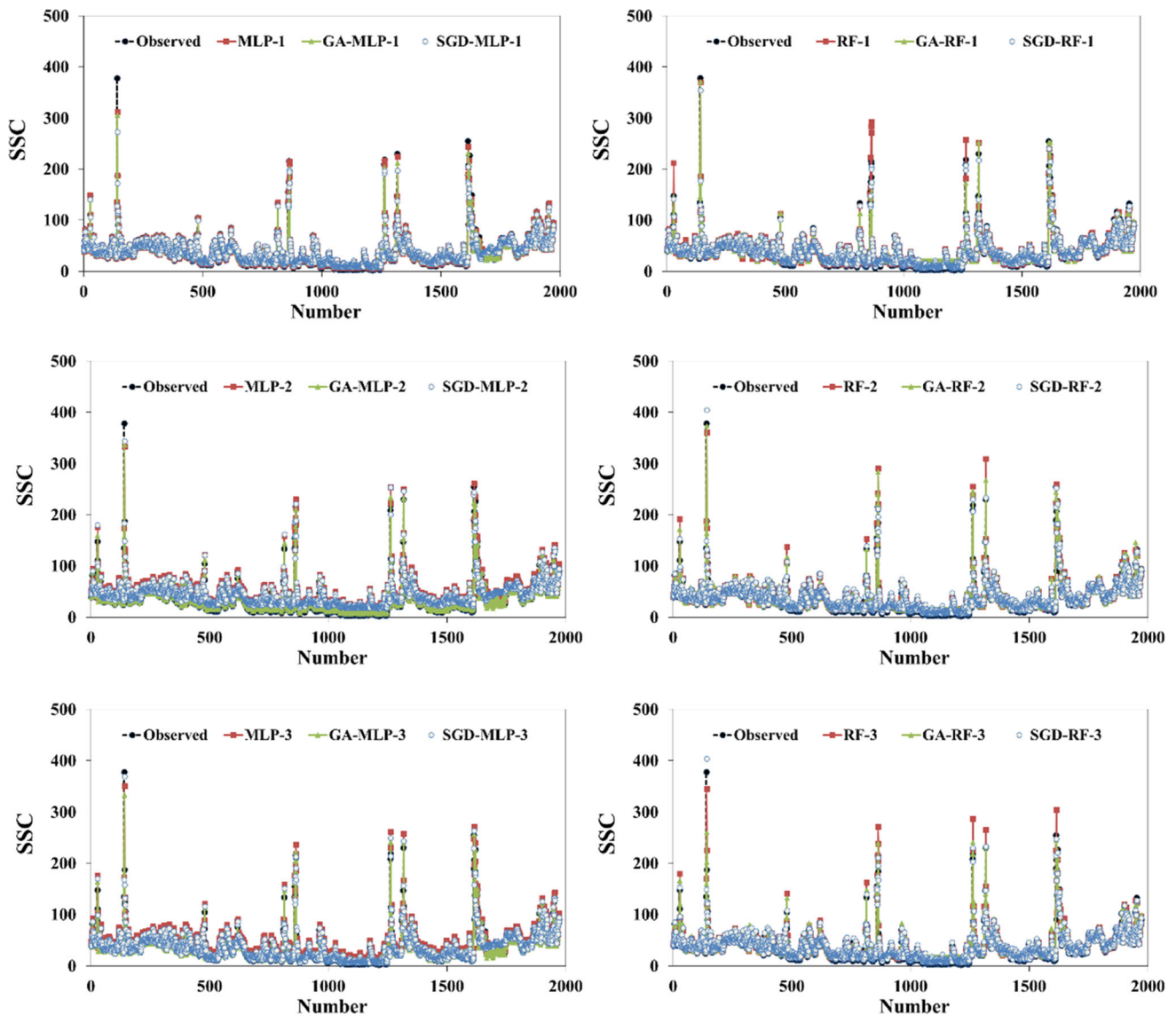
Fig. 8 Observed and estimated SSC values with MLP, GA-MLP, SGD-MLP, RF, GA-RF, and SGD-RF models at San Joaquin station

Joaquin station, thus providing the best estimates for the SSC.

Sediment transport has complicated process, and estimation of *SSC* is a quite difficult issue as a fundamental hydrological problem. Numerous conventional regression alternatives are available in the literature; however, their applicability on rivers in different climate conditions is a challenging task. In recent decade, runoff-suspended sediment load modeling attracts interests in implementation of machine learning algorithms to develop robust models with high computational ability. So, this study focused on the

applicability of MLP and RF; hybridized with GA and SGD methods for SSC prediction. Variety of scenarios were implemented for the modeling to find the superior combination from historical records. It was found that SSC_{t-1}, Q_t, Q_{t-1} and SSC_{t-1}, SSC_{t-2}, Q_t, Q_{t-1}, Q_{t-2}, combinations provided more accurate results which means that one and two days ahead records could be joined the upcoming day's SSC value. The results obtained in this study showed that GA-MLP, GA-RF, SGD-MLP, and SGD-RF models successfully estimated SSC in two different rivers. Extension of the present study may be
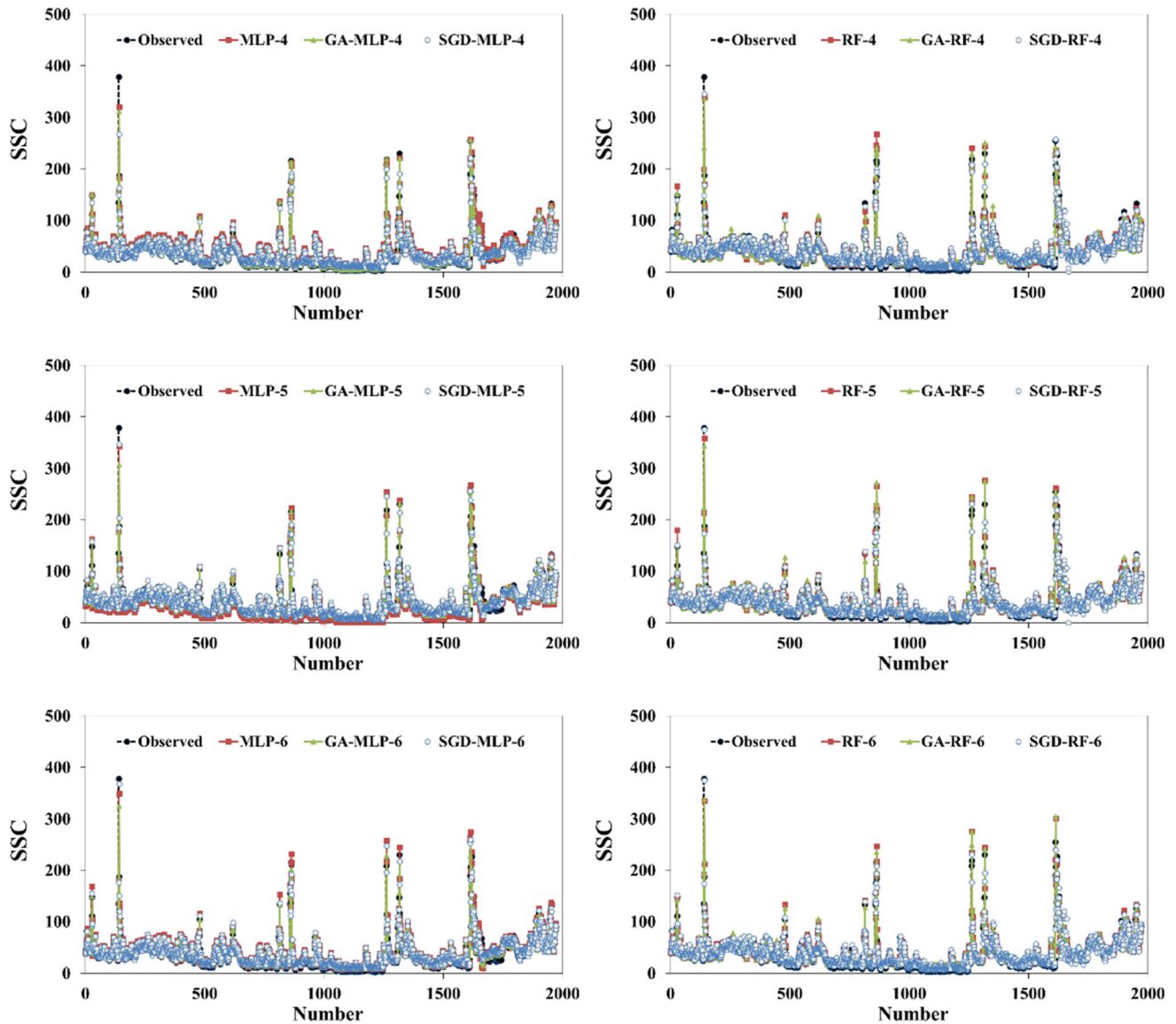
Fig. 8 continued

considered as the application of the suggested algorithms in other rivers with different climate conditions. In the current research, GA and SGD algorithms as a metaheuristic algorithms were implemented for optimization of the MLP and RF models. Future studies may consider application of the alternative optimization algorithms for SSC modeling.

4 Conclusions

The sedimentation problem is an essential issue in hydrological sciences, due to its imperative role in river hydraulic. In the current research, MLP and RF methods and their hybrid forms with GA and SGD are proposed to estimate SSC at of Minnesota and San Joaquin rivers. For each method, different combinations as input parameters were implemented for the modeling. The performance of
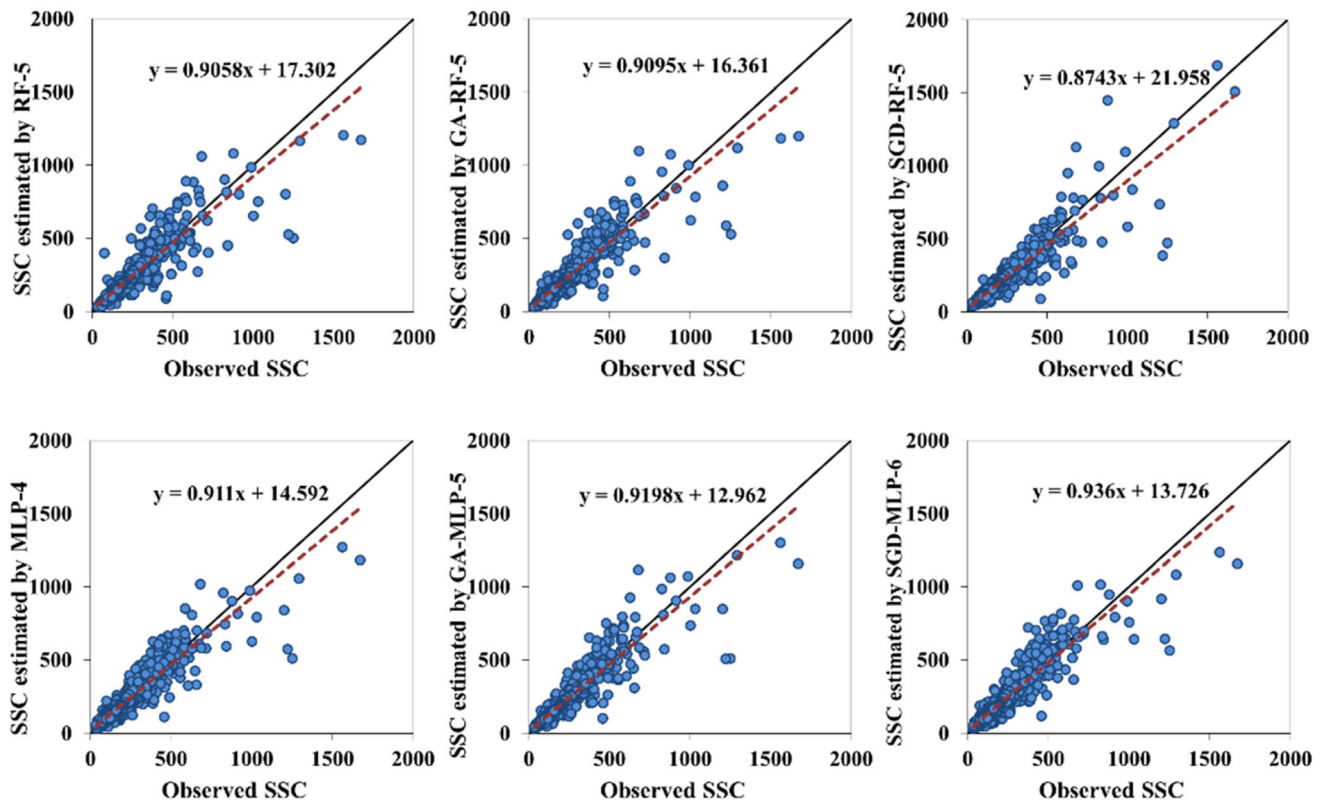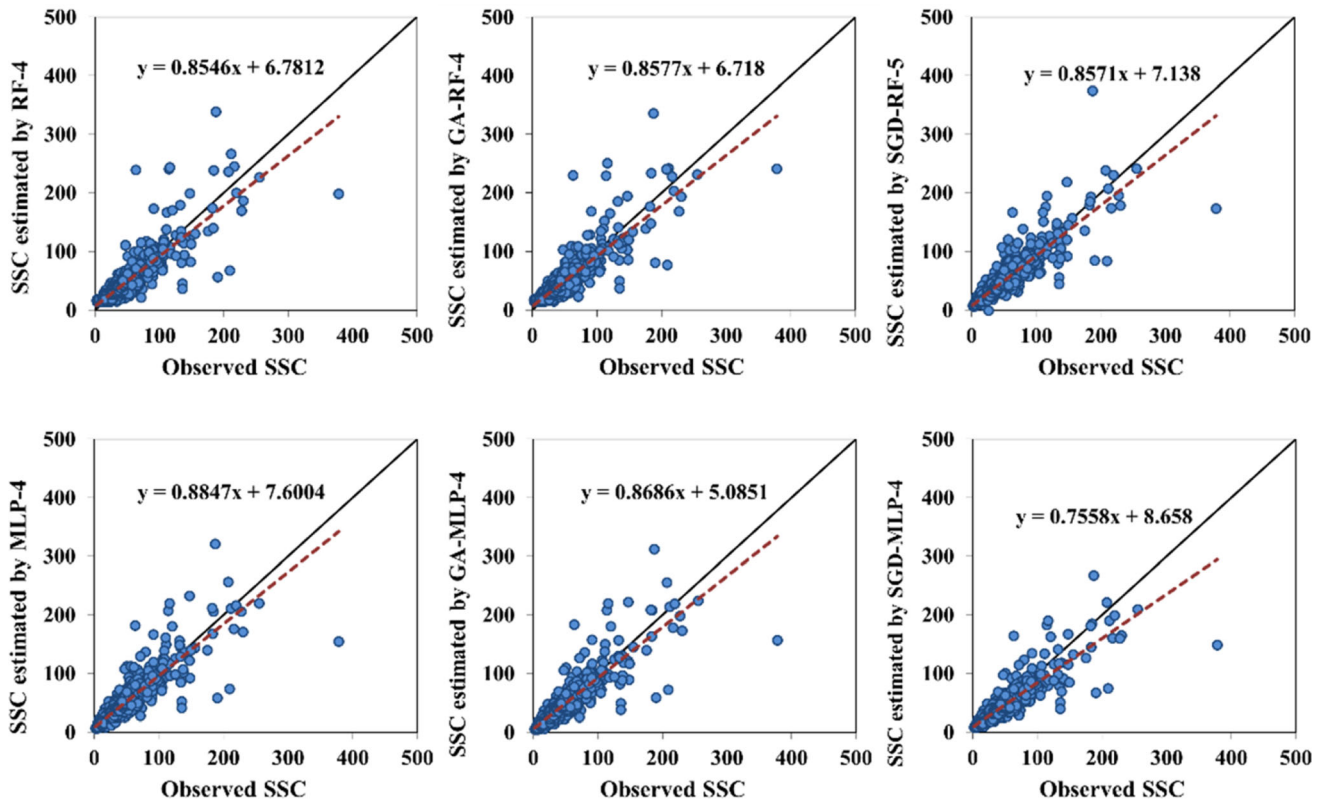
Fig. 9 The scatter plots of observed and estimated SSC values of the most accurate MLP, GA-MLP, SGD-MLP, RF, GA-RF, SGD-RF models at Minnesota station
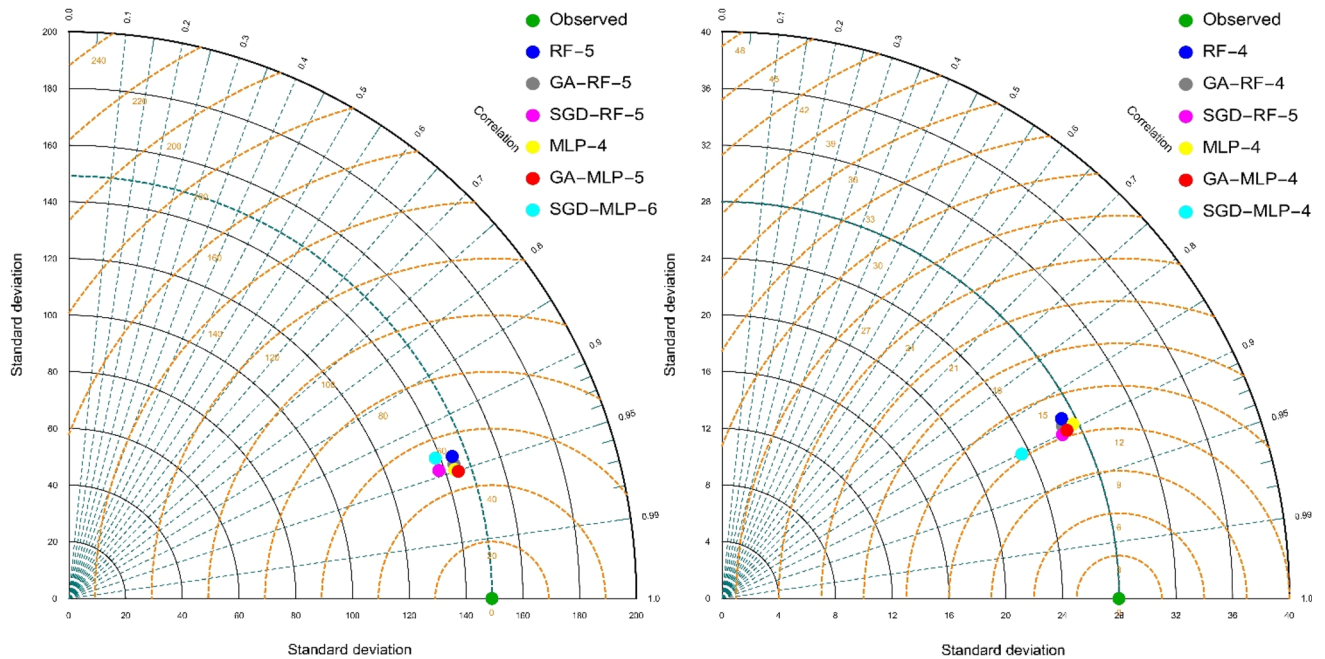
models was examined in order to discover the best one in this process. The results demonstrated that the accuracy of standalone models was increased using GA and SGD optimized models. Overall results indicated that GA-MLP and SGD-MLP were found to be the robust techniques for modeling SSC relying on the statistical results obtained based on various indexes, including SI, CC, and WI. Moreover, assessing the precision of models in estimating

SSC revealed that the standalone model's superiority in predicting *SSC* was less than their hybrid counterparts. Conclusively, these models can be used in water resource management and alternative fields of engineering with a high degree of confidentiality. Additional to the streamflow and suspended sediment load variables, incorporating the further hydrological parameters for the modeling seems to be worthy for improving the model credibility.

Fig. 10 The scatter plots of observed and estimated SSC values of the most accurate MLP, GA-MLP, SGD-MLP, RF, GA-RF, SGD-RF models at San Joaquin station

Fig. 11 Taylor diagrams of the studied models at both stations

Declarations

References

1. Adnan MN, Islam MZ (2016) Optimizing the number of trees in a decision forest to discover a subforest with high ensemble accuracy using a genetic algorithm. Knowl-Based Syst 110:86–97

2. Altunkaynak A (2009) Sediment load prediction by genetic algorithms. Adv Eng Softw 40(9):928–934

3. Asadi E, Isazadeh M, Samadianfard S, Ramli MF, Mosavi A, Nabipour N, Shamshirband S, Hajnal E, Chau KW (2020) Groundwater quality assessment for sustainable drinking and irrigation. Sustainability 12:177

4. Bäck T, Fogel DB, Michalewicz Z (2000) Evolutionary computation 1: Basic algorithms and operators. Institute of Physics Pub, Bristol

5. Breiman L (2001) Random forests. Mach Learn 45:5–32

6. Chong EK, Zak SH (2013) An introduction to optimization. Wiley, NY

7. Choubin B (2020) Spatial hazard assessment of the PM10 using machine learning models in Barcelona Spain. Sci Total Environ 701:134474

8. Cobaner M, Unal B, Kisi O (2009) Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. J Hydrol 367:52–61

9. Cutler A, Cutler DR, Stevens JR (2011) Random forests. In: Ensemble Machine Learning, pp 157–176

10. Dang MN (2021) Integration of ANFIS with PCA and DWT for daily suspended sediment concentration prediction. Water SA 47:200–209

11. Dodangeh E, Choubin B, Eigdir AN (2019) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Sci Total Environ 9:135983

12. Douglas RK, Nawar S, Alamar MC, Mouazen AM, Coulon F (2018) Rapid prediction of total petroleum hydrocarbons concentration in contaminated soil using vis-NIR spectroscopy and regression techniques. Sci Total Environ 616–617:147–155

13. Du KL, Swamy MN (2006) Neural networks in a soft computing framework. Springer Science & Business Media, Berlin

14. Frings RM, Kleinhans MG (2008) Complex variations in sediment transport at three large river bifurcations during discharge waves in the river Rhine. Sedimentology 55:1145–1171

15. Gallagher K, Sambridge M (1994) Genetic algorithms: a powerful tool for largescale nonlinear optimization problems. Comput Geosci 20(7):1229–1236

16. Ghose D, Samantaray S (2018) Modelling sediment concentration using back propagation neural network and regression coupled with genetic algorithm. Procedia Comput Sci 125:85–92

17. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley Longman Publishing Co., Inc.

18. Holland JH (1992) Genetic algorithms. Sci Am 267:66–72

19. Kargar K, Samadianfard S, Parsa J, Nabipour N, Shamshirband S, Mosavi A, Chau KW (2020) Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms. Eng Appl Comput Fluid 14(1):311–322

20. Kisi O (2010) River suspended sediment concentration modeling using a neural differential evolution approach. J Hydrol 389:227–235

21. Kisi O, Guven A (2010) A machine code-based genetic programming for suspended sediment concentration estimation. Adv Eng Softw 41:939–945

22. Kisi O, Zounemat-Kermani M (2016) Suspended sediment modeling using neurofuzzy embedded fuzzy c-means clustering technique. Water Resour Manage 30:3979–3994

23. Kumar D, Pandey A, Sharma N, Flügel WA (2016) Daily suspended sediment simulation using machine learning approach. CATENA 138:77–90

24. Liu QJ, Shi ZH, Fang NF, Zhu HD, Ai L (2013) Modeling the daily suspended sediment concentration in a hyperconcentrated river on the Loess Plateau, China, using the Wavelet–ANN approach. Geomorphology 186:181–190

25. Liu QJ, Zhang HY, Gao KT, Xu B, Wu JZ, Fang NF (2019) Time-frequency analysis and simulation of the watershed suspended sediment concentration based on the Hilbert-Huang transform (HHT) and artificial neural network (ANN) methods: A case study in the Loess Plateau of China. CATENA 179:107–118

26. Malik A, Kumar A, Piri J (2017) Daily suspended sediment concentration simulation using hydrological data of Pranhita River Basin, India. Comput Electron Agric 138:20–28

27. McBean EA, Al-Nassri S (1988) Uncertainty in suspended sediment transport curves. J Hydrol Eng, ASCE 114(1):63–74

28. Mehri Y, Nasrabadi M, Omid MH (2021) Prediction of suspended sediment distributions using data mining algorithms. Ain Shams Engineering Journal

29. Meshram SG, Safari MJS, Khosravi K, Meshram C (2021) Iterative classifier optimizer-based pace regression and random forest hybrid models for suspended sediment load prediction. Environ Sci Pollut Res 28(9):11637–11649

30. Mohammadi B, Guan Y, Moazenzadeh R, Safari MJS (2021) Implementation of hybrid particle swarm optimization-differential evolution algorithms coupled with multi-layer perceptron for suspended sediment load estimation. CATENA 198:105024

31. Prasad AM, Iverson LR, Andy L (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9(2):181–199

32. Qasem SN, Samadianfard S, Sadri Nahand H, Mosavi A, Shamshirband S, Chau KW (2019) Estimating daily dew point temperature using machine learning algorithms. Water 11:582

33. Rajaee T, Mirbagheri SA, Zounemat-Kermani M, Nourani V (2009) Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Sci Total Environ 407:4916–4927

34. Robbins H, Monro S (1951) A Stochastic Approximation Method. Ann Math Stat 22:400–407

35. Rodriguez-Galiano V, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez J (2012) An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens 67:93–104

36. Roushangar K, Aghajani N, Ghasempour R, Alizadeh F (2021) The potential of ensemble WT-EEMD-kernel extreme learning machine techniques for prediction suspended sediment concentration in successive points of a river. J Hydroinf 23:655–670

37. Safari MJS (2020) Hybridization of multivariate adaptive regression splines and random forest models with an empirical equation for sediment deposition prediction in open channel flow. J Hydrol 590:125392

38. Safari MJS, Aksoy H, Mohammadi M (2016) Artificial neural network and regression models for flow velocity at sediment incipient deposition. J Hydrol 541:1420–1429

39. Samadianfard S, Hashemi S, Kargar K, Izadyar M, Mostafaeipour A, Mosavi A, Nabipour N, Shamshirband S (2020) Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. Energy Rep 6:1147–1159

40. Samantaray S, Sahoo A (2021) Prediction of suspended sediment concentration using hybrid SVM-WOA approaches. Geocarto International.

41. Shabani S, Samadianfard S, Sattari MT, Mosavi A, Shamshirband S, Kmet T, Várkonyi-Kóczy AR (2020) Modeling pan evaporation using gaussian process regression K-nearest neighbors random forest and support vector machines. Comparative Anal Atmos 11:66

42. Shirzad A, Safari MJS (2019) Pipe failure rate prediction in water distribution networks using multivariate adaptive regression splines and random forest techniques. Urban Water J 16(9):653–661

43. Singh N, Chakrapani GJ (2015) ANN modelling of sediment concentration in the dynamic glacial environment of Gangotri in Himalaya. Environ Monit Assess 187(8):494

44. Sivakumar B, Jayawardena AW (2002) An investigation of the presence of low-dimensional chaotic behaviour in the sediment transport phenomenon. Hydrol Sci J 47:37–41

45. Taddy M (2019) Business data science: Combining machine learning and economics to optimize, automate, and accelerate business decisions. McGraw-Hill, New York

46. Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res: Atmos 106:7183–7192

47. Verstraeten G, Poesen J (2001) Factors controlling sediment yield from small intensively cultivated catchments in a temperate humid climate. Geomorphology 40:123–144

48. Ward P, Balen RT, Verstraeten G, Renssen H, Vandenberghe J (2009) The impact of land use and climate change on late Holocene and future suspended sediment yield of the Meuse catchment. Geomorphology 103:389–400

49. Willmott CJ (1982) Some comments on the evaluation of model performance. Bull Am Meteor Soc 63:1309–1313

50. Zhang FX, Wai OWH, Jiang YW (2010) Prediction of sediment transportation indeep bay (Hong Kong) using genetic algorithm. J Hydrodyn, Ser B 22(5):599–604

51. Zhou ZH, Wu J, Tang W (2002) Ensembling neural networks: many could be better than all. Artif Intell 137:239–263

52. Zounemat-Kermani M, Kisi O, Adamowski J, Ramezani-Charmahineh A (2016) Evaluation of data driven models for river suspended sediment concentration modeling. J Hydrol 535:457–472

53. Zounemat-Kermani M, Seo Y, Kim S, Ghorbani MA, Samadianfard S, Naghshara S, Kim NW, Singh VP (2019) Can decomposition approaches always enhance soft computing models? Predicting the dissolved oxygen concentration in the St. Johns River, Florida. Applied Sciences, 9:2534.

Authors and Affiliations

Saeed Samadianfard[1] (ID) · **Katayoun Kargar[2]** · **Sadra Shadkani[1]** · **Sajjad Hashemi[1]** · **Akram Abbaspour[1]** · **Mir Jafar Sadegh Safari[3]**

✉ Saeed Samadianfard
s.samadian@tabrizu.ac.ir

Katayoun Kargar
katayoun.kargar@ryerson.ca

Sadra Shadkani
sshadkani@gmail.com

Sajjad Hashemi
hashemisajjad2009@gmail.com

Akram Abbaspour
akabbaspour@yahoo.com

Mir Jafar Sadegh Safari
jafar.safari@yasar.edu.tr

[1] Department of Water Engineering, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

[2] Department of Civil Engineering, Ryerson University, Toronto, ON, Canada

[3] Department of Civil Engineering, Yaşar University, Izmir, Turkey