Advances in Environmental Technology 10(2) 2024, 102-117.

Iranian Reasearch Organization for Science and Technology (IROST)

Advances in Environmental Technology

Journal home page: https://aet.irost.ir

Estimating daily suspended sediment by intelligent and traditional models (Case Study: Kasalian and Rood Zard watersheds, Iran)

Adele Alijanpour Shalmani[1*], Ali Reza Vaezi[1], Mahmoud Reza Tabatabaei[2]

[1]*Department of Soil Science, University of Zanjan, Zanjan, Iran*

[2]*Soil Conservation and Watershed Management Research Institute, Agricultural Research, Education and Extension Organization (AREEO), Tehran, Iran*

ARTICLE INFO

Document Type:
Research Paper

Article history:
Received 13 April 2021
Received in revised form 26 February 2024
Accepted 4 March 2024

Keywords:
Daily flow discharge
Daily precipitation
Intelligent model
Traditional model
Watershed

ABSTRACT

Suspended sediment load is an indicator of erosion in watersheds. Therefore, accurately estimating the daily suspended sediment load (DSSL) is an important issue in river engineering. In this research, Artificial Neural Networks (ANN), Genetic Expression Programming (GEP) intelligent models, and the traditional Sediment rating curve (SRC) model were used to estimate DSSL in the Kasilian and Rood Zard watersheds in Iran. The input data to the models included instantaneous flow discharge (Q), average daily flow discharge (Q_i), average daily flow discharge with a delay of three days (Q_{i-1},Q_{i-2},Q_{i-3}), average daily precipitation (P_i), and average daily precipitation with a delay of three days (P_{i-1},P_{i-2},P_{i-3}); the output data was DSSL. In this research, the self-organizing map (SOM) artificial neural network was used for data clustering, and gamma test (GT) methods were used to obtain the best combination of input variables to intelligent models. The results showed that the best models for estimating DSSL in the Kasilian and Rood Zard watersheds were respectively the ANN model with an activation function of tangent sigmoid with the best combination of input variables (Q_{i-1},Q_{i-2},Q_{i-3},P_i,P_{i-1},P_{i-2},P_{i-3}) and the GEP model with the input variables Q_i,Q_{i-1},Q_{i-2},P_i,P_i,P_{i-1},P_{i-2},P_{i-3}. The statistical values of the ANN-tangent sigmoid model for the Kasilian watershed were MAE=231.4 (ton day^{-1}), RMSE=578.6 (ton day^{-1}), NSE =0.98, and R^2=0.98; these values for the GEP model in the Rood Zard watershed were MAE=475.7 (ton day^{-1}), RMSE=1671.9 (ton day^{-1}), NSE=0.99, and R^2=0.99. The SRC model in the Kasilian watershed with R^2=0.34 and NSE=0.08 and the Rood Zard watershed with R^2=0.59 and NSE=-0.11 showed the low accuracy of this model in estimating DSSL.

*Corresponding author Tel.: +989112373413
E-mail: adele.alijanpour@gmail.com

DOI: 10.22104/AET.2024.4846.1309

1. Introduction

Erosion due to the degradation of forests and pastures, the changes in land use, and improper agriculture increase the sediment in rivers in the watershed. Eroded sediments, after entering the river, usually enter the reservoirs of dams and cause many problems, such as reducing the volume of the reservoir, reducing the reservoir water supply, flood control, etc. [1]. Therefore, accurately recognizing the amount of DSSL is very useful for soil management and protection, improvement of water quality, river engineering, and damping [2]. The estimate of DSSL is done directly and indirectly. In the direct method, the data do not have the required quantity and quality. The problems with data quantity are the lack of sufficient specialized personnel, the failure of sedimentation devices, and the high cost of direct measurement. Also, since most DSSL measurements are done at low flow discharge, the data does not have the required quality [3]. Due to the quantitative and qualitative problems of direct measurement of DSSL, methods for modeling and estimating it in watersheds are used. One of the methods used in this field is Sediment rating curve (SRC) [4]. SRC is a simple and traditional method for estimating DSSL. This method establishes regression assumptions, such as predicted mathematical relationships between variables, data normality, data independence, and data reliability [5]. Also, using the SRC, due to the conversion of data from logarithmic space to the arctic space, makes it possible to estimate high sediment values below the real value [6]. The relationship between the formation of DSSL in the river watershed and the physical, geochemical, and biological processes and human interventions that cause it are often very complex and nonlinear. All these factors are very difficult to estimate [7]. Hence, due to the lack of confidence in the SRC method to accurately estimate DSSL and the uncertainty in the full recognition of processes affecting the erosion and sedimentation of watersheds, the DSSL modeling is the focus rather than a quantitative relationship. It is important to pay attention to the response of the watershed to the input factors that create different behaviors in them. In this regard, the use of Artificial Neural Networks (ANN) and Genetic Expression Programing (GEP) are good tools for the accurate modeling of DSSL in watersheds [8,9,10]. Some researchers have used these models in a wide range of topics. Emamgholizadeh and Karimi Demneh [1] compared three intelligence models (GEP, ANN, and ANFIS models) with the SRC method. The data used were daily flow rate and sediment flow rate in two hydrometric stations of the Kasilian and Talar Rivers. Their results showed that all the intelligent models performed better in estimating suspended sediment load than the SRC method. Rajaee et al. [11] used Multilayer Perceptron Neural Network (MLP) models, multivariate linear regression, and SRC to estimate the amount of DSSL in black rivers in the United States. Daily flow discharge data were used as input and DSSL as output. The results showed that the MLP neural network model was more accurate than the regression and SRC models in DSSL estimation. To estimate the amount of DSSL in Rio Valenciano and Quebrada Blanca Stations in the United States, Kisi and Aytac [12] used linear genetic programming, ANN, and SRC models. The input and output data of the models were the flow discharge and Suspended sediment concentration (SSC), respectively. The results at both stations indicated the superiority of the linear genetic programming model compared to other models. Boukhrissa et al. [13]estimated the amount of DSSL of the El Kebir catchment in Algeria using two methods: SRC and ANN. The models used daily water discharge and daily suspended sediment data as inputs and outputs. Their results showed that the ANN model with R^2=0.99 and RMSE=0.045 (ton day^{-1}) had a more accurate estimate of DSSL. Sheikhipor et al. [14] examined suspended sediment load using GEP in the Sistan River during the statistical period from 1996 to 2012. Results for the training data with RMSE=0, MBE=4.69×10^{-4}, and R^2=1, and those for the validation data with RMSE=0, MBE=2.4×10^{-4}, and R^2=1 represented the estimation of high accuracy DSSL using this method. Abbaspour et al. [15] compared the ANN and SRC methods to estimate the DSSL at Cham Anjir Hydrometric Station in Lorestan Province, Iran. Daily flow discharge and daily DSSL were used for modeling. The results showed that the ANN model had more power in estimating DSSL with MSE=0.0187 and R^2=0.95. Joshi et al. [16] stated that the SSC had a nonlinear relationship with the hydrological characteristics of the watershed. For this reason, the use of traditional SRC methods could not accurately estimate the SSC. As a result, they used ANN methods to model it. The region studied was the Gangotri glacier in the Himalayas. Their results showed that the ANN model with R^2=0.81 and RMSE=1.6 g/l had a more accurate estimation of SSC than SRC with R2=0.25 and RMSE=6.4 gr/l. Nivesh and Kumar [17] used ANN and Multiple linear regression (MLR) models to estimate DSSL in the Vamsadhara catchment in India. The study period from 1997 to 2000 included rainfall and discharge data as inputs and sediment concentrations as output data for the models. Seventy percent of the data was used for training and 30% to validate the models. The results showed that the ANN model with R^2=0.97 and RMSE=110.15 gr/l had a high accuracy in DSSL estimation. Because of the quantitative and qualitative problems of the data in the direct measurement of DSSL, its estimation is done using various methods. In most studies, modeling methods have been used to estimate the DSSL. However, in these studies, the pre-processing of data has been neglected or given less attention. Also, the inappropriate division of data was done randomly in the training and validation groups for entering the models. Also, in most studies, the role of dynamic and effective daily precipitation data in the production of DSSL is ignored, and only daily flow discharge is used to estimate the amount of DSSL in the watersheds. Lack of attention to these cases will lead to an error in estimating DSSL and the inability to generalize it for the watersheds studied. Also, the use of the SRC method due to the regression nature of the model and the expansion of the mathematical relationship; the inadequate understanding of the relationship between DSSL and other variables in nature will lead to errors in the estimation of DSSL. Therefore, the objectives of this research include:

1. Data clustering using a Self-organizing map (SOM) neural network to increase the power of model generalization.
2. Using the WinGamma Method to reduce the dimensions of variables and the best variable composition for entering into the intelligent models.
3. Use of daily precipitation and daily flow discharge variables with a delay of three days for more accurate DSSL estimation.
4. Use intelligent models to estimate DSSL and compare them with the SRC model.

2. Material and methods

2.1. Study area and data

The areas studied in this research included the Kasilian watershed, with a humid climate in the north of Iran, and the Rood Zard watershed, with semi-arid climate in the southwest of Iran. The Kasilian watershed consists of vast mountainous and jungle areas located in Mazandaran Province. The watershed is located at 35° 58' to 36° 19' in north latitudes and 52° 53' to 53° 15' in the eastern longitude (Figure 1). The Rood Zard watershed is located at 31° 21' to 31° 41' in the north latitudes and 49° 39' to 50°10' in the eastern longitude east of Khoozestan Province (Figure 2). Physiographic characteristics for the Kasilian and Rood Zard watersheds are shown in Table 1. This study used the Shirgah-Kasilian hydrometric station in the Kasilian watershed and the Mashin hydrometric station in the Rood Zard watershed. The data used in this research for the Kasilian watershed included 491 information records for a 41-years (1971–2012) statistical period. The average daily precipitation for the Kasilian watershed was 1.78 (mm), the average daily flow discharge was 2.84 (m^3/s), the minimum and maximum amount of instantaneous flow discharge was 0.05 and was 12.46 (m^3/s), the minimum amount of DSSL was 0.42 and the maximum was 1190.45 (ton/day). The data used for the Rood Zard watershed included 458 information records for a 36-years (1977–2012) statistical period. The average daily precipitation for the Rood Zard watershed was 2.35 (mm), the average daily flow discharge was 10.21 (m^3/s), the minimum and maximum amount of instantaneous flow discharge was 0.22 and 109.00 (m^3/s), the minimum and maximum amount of DSSL was 0.38 and 57300.46 (ton/day). The input data to the models included the instantaneous flow discharge (Q), average daily flow discharge (Q_i), average daily flow discharge for one day ago (Q_{i-1}), average daily flow discharge for two days ago (Q_{i-2}), average daily flow discharge for three days ago (Q_{i-3}), average daily precipitation (P_i), average daily precipitation for one day ago (P_{i-1}), average daily precipitation for two days ago (P_{i-2}), and average daily precipitation for three day ago (P_{i-3}). The output data to the models was the daily suspended sediment load (DSSL).

Table 1. Physiographic characteristics for Kasilian and Rood Zard watersheds.

Physiographic characteristics	Kasilian watershed	Rood Zard watershed
Minimum elevation above sea level (m)	300	400
Maximum elevation above sea level (m)	3000	3300
Surface area (km^2)	336.7	861
Perimeter (km)	114.2	173
Total length of Stream watershed (km)	232.6	453.9
Slop (%)	19.7	27.4
length of Major River (km)	48	47.5
Gravelius factor	1.7	1.7
Form factor	6.8	2.6
Stream condensation	0.7	0.5
Time of Concentration, Kirpich (h)	4	3.8
Bifurcation ratio	2.7	1.6

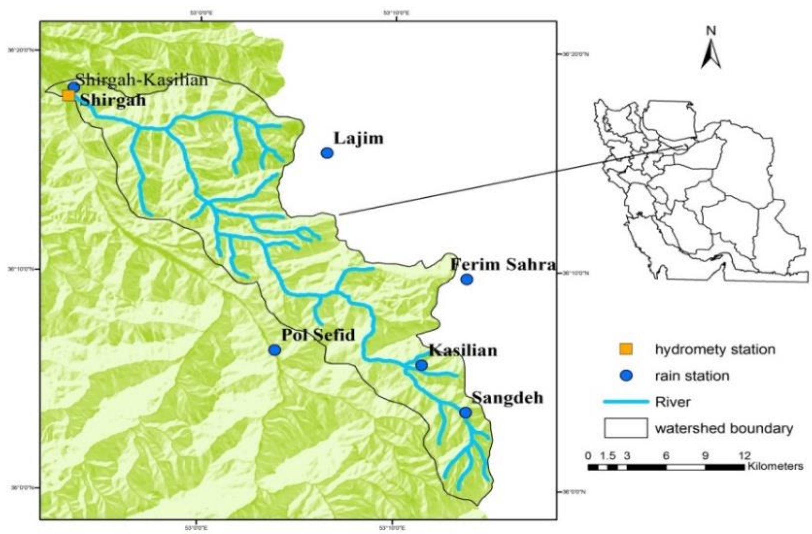

Fig. 1. Kasilian watershed map.

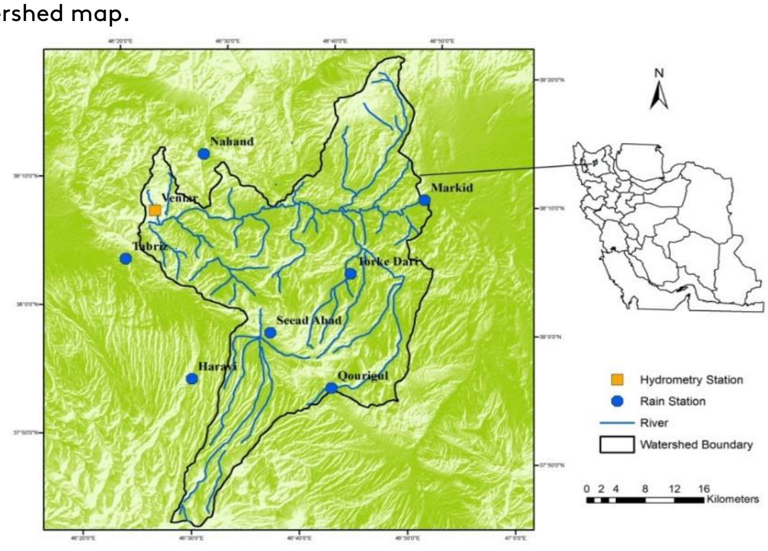

Fig. 2. Rood Zard watershed map.

2.2. Preprocessing data

In this study, pre-processing was performed on the models before using the data in the models. The first step was data clustering; SOM was used for data clustering. This network is one of a variety of unstructured ANNs. The network consists of two layers: the input layer is the location of the input variables, and the output layer includes a network of neurons. In this network, the Euclidean distance (Equation 1), each of the data, is calculated from the center of the output layer neurons; the neuron with the least distance to the variable is selected as the winning neuron (the competition phase). In the next step, the winning neuron stimulates the neighboring neurons to adapt to their new input, which is called the Cooperative phase. Ultimately, the network performs the best matching with the data (Adaptation phase) [18]. The Euclidean distance is calculated from Equation 1:

$$D_j = |x - w_j| = \sum_{i=1}^{N} \left[(x_i - w_{ij})^2\right]^{\frac{1}{2}} \quad (1)$$

where D_j is the distance between the output vector of the input vector, N is the number of vector variables, M is the number of output layer neurons, W_{ij} is the weight of the output neuron, and the sign $|x- w_j|$ represents the distance [19]. MATLAB R2013a software was used to design this model. After the first stage of data preprocessing, the next stage is input variable selection. In order to reduce data dimensions, input data selection, and the best combination of input into models, the gamma test software package was used in the WinGamma^{TM} software. In this method, all variables are entered into the software, and based on the V_{ratio}, the gamma coefficient and standard error values for all possible combinations are calculated; the best combination of variables is obtained based on the lowest values of the statistics. This method is very effective for variables that have nonlinear relationships with each other [20,9]. Another stage of data preprocessing is the standardization of data in order to know the scaling data before the combination of variables is entered into the models, making it possible to compare the data with different measurement criteria. In this study, for data entry into the WinGamma^{TM} software, the standardization of data between [0 1] and the use of activation functions log sigmoid or tangent sigmoid in artificial neural networks, data standardization was performed between [0.1 0.9] and [-0.9 0.9].

$$Z = \frac{(X_i - X_{imin})}{(X_{imax} - X_{imin})} \quad (2)$$

$$Z = 0.1 + 0.8 * \frac{(X_i - X_{imin})}{(X_{imax} - X_{imin})} \quad (3)$$

$$Z = \left(1.8 * \frac{(X_i - X_{imin})}{(X_{imax} - X_{imin})}\right) - 0.9 \quad (4)$$

where Z is a standardized variable, X_i is the initial variable, X_{imin} is the minimum value, and X_{imax} is the maximum value.

2.3. ANNs mode

One of the methods of intelligent data processing is the use of ANN, which is due to the analysis of information in a similar way to the human brain, generalization power, no need for a predetermined mathematical model, and the ability to learn and learn. It can be used to estimat,e latency parameters [21,22]. An artificial neural network is a subset of Calculation intelligence (CI) methods. CI means the extraction of algorithms, mathematical relations, and mappings in numerical data. CI systems, in principle, offer free model dynamic systems for approximating functions and mappings [23]. ANN is modeling the structure of neural computations and synaptic connections similar to the human brain. The Feed-forward Multi-layer Perceptron Neural Networks model was used in this research. This model consists of three layers: input, intermediate (hide), and output. The input layer is a transmitter layer and a device for data acquisition. The final layer or output layers include the values predicted by the network and, therefore, represent the pattern output; the middle and hidden layers composed of processor nodes are the locations of the data processing. The number of hidden layers and the number of nodes in each hidden layer are typically identified by the validation and strain method [8]. The process of training in Feed-forward Multi-layer Perceptron Neural Networks follows the Delta rule or back propagation rule. In order to succeed in network training, the output should be gradually closer to the optimal output to reduce the amount of error function. For this purpose, the weight coefficients of the communication lines of the units are adjusted using the Delta general rule. Delta's rule calculates the value of the error function and releases it from the back of a layer to the previous layer [24]. Delta law is defined by Equation 5:

$$W_{ij}^{new} = W_{ij}^{old} - \eta \frac{\partial E}{\partial W_{ij}} \quad (5)$$

where W_{ij}^{old} and W_{ij}^{new} are respectively the weight between the neurons i and j before and after the specified repetition, η is the learning rate, and E is the error function. The Lewenberg Marquardt method was used for learning the neural network. The activation functions in the hidden layer neurons and the output layer were respectively considered log sigmoid or tangent sigmoid and linear. In this research, MATLAB R2013a software was used for ANN modeling, data clustering and calculating the data cluster validity index.

2.4. GEP model

The GEP model, a form of extended genetic programming, was presented by Ferreira [25]. This model is one of the methods of circular algorithm and one of a variety of intelligent models based on Darwin's theory of evolution. In this method, the algorithms attempt to define a goal function in the form of qualitative criteria and then apply the function to compare different problem-solving solutions in a step-by-step process of data structure, ultimately providing the appropriate answer. The fundamental difference between the genetic algorithm and GEP is the nature of each individual; individuals in the genetic algorithm linear rows are fixed-length (chromosomes), but in the GEP, they are separate branches [25]. GEP is also emphasized on the tree structure of the sets, but the genetic algorithm is based on the system of binary cultivars. The first step in the model algorithm is to generate the initial population of solutions, which can be done by random process or taking input information about the problem. Then, the chromosomes are expressed as tree expressions and evaluated by fit function. Evolution is stopped if the desired solution is reached or generations reach a certain number, and the best solution is presented [25]. If the conditions do not stop, elitism will be performed, and the remaining solutions will be assigned to the selective process. This process is repeated for several generations, and with the advance of generations, the quality of the population is also improved in relative terms [26]. In GEP, various operators, such as mutation and combination, are used. The goal of the mutation operator is to randomly regenerate within certain chromosomes. The function of this operator is that it performs some defective operations to prevent the creation of defective individuals in terms of rules. In this model, one-point, two-point, and gene combinations are used. Since the two-point combination is able to turn the non-coded areas into chromosomes more efficiently, it is more favorable. Another operator used in GEP is transposition. In this method, various phenomena are modeled using a set of arithmetic functions, trigonometry or functions defined by users and terminals. The set of terminals consists of fixed values and independent variables of the problem [27,10]. In this research, GEPXpro Tools5.0 software was used for GEP modeling.

2.5. SRC model

SRC is an exponential regression relation between instantaneous flow (Q) in terms of (m^3/s) and its daily suspended sediment load (DSSL) in terms of (ton day^{-1}). The coefficients a and b are the coefficients of SRC for the hydrometric station, which is calculated by the least squares error method [23].

$$\mathrm{DSSL} = aQ^b \quad (6)$$

2.6. Evaluation criteria

In this research, for the evaluation and performance of the models, quantitative indices including Coefficient of Determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and Nash-Sutcliffe (NS) were used, which are shown in Equations 7 to 11, respectively.

$$R^2 = \left[\frac{\sum_{i=1}^{n}(s_o - \overline{s_0})(s_M - \overline{s_M})}{\text{[illegible]}}\right]^2 \quad (7)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_M - s_O)^2} \quad (8)$$

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}|(s_O - s_M)|}{n} \quad (9)$$

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n}(s_M - s_O)^2}{\sum_{i=1}^{n}(s_o - \overline{s_0})^2} \quad (10)$$

where s_O and s_M respectively are the observed and predicted suspended sediment load, $\overline{s_O}$ is the average of the observed suspended sediment load, $\overline{s_M}$ is the average of predicted suspended sediment load, and n is the number of data. In this research, MATLAB R2013a software and SPSS22 software were used for statistical analysis.

3. Results and Discussion

3.1. Statistical data

The statistical characteristics of the data used in the ANN, GEP, and SRC models for the Kasilian and Rood Zard watersheds are presented in Tables 2 and 3.

Table 2. Statistical data for the Kasilian watershed.

Data set	Minimum	Maximum	Mean	Coefficient of variation
Q (m^3/s)	0.05	12.46	2.84	0.84
Q_i (m^3/s)	0.00	14.40	2.84	0.91
Q_{i-1} (m^3/s)	0.01	19.70	3.01	0.99
Q_{i-2} (m^3/s)	0.01	22.00	3.02	1.05
Q_{i-3} (m^3/s)	0.00	62.45	3.08	1.28
P_i (mm)	0.00	38.94	1.78	2.08
P_{i-1} (mm)	0.00	60.26	2.23	2.37
P_{i-2} (mm)	0.00	39.75	2.31	2.05
P_{i-3} (mm)	0.00	47.98	2.45	2.07
DSSL (ton/day)	0.42	1190.45	70.08	2.19

Table 3. Statistical data for the Rood Zard watershed.

Data set	Minimum	Maximum	Mean	Coefficient of variation
Q (m^3/s)	0.22	109.00	9.39	1.50
Q_i (m^3/s)	0.20	174.00	10.21	1.89
Q_{i-1} (m^3/s)	0.20	370.00	11.73	2.55
Q_{i-2} (m^3/s)	0.15	533.00	9.98	2.84
Q_{i-3} (m^3/s)	0.15	443.00	9.74	2.60
P_i (mm)	0.00	55.50	2.35	3.31
P_{i-1} (mm)	0.00	91.23	2.85	3.45
P_{i-2} (mm)	0.00	72.85	1.92	3.46
P_{i-3} (mm)	0.00	51.29	1.57	3.59
DSSL (ton/day)	0.38	57300.46	1218.43	4.86

3.2. Results of preprocessing of data

In this research, data clustering was performed in three groups of 70% training data (345 data for the Kasilian watershed and 320 for the Rood Zard watershed), 15% cross-validation data (73 data for the Kasilian watershed and 69 for the Rood Zard watershed), and 15% validation data (73 data for the Kasilian watershed and 69 for the Rood Zard watershed) using the SOM method. In this method, the optimal number of clusters was first obtained using Davis Bouldin index (DBI), and then the data were divided into three groups. Figures 3 and 4 show the DBI charts for the Kasilian and Rood Zard watersheds, respectively. The optimal cluster number was obtained when the DBI was the least. According to Figures. 3 and 4, the number of optimal clusters of the Kasilian watershed was 46 with DBI= 1.01 and the optimal number of clusters of the Rood Zard watershed was 27 with DBI=0.85. Figures 6 and 7 show the minimum, maximum, mean, and coefficient of variation (CV) charts in three groups of training data, cross-validation data, and validation data for the Kasilian and Rood Zard watersheds, respectively.

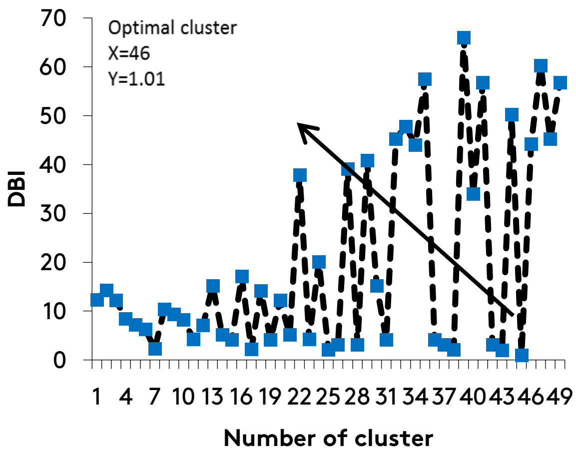

Fig. 3. DB index for the Kasilian watershed.

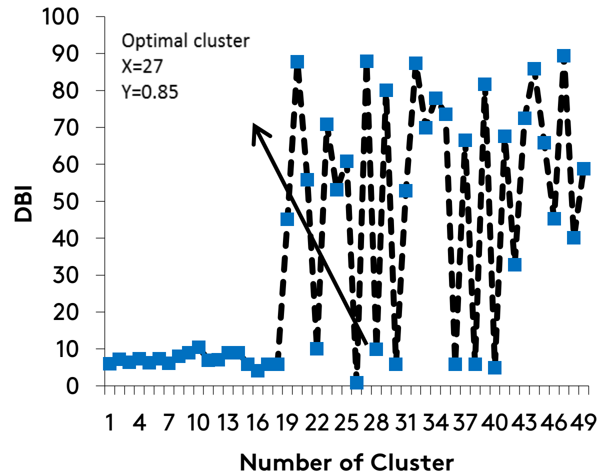

Fig. 4. DB index for the Rood Zard watershed.

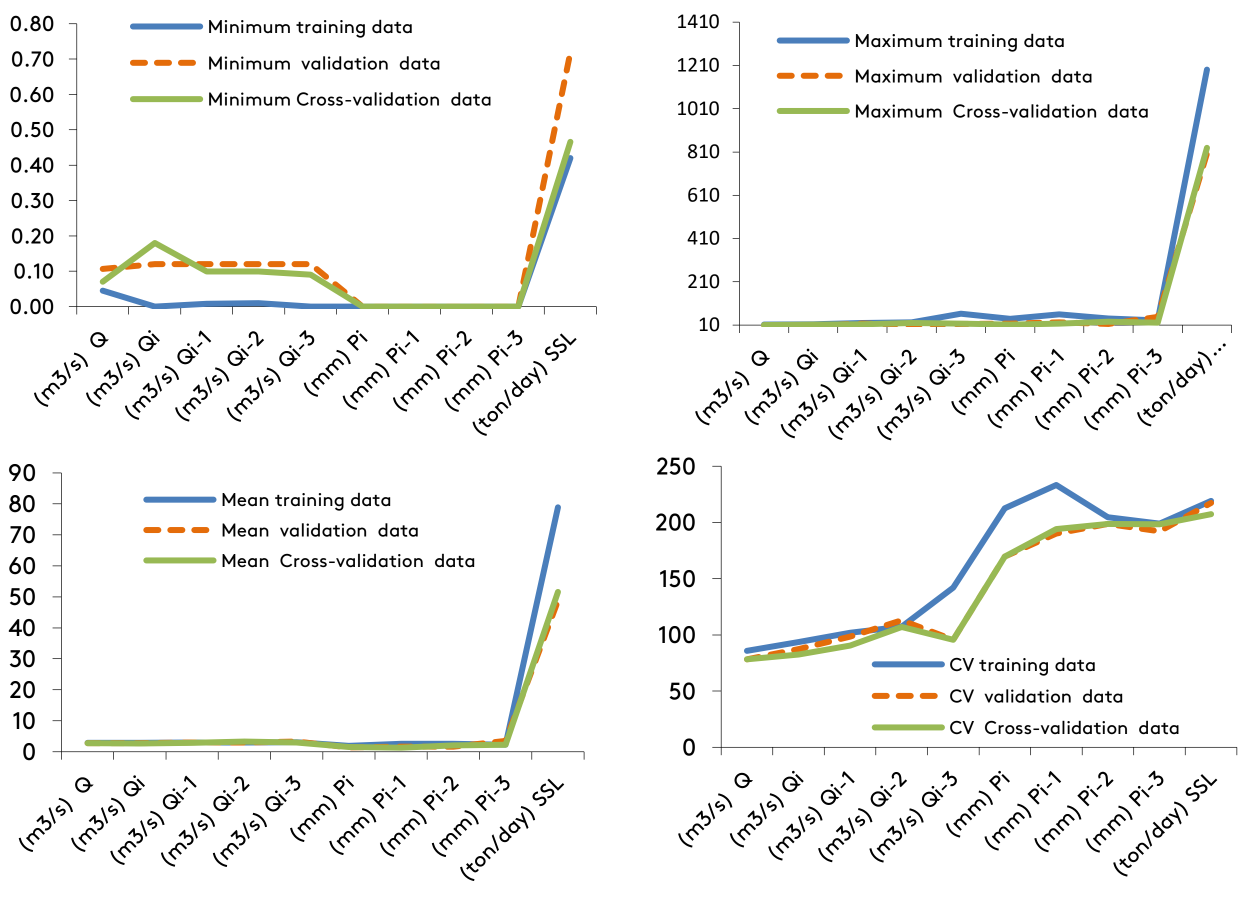

Fig. 5. Statistical data charts for training, cross-validation, and validation data for the Kasilian watershed.

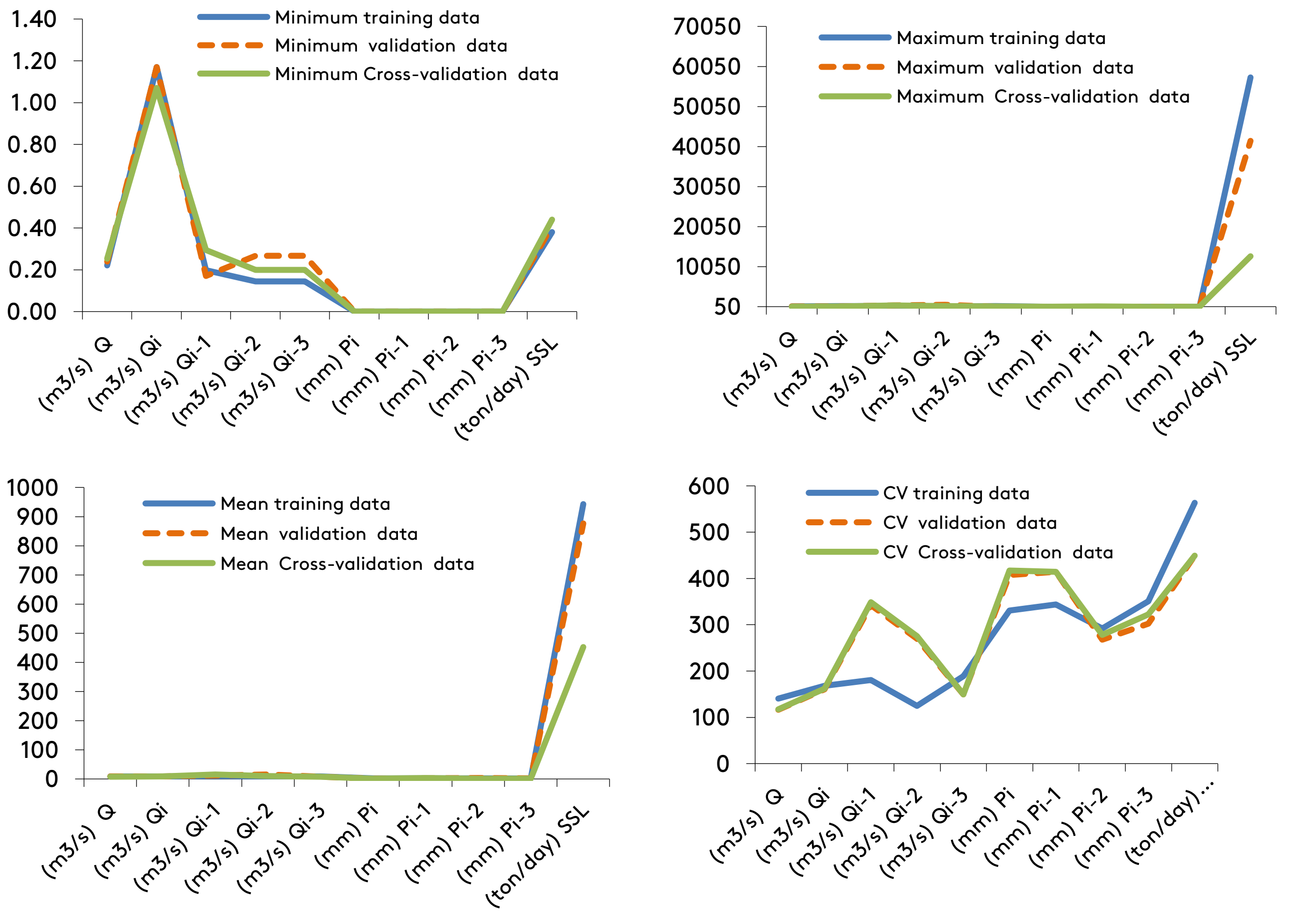

Fig. 6. Statistical data charts for training, cross-validation, and validation data for the Rood Zard watershed.

These charts show that the distribution of data in all three groups was uniform. That is, the data in each training, cross-validation, and validation group could be representative of the total data during the statistical period. In fact, the use of the SOM method for clustering and dividing data makes the data even more homogeneous. This will increase the power of generalization and performance of the models, and the accuracy of DSSL estimates increased; the results of this study corresponded with the research by Chaudhary et al. [19] regarding the role of the SOM method in improving the results of the models. The results of this study were consistent with the results of Tabatabaei and Salehpour Jam [28]. These researchers calibrated the SRC model using an evolutionary algorithm for DSSL estimation in the Shalman Rood watershed in the north of Iran. They clustered the data using the SOM neural network to increase the power of generalization and accuracy of the model. The results showed the effect of this method on reducing the amount of RMSE from 5754 to 1681 (ton day^{-1}). The best combination of input variables to the intelligent models in the Kasilian watershed was the gamma test and genetic algorithm methods in WinGamma^{TM} software. These variables included $Q_{i-1}, Q_{i-2}, Q_{i-3}, P_i, P_{i-1}, P_{i-2}, P_{i-3}$, with the least amount of gamma statistic equal to 0.0015, the standard error equal to 0.0008, and V_{ratio} equal to 0.0227; the best combination of input variables for the Rood Zard watershed included the $Q_i, Q_{i-1}, Q_{i-2}, P_i, P_{i-1}, P_{i-2}, P_{i-3}$ variables, with the least amount of gamma statistic equal to 9.2×10^{-5}, the standard error equal to zero, and V_{ratio} equal to 0.0126. Therefore, it was observed that the gamma test and genetic algorithm methods were able to save time and cost in developing the best model. This method is a suitable and impartial technique to assess the potential of each input to the models, providing the best combination of input variables to the ANN and GEP models. Other researchers have also used this technique [29,9].

3.3. Results of modeling

In this research, the best combination of input variables obtained from the gamma test and genetic algorithm was used to enter GEP and ANN models in the Kasilian and Rood Zard watersheds. The values of the parameters used in the GEP model are shown in Table 4. The chromosomes in the GEP model consisted of more than one gene of the same length. The number of genes was optional. In this research, the number of genes for an optimal response in the GEP model was three, and the number of chromosomes was 30. Also, this model was evaluated with the RMSE error function criterion. The mathematical operators used in this model included (+, - , /, ×, Exp, Ln, Inv, x2 and 3RT). The results of the ANN with two activation functions, namely the log sigmoid and tangent sigmoid, and the GEP and SRC models for the Kasilian and Rood Zard watersheds are presented in Tables 5 and 6, respectively.

Table 4. The values of the parameters used for GEP modeling.

parameter	Value	Parameter	Value
Number of chromosomes	30	One-point recombination rate	0.3
Number of Genes	3	Two-point recombination rate	0.3
Linking function	+	Gene recombination rate	0.1
Mutation rate	0.044	Number of head	10-18
Inversion rate	0.1	Gene transposition rate	0.1

Table 5. The results of the models for the Kasilian watershed.

Number of model	Model	Input variables combination	MAE	RMSE (ton day^{-1})	NSE	R^2
1	ANN-Log sigmoid		250.6	591.4	0.96	0.97
2	ANN-tangent sigmoid	$Q_{i-1},Q_{i-2},Q_{i-3},P_i,P_{i-1},P_{i-2},P_{i-3}$	231.4	578.6	0.98	0.98
3	GEP		252.1	611.2	0.96	0.96
4	SRC	Q	964.7	2420.9	0.08	0.34

Table 6. The results of the models for the Rood Zard watershed.

Number of models	Model	Input variables combination	MAE	RMSE (ton day^{-1})	NSE	R^2
1	ANN-Log sigmoid		564.8	1771.9	0.98	0.98
2	ANN-tangent sigmoid	$Q_i,Q_{i-1},Q_{i-2},P_i,P_{i-1},P_{i-2},P_{i-3}$	781.9	2276.1	0.94	0.95
3	GEP		475.7	1671.9	0.99	0.99
4	SRC	Q	3520.6	9982.4	-0.11	0.01

Figures 6 and 7 respectively show the scatter plot of the results of the predicted DSSL by ANN with the activation function of log sigmoid and tangent sigmoid and the GEP and SRC models versus the observed values for test data set for the Kasilian and Rood Zard watersheds. According to Table 5 and Figure 6, between the intelligent models (ANN and GEP) with best combination of input variables ($Q_{i-1},Q_{i-2},Q_{i-3},P_i,P_{i-1},P_{i-2},P_{i-3}$) and also the traditional SRC model with input variable Q, the ANN model with the activation function of tangent sigmoid with the statistical values MAE=231.4 (ton day^{-1}), RMSE=578.6 (ton day^{-1}), NSE =0.98 and R^2=0.98 was able to estimate DSSL with high accuracy in the Kasilian watershed. Regarding the regression equation of this model (y=0.9677x+5.8462), the line slope was close to one, which indicated the high power of this model in estimating the DSSL value. In this formula, by setting the observed DSSL value instead of X, the predicted DSSL could be obtained with high precision. Also, the proper distribution of data around the regression line in this figure indicates the proper data clustering so that it covers all the DSSL values from low to high. These results were consistent with the results of Kumar et al. [30] and the Samantaray and Ghose [31] research. Kumar et al. [30] used ANN and regression models to estimate DSSL in the Kopili watershed in India. Their results showed that the ANN model compared with the regression model, with NSE=0.89 and R^2=0.92, estimated the DSSL value with high accuracy. Samantaray and Ghose

[31] estimated the value of DSSL using artificial neural network models in the rivers of Salebhata, India. Their results showed that the Feed-forward Multi-layer Perceptron (FFMLP) of the ANN model had an accurate estimate of DSSL with RMSE=0.00873 and R^2=0.93. Also, the best model for estimating DSSL in the Rood Zard watershed was the GEP model, with the input variables Q_i,Q_{i-1},Q_{i-2},P_i,P_{i-1},P_{i-2},P_{i-3}. The statistical values of this model were MAE=475.7 (ton day^{-1}), RMSE=1671.9 (ton day^{-1}), NSE=0.99, and R^2=0.99 (Table 6 and Figure 7). Azamathulla et al. [32], Sheikhalipour and Hassanpour [14], and Emamgholizadeh and Karimi Demneh [33] obtained results similar to those of the current research. Azamathulla et al. [32] used Adaptive neuro fuzzy inference system (ANFIS), GEP, and regression methods to estimate DSSL in the Muda, Langat, and Kurau rivers in Malaysia. The results showed that the ANFIS and GEP models performed DSSL estimation with high accuracy compared to the regression model. Sheikhalipour and Hassanpour [14] estimated the suspended sediment load in the Sistan River using the GEP model. Results for test data with RMSE=2305.45 (ton day^{-1}), MBE=1400.12 (ton day^{-1}), and R^2=0.88 showed that DSSL was estimated using this method with high accuracy. Emamgholizadeh and Karimi Demneh [33] compared three intelligence models (GEP, ANN, and ANFIS models) with the SRC method to estimate the daily suspended sediment load in the Kasilian and Talar Rivers. The results showed that the GEP model with a high coefficient of determination (R^2) and a low mean absolute error (MAE) was better than both the ANN and ANFIS models for estimating the daily suspended sediment load of the two sub-basins. The results of the models in the Kasilian and Rood Zard watersheds showed that the intelligent models had higher power than the traditional SRC model in DSSL estimation. The SRC model in the Kasilian watershed with R^2=0.34 and NSE=0.08 and the Rood Zard watershed with R^2=0.01 and NSE=-0.11 showed the inability of this model to estimate DSSL. Since the SRC method is based on regression methods and definitive variables, and because in nature the relationship between variables is vague and nonlinear, the SRC model was not able to estimate the correct DSSL. Figures 8 and 9 show the results of the best model derived from the DSSL estimate versus the observation values for the Kasilian and Rood Zard watersheds, respectively. These graphs showed that the observed and predicted data were well-matched. Therefore, it was concluded that intelligent models such as artificial neural networks and GEP were able to estimate DSSL value in watersheds with high accuracy and efficiency.

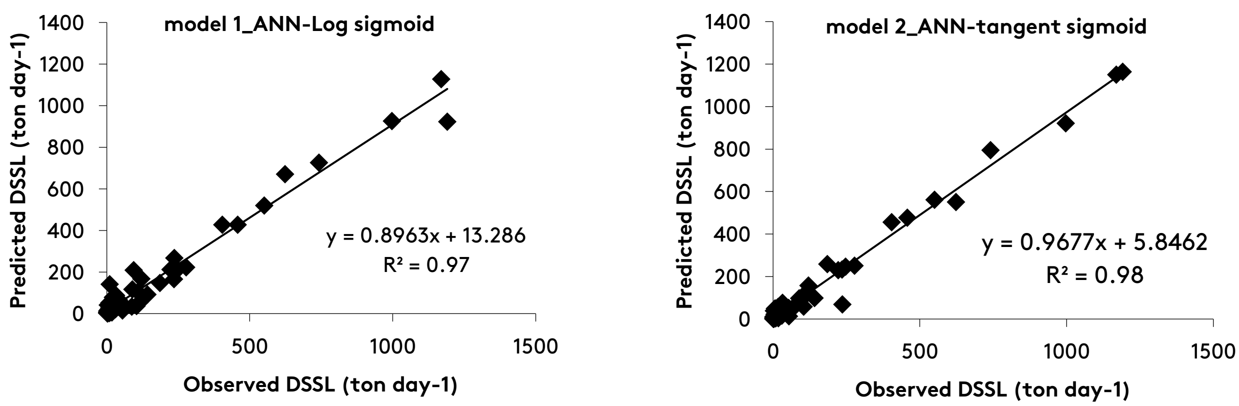

Fig. 6. Scatter plot of predicted versus observed DSSL by models for the Kasilian watershed.

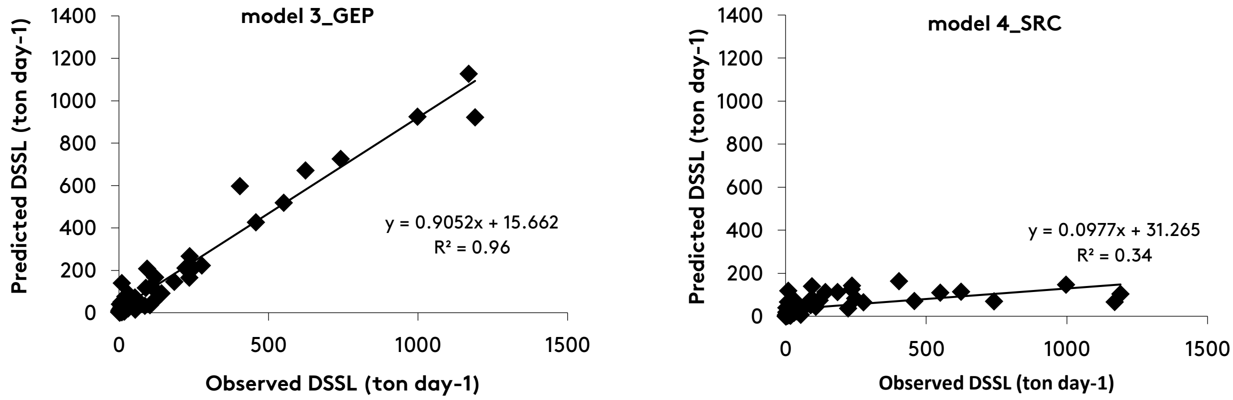

Fig. 6. (Continued) Scatter plot of predicted versus observed DSSL by models for the Kasilian watershed.

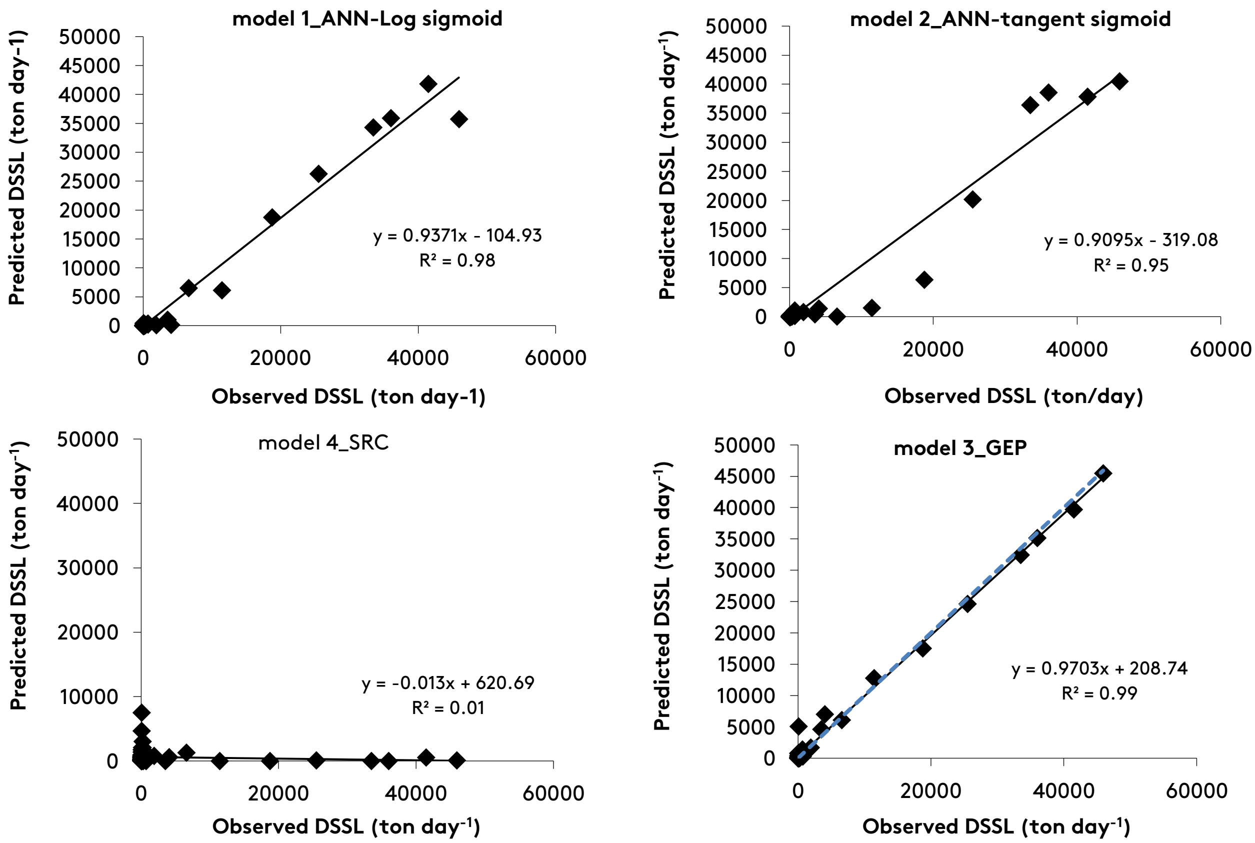

Fig. 7. Scatter plot of predicted versus observed DSSL by models for the Rood Zard watershed.

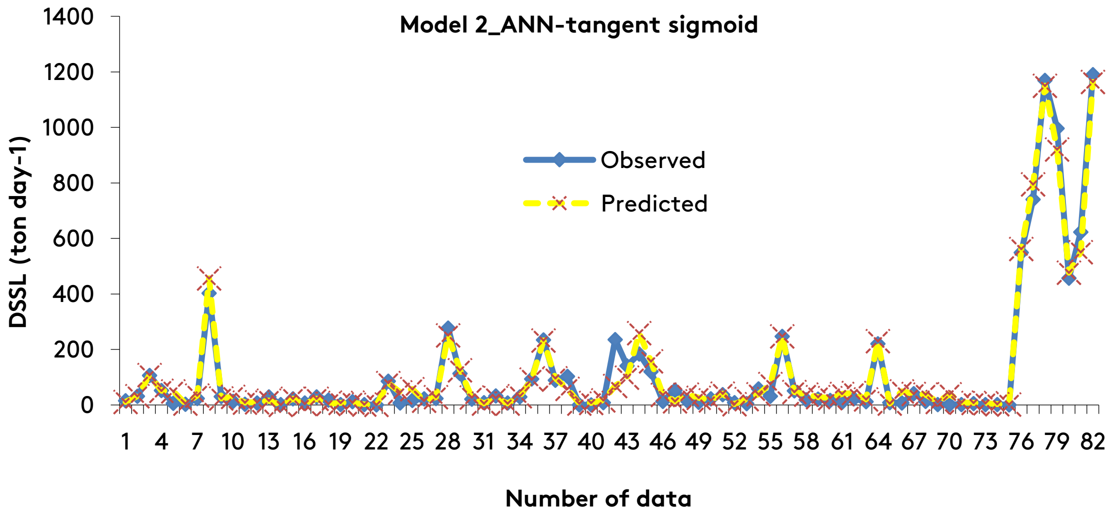

Fig. 8. Graph of the results of the best model for the predicted DSSL versus observational values for the Kasilian watershed.

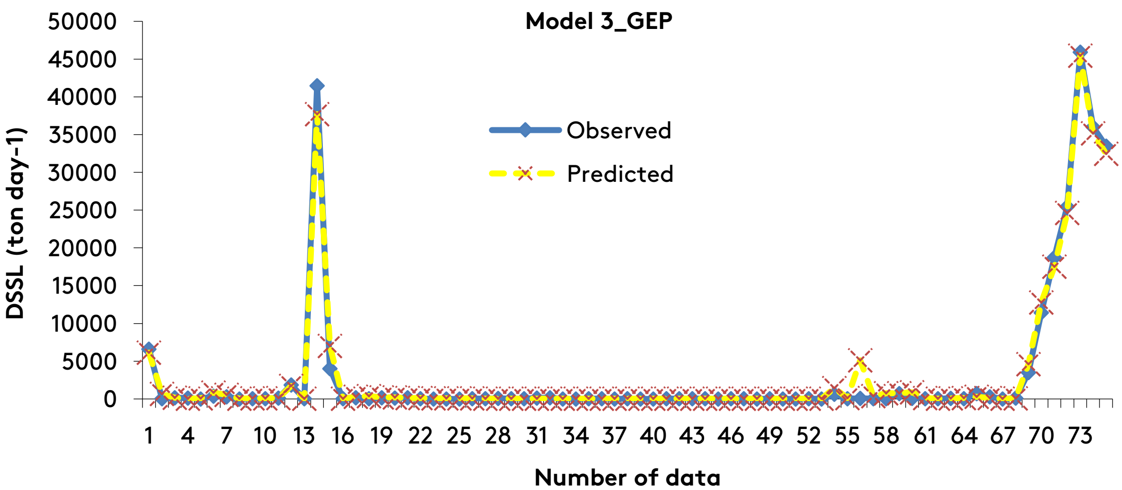

Fig. 9. Graph of the results of the best model for the predicted DSSL versus observational values for the Rood Zard watershed.

4. Conclusions

In this research, the ANN and GEP intelligent models and the traditional SRC model were used to estimate the suspended sediment load in the Kasilian watershed, with a humid climate in the north of Iran, and the Rood Zard watershed, with a semi-arid climate in the southwest of Iran. The SOM method was used for data clustering, and the data was divided into three groups: 70% training data, 15% cross-validation data, and 15% validation data. Contrary to the random methods in the data divide, the data in each group in the SOM method are representative of the total data during the studied statistical period; this will increase the modeling power in the correct DSSL estimation. Also, the gamma test and genetic algorithm were used to reduce the dimensions of input data into models to increase the speed of the algorithms and obtain the best combination of input variables in the watersheds to save time and cost. Modeling using this method reduced the likelihood of over fitting and, therefore, the generalization power for learning algorithms increased. The results showed that in the Kasilian watershed, the ANN model with the activation function of the tangent sigmoid and input variables $Q_{i-1},Q_{i-2},Q_{i-3},P_i,P_{i-1},P_{i-2},P_{i-3}$ and in the Rood Zard watershed, the GEP model with input variables $Q_i,Q_{i-1},Q_{i-2},P_i,P_{i-1},P_{i-2},P_{i-3}$, were the best models in estimating DSSL. The results of the models in the watersheds showed the superiority of intelligent models compared to the SRC model in DSSL estimation. In intelligent models, the relationships between input and output variables, regardless of the explicit physical laws between them, were detected. As a result, intelligent models increased the ability to estimate the DSSL of the watersheds in conditions where there is always uncertainty in understanding the problem and the responses between the watershed resources with the dynamic variables. Therefore, considering the accuracy of smart methods in data estimation, it is suggested to use SOM models, gamma tests, genetic algorithms, ANN and GEP models in other studies such as watershed erosion and soil aggregate stability estimation.

References

[1] Emamgholizadeh, S., Bateni, S. M., Nielson, J. R. (2018). Evaluation of different strategies for management of reservoir sedimentation in semi-arid regions: a case study (Dez Reservoir). *Lake and reservoir management*, *34*(3), 270-282. https://doi.org/10.1080/10402381.2018.1436624

[2] Chamoun, S., De Cesare, G., Schleiss, A. J. (2017). Venting of turbidity currents approaching a rectangular opening on a horizontal bed. *Journal of hydraulic research,* 1–15. https://doi.org/10.1080/00221686.2017.1289266

[3] Tayfur, G. (2014). Soft computing in water resources engineering: Artificial neural networks, fuzzy logic and genetic algorithms. WIT Press.

[4] Harrington, S. T., Harrington, J. R. (2013). An assessment of the suspended sediment rating curve approach for load estimation on the Rivers Bandon and Owenabue, Ireland. *Geomorphology, 185*, 27–38. https://doi.org/10.1016/j.geomorph.2012.12.002

[5] Shiau, J. T., Chen, T. J. (2015). Quintile regression-based probabilistic estimation scheme for daily and annual suspended sediment loads. *Water resources management,* 29(8), 2805–2818. https://doi.org/*10.1007/s11269-015-0971-5*

[6] Heng, S., Suetsugi, T. (2014). Comparison of regionalization approaches in parameterizing sediment rating curve in ungauged catchments for subsequent instantaneous sediment yield prediction. *Journal of hydrology, 512*, 240–253. https://doi.org/10.1016/j.jhydrol.2014.03.003

[7] Li, X., Nour, M. H., Smith, D. W., Prepasc, A. A. (2010). Neural networks modeling of nitrogen export: model development and application to unmonitored boreal forest watersheds. *Environmental technology*, *31*(5), 495–510. https://doi.org/10.1080/09593330903527880

[8] Chen, X. Y., Chau, K. W. (2016). A hybrid double feed forward Neural Network for suspended sediment load estimation. *Water resources management*, *30*, 2179–2194. https://doi.org/10.1007/s11269-016-1281-2.

[9] Shamim, M. A., Hassan, M., Ahmad, S., Zeeshan, M. (2016). A comparison of Artificial Neural Networks (ANN) and Local Linear Regression (LLR) techniques for predicting monthly reservoir levels. *KSCE journal of civil engineering, 20*(2), 971–977. https://doi.org/10.1007/s12205-015-0298-z

[10] Thompsona, J., Sattar, A., Gharabaghi, B., Warner, R. (2016). Event-based total suspended sediment particle size distribution model. *Journal of hydrology, 536*, 236-246. https://doi.org/10.1016/j.jhydrol.2016.02.056

[11] Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M., Nourani, V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of the total environment*, *407*(17), 4916-4927. https://doi.org/10.1016/j.scitotenv.2009.05.016

[12] Kisi, O., Aytac, G. (2010). A machine code-based genetic programming for suspended sediment concentration estimation. *Advances in engineering software, 41*(8), 939-945. https://doi.org/10.1016/j.advengsoft.2010.06.001

[13] Boukhrissa, Z. A., Khanchoul, K., Bissonnais, Y. L., Tourki, M. (2013). Prediction of sediment load by sediment rating curve and neural networ.k (ANN) in El Kebir catchment, Algeria. *Journal of earth system science, 122*(5), 1303–1312.

https://doi.org/10.1007/s12040-013-0347-2

[14] Sheikhalipour, Z., Hassanpour, F. (2013). Estimation of suspended sediment load using genetic expression programming. *Journal of civil engineering and urbanism, 3*(2), 292-299.

[15] Abbaspour, B., Haghiabi, A. H. (2015). Comparing the estimation of suspended load using two methods of sediments rating curve and artificial neural network (A Case Study: Cham Anjir Station, Lorestan Province). *Journal of environmental treatment techniques*, *3*(4), 215-222.

[16] Joshi, R., Kumar, K., Adhikari, P. S. (2016). Modelling suspended sediment concentration using artificial neural networks for Gangotri glacier. *Hydrology process, 30*(2), 1354–1366.

[17] Nivesh, Sh., Kumar, P. (2017). Modelling River suspended sediment load using artificial neural network and multiple linear regression: Vamsadhara River Basin, India. *International journal of chemical studies, 5*(5), 337-344. https://doi.org/10.1002/hyp.10723

[18] Chen, I. T., Chang, L. C., Chang, F. J. (2017). Exploring the spatio-temporal interrelation between groundwater and surface water by using the Self-Organizing Maps. *Journal of hydrology*, *556*, 131-142. https://doi.org/10.1016/j.jhydrol.2017.10.015

[19] Chaudhary, V., Bhatia, R. S., Ahlawat, A. (2014). The self-organizing map learning algorithm with inactive and relative winning frequency of active neurons. *Journal HKIE transactions, 21*(1), 62–67. http://dx.doi.org/10.1080/1023697X.2014.883680

[20] Kakaei Lafdani, E., Moghaddam Nia, A., Ahmadi, A. (2013). Daily suspended sediment load prediction using artificial neural networks and support vector machines. *Journal of hydrology, 478*(4), 25-50. https://doi.org/10.1016/j.jhydrol.2012.11.048

[21] Emamgholizadeh, S. (2012). Neural network modeling of scour cone geometry around outlet in the pressure flushing. *Global NEST journal. 14*(4), 540-549. https://doi.org/10.30955/gnj.000765

[22] Emamgholizadeh S, Bateni, M. F.M. Jeng D. S. (2013). Artificial intelligence-based estimation of flushing half-cone geometry. *Engineering applications of artificial intelligence*, *26*(10), 2551-2558. *https://doi.org/10.1016/j.engappai.2013.05.014*

[23] Tfwala, S. S., Wang, Y. M. (2016). Estimating sediment discharge using Sediment Rating Curves and Artificial Neural Networks in the Shiwen River, Taiwan. *Water, 8*(53), 1-15. https://doi.org/10.3390/w8020053

[24] Zounemat-Kermani, M., Kişi, O., Adamowski, J., Ramezani-Charmahineh, A. (2016). Evaluation of data driven models for river suspended sediment concentration modeling. *Journal of hydrology,* 16(2), 1-40. https://doi.org/10.1016/j.jhydrol.2016.02.012

[25] Ferreira, C. (2006). *Gene expression programming: mathematical modeling by an artificial intelligence* (Vol. 21). Springer.

[26] Shiri, J., Kisi, O. (2011). Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Computers and geosciences, 37*(10), 1692-1701. https://doi.org/10.1016/j.cageo.2010.11.010

[27] Emamgholizadeh, S. M., Bateni, S.M., Shahsavani, D., Ashrafi, T. Ghorbani, H. (2015). Estimation of soil cation exchange capacity using Genetic Expression Programming (GEP) and Multivariate Adaptive Regression Splines (MARS). *Journal of hydrology*, *529*, 1590–1600. *https://doi.org/10.1016/j.jhydrol.2015.08.025*

[28] Tabatabaei, M., Salehpour Jam, A. (2017). Optimization of sediment rating curve coefficients using evolutionary algorithms and unsupervised artificial neural network. *Caspian journal of environmental sciences, 15*(4), 387-401. https://doi.org/10.22124/CJES.2017.2665

[29] Rashidi, S., Vafakhah, M., Kakaei Lafdani, E., Javadi, M. R. (2016). Evaluating the support vector machine for suspended sediment load forecasting based on gamma test. *Arabian journal of geosciences, 9*, 1-15. http://dx.doi.org/10.1007/s12517-016-2601-9

[30] Kumar, D., Pandey, A., Sharma, N., Flugel, W. A. (2016). Daily suspended sediment simulation using machine learning approach. *CATENA, 138*, 77–90. https://doi.org/10.1016/j.catena.2015.11.013

[31] Samantaray, S., Ghose, D. K. (2018). Evaluation of suspended sediment concentration using descent neural networks. *Procedia computer science*, *132*, 1824–1831. https://doi.org/10.1016/j.procs.2018.05.138

[32] Azamathulla, H. M., Cuan, Y. C., Ghani, A. b., Chang, C. K. (2013). Suspended sediment load prediction of river systems:GEP approach. *Arabian journal of geosciences, 6*, 3469–3480. https://doi.org/10.1007/s12517-012-0608-4

[33] Emamgholizadeh, S. Karimi Demneh, R. (2019). A comparison of artificial intelligence models for the estimation of daily suspended sediment load: a case study on the Telar and Kasilian rivers in Iran. *Water supply, 19*(1), 165-178. *https://doi.org/10.2166/ws.2018.062*

How to site this paper:

Alijanpour Shalmani, A., vaezi, A., & Tabatabaei, M. R. (2024). Estimating daily suspended sediment by intelligent and traditional models (Case Study: Kasalian and Rood Zard watersheds, Iran). *Advances in Environmental Technology*,10(2), 102-117, doi: 10.22104/aet.2024.4846.1309