Suspended Load Modeling of River Using Soft Computing Techniques

Amir Moradinejad[1]

Abstract

The phenomenon of sediment transport has always affected many river and civil structures. Not knowing the exact amount of sediment, causes much damage. Correct estimation of river sediment concentration is essential for planning and managing water resources projects and environmental issues. For this, you can use the artificial intelligence method, which has high flexibility. In this research, adaptive neuro-fuzzy models (ANFIS), gene expression programming (GEP), support vector regression (SVR), Group Method of Data Handling (GMDH), and the classical method of sediment rating curve (SRC) were used to model and prediction. For this purpose, the daily data of temperature, rainfall, sediment, and discharge of the Jalair station located in the Markazi province of Iran were used. The results obtained from these five methods were compared with each other and with the measured data. To evaluate the methods used, correlation coefficient, root mean square error, mean absolute error, and Taylor diagram were used. The results show the acceptable performance of data mining methods compared to the Sediment rating curve. Also, the model's superiority (GEP) was shown with the highest coefficient of determination R^2 with a value of 0.98 and the lowest root mean square error RMSE in terms of tons per day with a value of 3721. The efficiency of the ANFIS and GMDH model with R^2 values of 0.93, 0.98, and RMSE values of 16556, and 18638 was somewhat better than the SVR model with an R^2 value of 0.90 and RMSE value of 35158.

1 Introduction

The sediment transport phenomenon is one of the hydrodynamic processes that affect many river structures and civil facilities. A river is a dynamic system controlled by hydraulic processes and sediment transport. At the same time, rivers change their status by changing channel sections, increasing or decreasing sediment transport, erosion and

✉ Amir Moradinejad
amir_24619@yahoo.com

1   Soil Conservation and Watershed Management Research Department, Markazi Agricultural and Natural Resources Research and Education Center, Arak, Agricultural Research Education & Extention Organization (AREEO), Arak, Iran

sedimentation along their channels; all of which affect sustainability, the strength and power of banks and the shape of rivers. Delayed load analysis is widely used in reservoir design, sediment transport, water pollution, access control, and watershed management (Melesse et al. 2011). Removal of excess sediment causes a decrease in section water, a change in water flow in the plan, and a decrease in water in the water (Adnan et al. 2022). Therefore, it is essential to estimate the amount of sediment removed from various factors accurately. Experts have prepared many relationships to predict water suspension transport, however, it is generally accepted that in most cases, due to the complexity of the sediment transport process, the results obtained from experimental relationships are inconsistent with reality (Azamathulla et al. 2013). Almost all equations that predict sediment loads are derived for extreme conditions of steady flow and sediment transport equilibrium, which rarely occur in rivers. Therefore, the output prediction is not accurate according to the basic assumptions above (Hesavi and Shafaee-bajestan 2010). Estimating actual sediment load is complicated by its relationship to flow, nonlinear mechanisms, and complex interactions between events (Kisi et al. 2012). To simulate such situations, you can use data-driven methods, which are physical process methods and do not require relationships that control the outcome (Ivakhnenko 1968). Intelligent models are essential tools for understanding abnormal behavior of conditions, so researchers use these models to predict displacement loads and deposits in rivers. It is non-linear and complex; therefore, calculating sediment concentration is important in water management (Rahul et al. 2021). Alp and Cigizoglu (2005) used the generalized neural network (GRNN) model to predict daily loads in the Janiata basin in the United States and compared the performance of the model with the regression model. Their results show that the GRNN model is more accurate in predicting delays. Kisi et al. (2008) examined the performance of neurofuzzy adaptive inference in predicting monthly sediment removal in the Keulos and Salurkoprosu rivers in Turkey and compared it with artificial neural network (ANN) model and sediment rating curves. Research results show that the neuro-fuzzy adaptive inference system method performs better in predicting the load delay rate than other methods. These researchers concluded that neural networks outperformed statistical models of sediment transport. Kisi and Shiri (2012) evaluated the accuracy of a genetic process (GP) model, a neural network, a modified neuro-fuzzy inference system, and a support vector machine in predicting the daily load of two aquatic centers in the United States. The GP model gives better results than other models. Many researchers such as Aytek and Kisi (2007), Kisi (2005), Ghani et al. (2010), and Fırat and Gungor (2009) on artificial neural networks, fuzzy logic, algorithmic genetics (GA), adaptive fuzzy neural inference system (ANFIS), error backpropagation algorithm (FFBP), feed-forward algorithm (FFNN) and … It is used to simulate suspended load, bedload, and determine the relationship between flow and sediment load. It is already used in rivers. Rajai et al. (2009), based on flow and head concentration data of Liqvan-Chai station, the use of neural networks and neuro-fuzzy networks for sediment concentration prediction was investigated. Historical discharge data were used to predict future suspended sediment concentrations. The results of this study show the superiority of the neurophysical model over other models. Compared with the classical canonical curve model, the smart neural and neural phase models perform very well. Eder et al. (2010) considered the residual effect and found that the sediment rating curve method was less than the above methods. Onderka et al. (2012) used different models to predict suspended loads and found that the M5 tree model had a good ability to predict sediment loads. Duan et al. (2015) used the SPARROW method to model sediment removal and their transport in the Shiluo basin, analyzed the processes of sediment formation and transport in rivers, and reported that this method can be used to manage water resources. Nikpour and Sanikhani

(2017) modeled and predicted the rock removal rate in the Darhroud basin in Ardabil Province through adaptive neuro-fuzzy modeling (ANFIS), gene expression programming (GEP), and support vector regression (SVR). The results show that the performance of the model is good, the SVR model has the highest coefficient of determination ($R^2 = 0.97$), the lowest root mean square error (RMSE = 17343 tons/day), and the best of the Nash–Sutcliffe (NS = 0.97) and Wilmot index (WI = 0.98) is in the verification phase. Kisi and Ozkan (2017) Keulos and Salurkoprosu, Comparison of the performance of a neuro-fuzzy adaptive inference system with a network model in predicting the monthly removal of sediments from two rivers in Turkey. They learn about neural networks and learning sediment rating curves. This study shows that the neuro-fuzzy adaptive inference system performs better than the other two methods in estimating the freight rate in rivers. Water sediment load is one of the most important factors to control hydraulic functions and predict soil and water erosion (Asadi and Fathzadeh 2017). Moradinejad et al. (2018) investigated the effectiveness of the sediment load estimation method in the Qara Çhai River. They use methods of neural network design, neural networks, econometric sedimentation curves, and multiple regression models. The results show that the neurophysical method based on flow and input material and the neural network model based on input material are more accurate than the multivariate and sediment rating curves. Nourani et al. (2020) estimated the solid waste removal rate in the Qatar River Basin in Ethiopia using an integrated artificial intelligence model. This study uses three smart models, ANFIS, SVM, and FFNN, and multilinear MLR with regression to model sediment load (SSL), and uses the coefficient of determination and square root error to evaluate the model's effectiveness. Doroudi et al. (2021) predicted the daily suspended load in the Cham-Siyah river basin in Kekhiloyeh and Boyer Ahmad (Iran) provinces using the joint support vector regression model and learning-learning analysis and evaluated the (teacher's) performance. In this study, since the uncertainty of SVR is unknown, a new hybrid model is proposed by combining the observation-based optimization method with the SVR model. The prediction model uses different measurement parameters. The results show that the SVR-OTLBO model performs better than other models. Beiranvand et al. (2023) used machine learning algorithms to examine the effectiveness of GP-RBF, GP-PUK, RepTree, RF, and M5P models in simulating water delayed loading during low water periods. They also paid too much for water in the Kashkan basin. The results showed that the two-core GP model, PUK and RBF, performed better than other models (RF, RepTree, M5P) during periods of water scarcity and abundance, according to the comparisons. Moreover, according to the test results, the GP-PUK model has the best results. Keshtegar et al. (2023) estimated water removal of sediments in Pakistan using a soft method. These researchers compared the results of the RM5Tree model with support vector regression (SVR), artificial neural network (ANN), multivariate adaptive regression spline (MARS), sediment scale curve (SRC), and response models (RSM). They use parameters such as water flow, snow cover, air temperature, evaporation and transpiration, and precipitation quality. Model accuracy was assessed using Pearson; correlation coefficient (R^2), root mean square error (RMSE) and mean percentage error (MAPE). Research results show that the RM5Tree model has a relative gain of 4.10% compared to the MARS, ANN, SVR, M5Tree, RSM and SRC models 80.62, 77.86, 81.90, 80.20, 74.58 and 62.49 respectively. In one study, Piraei et al. (2023) evaluated the total sediment load in rivers using XGBoost. The results show that XGBoost outperforms other methods when considering six performance metrics. In particular, the root mean square error and coefficient of determination are 216 and 0.95, respectively, while the ANN indicators are 316.23 and 0.87, respectively. XGBoost feature significance and Shapley Augmented Interpretation (SHAP) were used to interpret sediment estimates

and describe the significance of each feature. According to the significance analysis, the prediction of the XGBoost model is mostly (72%) affected by the good width. Additionally, SHAP analysis also checks the importance of water width in the final prediction. Finally, based on the results of this study, further use of XGBoost in water management was recommended. Sahoo et al. (2023) proposed a deep learning-based method to estimate daily suspended sediment load. LSTM networks are a type of recurrent neural network (RNN) that contains memory cells, making them ideal for learning temporal relationships over time. The model was developed using daily observed SSL time series for the Mississippi and Missouri rivers in the United States. The design was evaluated and compared with LSTM and RNN. The model was trained using four different SSL time series as input. The SM-LSTM model with 12 delay inputs outperforms other models with the lowest root mean square error (RMSE) = 32254 tonnes and mean absolute errors (MAE) = 19517 ton, and the highest Nash–Sutcliffe efficiency (NSE) = 0.99 for the Thebes Station while the model with three lagged inputs acted as the best with the lowest RMSE = 2244 ton, and MAE = 1370 ton, and the highest NSE = 0.989 for the Omaha Station. The comparison of prediction accuracies showed that the SM-LSTM model can more satisfactorily predict daily SSL time series than LSTM and RNN. The summary of the conducted research shows the capability of computational intelligence models compared to the classical method of sediment rating curve. This is because there are many computational intelligence models, and the behavior of each watershed is different in producing sediment. Therefore, it is necessary to evaluate and compare these models with classical models using a wide range of computational intelligence models in the conditions of different domains. Therefore, in this study, four computational intelligence methods and a classical method were used and compared with each other to estimate the suspended load of the river. Considering that this river supplies the drinking water of Saveh City and has different weather conditions compared to the studied areas, the study and measurement of its sediments are of great importance. Therefore, the purpose of this research is to investigate the application of soft methods in estimating the suspended sediment of Jalair station on the Qara Chai River and to choose the most appropriate method.

2 Materials and Methods

2.1 The Location of the Jalair Watershed

The area of the Jalair watershed is 16509.30 he, and is located in the Markazi Province of Iran. The highest elevation point is Mount Siah Kemar in the southeast of the watershed, with a height of 3074 m above sea level, and the lowest point of the watershed at the outlet of the basin has a height of 1238 m above sea level. The project area ranges between 50°, 03′, 11.4″ to 50°, 06′, 38.3″ and East 34°, 44′, 58.9″ to 34°, 48′, 58.8″ north latitude and it is located in the UTM system between longitudes 413401 to 418596 and latitudes 3845645 to 3852467 (Fig. 1).

2.2 Research Method

In this research, the performance of four types of support vector machine (SVR) models due to high efficiency and speed, gene expression programming (GEP) due to providing explicit relationships between input and output variables, adaptive neural fuzzy system (ANFIS) due
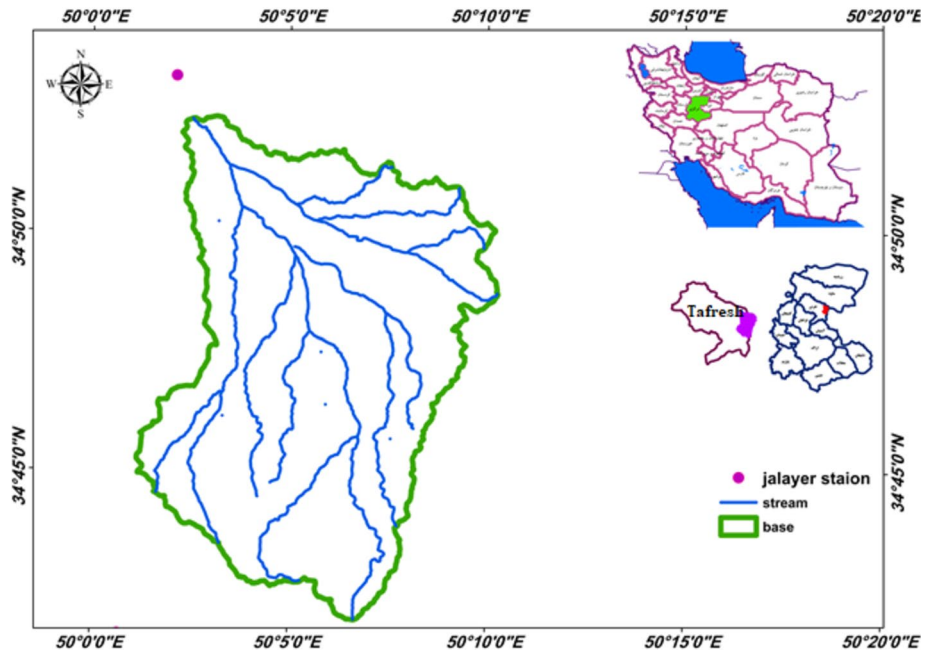
Fig. 1 The location of the study station of the Jalair area in Markazi ProvSince and Iran

to its simplicity and high efficiency and GMDH as a tool with high capability in tracking and diagnosing complex nonlinear trends, especially with a limited number of observations, were used to model the sediment load of Jalair station, Markazi Province. It is important to note that because GMDH operates based on the data obtained from the river system, the characteristics of the river affect the estimation of the results. Then, the results of the four methods were compared with each other and with the results of the sediment rating curve. Finally, the best method was suggested. For this purpose, literature, field studies, and a review of related sources, statistics, and information were collected. The statistics of temperature, rainfall, and daily average discharge of stream and sediment measured daily during a period of 40 years (1981–2021) at Jalair hydrometric station located on the Qarachai River were received from the Meteorology and Regional Water Department of Markazi Province. The received data were categorized and converted into the input format of the models. Based on the discharge and corresponding sediment data, the sediment rating curve was drawn, and its equation was obtained. Appropriate patterns of input variables were selected based on trial and error. Considering that the mentioned parameters have a historical course, the design of the input patterns of soft computing models should be done based on time delays (like what is discussed in the analysis and forecasting of time series). Then, the model was taken for each input and output pattern. In the next step, the most appropriate time delay of the input parameters in the modeling, which had a higher R^2 determination coefficient and a lower root mean square error (RMSE), was selected. In this research, 70% of the research data was used as training and 30% for validation and testing. Finally, four data mining methods were compared with each other as well as with the gauge curve and observational data. The range of changes and statistical characteristics of the parameters of stream flow, sediment flow, precipitation, and daily temperature are presented in Tables 1 and 2. Due to the importance of the basin's; response to the

Table 1 Summary statistics of quantitative data of Jalair Station

Variable	Observations	Minimum	Maximum	Mean	Std. deviation
P(mm)	402	0.000	47.000	5.766	9.270
T°(C)	402	-1.600	31.500	14.446	6.786
Q(m³/s)	402	0.020	190.000	12.941	24.034
S(kg/day)	402	0.000	610002.939	11157.821	51435.167

input variables to the models, in addition to the discharge and sediment variables, rainfall and temperature were also used. The rainfall dynamic variable was used because of its influential role in causing erosion and sediment production. The effect of temperature stands in controlling soil moisture in the area, which has a significant effect on infiltration and runoff generation and, therefore on the amount of suspended sediment.

2.3 Evaluation Criteria

In order to check the accuracy of the results of the models, four statistical criteria, including mean absolute error (MAE), root mean square error (RMSE), mean bias error (MBE), and coefficient of determination (R^2), were used to check the values of estimation error, overestimation, and correlation, respectively. Equations (1) to (3) and Taylor's; diagram were used.

$$NS = 1 - \frac{\sum_{I=1}^{n} (X_O - X_C)^2}{\sum_{i=1}^{n} (X_O - \overline{X}_O)^2} \tag{1}$$

$$MBE = \frac{1}{n} \sum_{i=1}^{n} (x_o - x_c) \tag{2}$$

$$(MAE = \frac{1}{n} \sum_{i=1}^{n} |x_o - x_c| \tag{3}$$

where x_o is the observed data, x_c is the predicted data, the number of observed data, $(\overline{x}_o)$ is the average of the observed data, and $\overline{x}c$ is the average of the predicted data.

Table 2 Pearson's; correlation matrix of data from Jalair station in 40 years

Variables	S(kg/day)	Q(m³/s)	P(mm)	T °(C)
S(kg/day)	**1**	**0.803**	0.062	0.003
Q(m³/s)	**0.803**	1	0.063	-0.011
P(mm)	0.062	0.063	**1**	**-0.342**
T °(C)	0.003	-0.011	**-0.342**	1

Bold values are different from zero with a significance level of $\alpha = 0.05$

3 Results and Discussion

The normality test for the data was investigated by XLSTAT statistical software, and the relevant results are presented in Table 3. Shapiro–Wilk, Anderson–Darling, Lillie-Force, and Jarko Bra tests were used to check whether the data had a normal distribution. In these tests, the null hypothesis is equal to the normality of the data, and the opposite hypothesis is equal to the non-normality of the data. In all the examined tests, the data was not normal. Temperature, rainfall, flow rate, and sediment load data show a significant deviation from the normal distribution. For Jalair station, the data shows a significant deviation from the normal distribution, but the deviation for the sediment load data is more than the flow rate data. It is worth noticing that the reason for using normal probability quantile charts is their ability to show the degree of deviation from the normal distribution. With the help of these graphs, one can comment on the amount of data deviation from normality and its effect on regression performance. Considering the importance of using correct statistical data, all available data were examined for homogeneity by the standard normal homogeneity test, one of the common methods for evaluating data homogeneity. The correlation maps between the parameters of the Jalair station have been presented in Fig. 2.

Due to the non-availability of accurate statistics of erosion and precipitation in the catchment, the sediment rating curve was used in most cases (Fig. 3A) The sediment rating curve for the gauging station at Jalair station was obtained from filed data. In Fig. 3B, the plot (P-P) curve was used to test the normality of the data. By drawing this curve, the cumulative probability of observations was plotted against the cumulative probability of the values calculated from the obtained equation. At first, the data was divided into two parts; then, 70% was considered for training and 30% for testing.

In this research, the performance of four intelligent algorithms, including SVR, GEP, GMD, and ANFIS models, were compared to predict the amount of suspended sediment in Jalair station. In order to estimate suspended sediment, different input patterns, including temperature values, rainfall, previous sediment discharge, and current and previous time step values, were used to determine the effect of each of these variables in suspended sediment modeling. These patterns were selected based on the results of other people's; research, weather conditions, river regime, and the mountainous nature of the region by trial and error method. When the variables are entered into the models in the form of trial and error, it is possible that a variable that has a negligible effect on the correct estimation of the output variable is used in the modeling process, and the influential variables are removed, so this type of modeling justifies and explains the results. It makes the models difficult (Wu et al. 2014). Among the investigated models, the SVR model was first chosen due to its higher efficiency and speed in the modeling process. In order to apply this model, the program developed in the MATLAB software environment was used. In the first step, 13 different scenarios (f1 to f13) were used as input patterns in the SVR model, as described in Table 4:

After sorting the data and determining the independent and dependent parameters for each pattern or scenario, the data was entered into the model, and the model was executed. In this step, for each model, the model was executed and outputted 15 times. Finally, the average R^2 and RMSE of these 15 models were selected and shown in Table 5. In this table, the performance of the SVR model for 13 different input patterns is shown in the form of statistical error indices, as well as the optimal values of the model parameters (σ) for each model. In the next step, 13 models (scenarios) were compared, and the best one with a higher explanation coefficient (R^2) and lower RMSE was

Table 3 The results of the tests to determine the normality of the data at Jalair station

| Jarko Bra test			Lillie-Force test			Anderson–Darling test			Shapiro–Wilk test			Test parameter
Confidence level α	Probability of rejecting the null hypothesis	P-value	Confidence level α	Probability of rejecting the null hypothesis	P-value	Confidence level α	Probability of rejecting the null hypothesis	P-value	Confidence level α	Probability of rejecting the null hypothesis	P-value	
0.05	0.01	0.0001	0.05	0.01	0.0001	0.05	0.01	0.0001	0.05	0.01	<0.0001	P
0.05	0.01	0.088	0.05	0.01	0.01	0.05	0.01	0.006	0.05	0.01	<0.005	T
0.05	0.01	0.0001	0.05	0.01	0.0001	0.05	0.01	0.0001	0.05	0.01	<0.0001	Q
0.05	0.01	0.0001	0.05	0.01	0.0001	0.05	0.01	0.0001	0.05	0.01	<0.0001	S
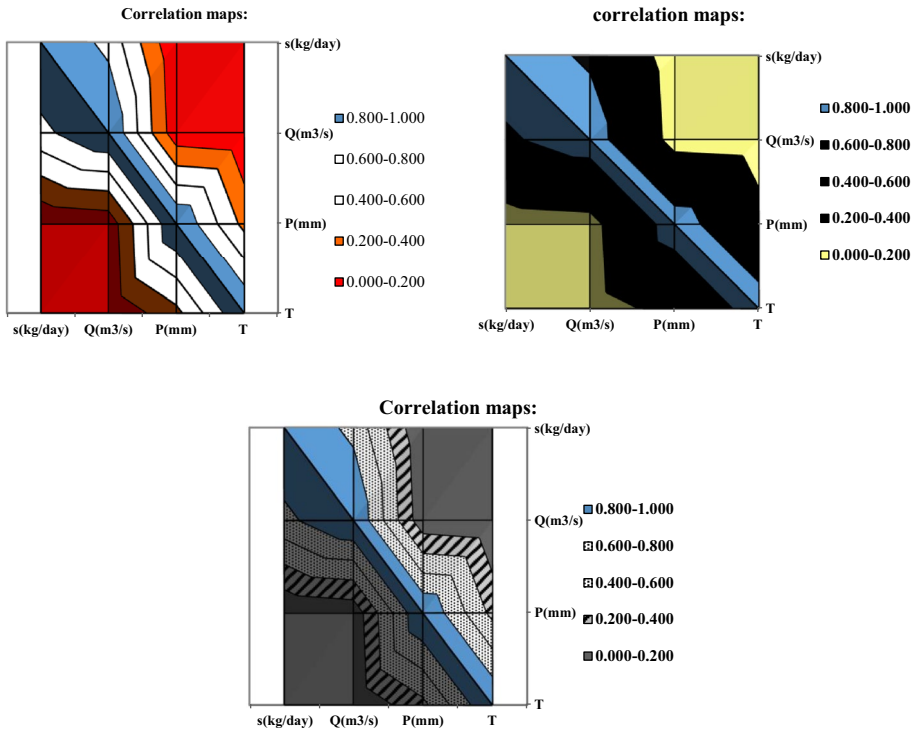
Fig. 2 Correlation maps between parameters of Jalair station

selected. According to the values of statistical error indicators, it can be seen that the best performance of the SVR model has been obtained for model number 4, in which the values of flow rate, temperature, and rainfall of the same day are used as inputs. For the best performance of the model, the values of R^2 and RMSE statistical indicators were obtained in the test phase equal to 0.90 and 35155 kg per day and in the training
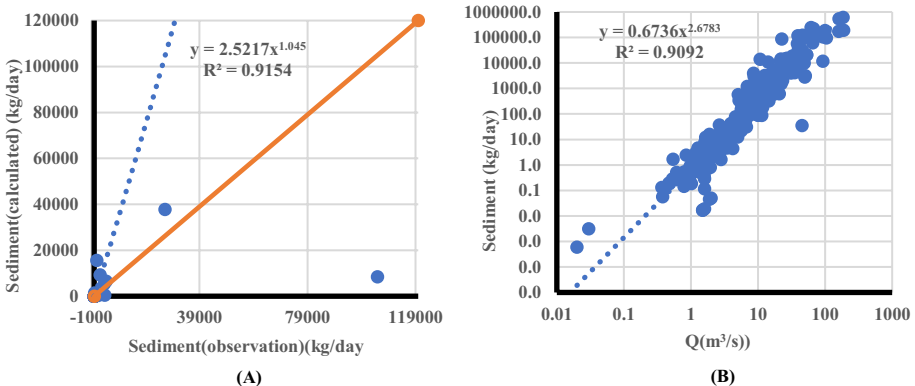

Fig. 3 **A** Sediment rating curve of the training part, **B** Calculation and observation values for the test part

Table 4 Different scenarios used for the models

function	Number	function	Number
Q_t, Q_{t-1}, S_t	(f8)	Q_t, S_t, T_t, P_t	(f1)
$Q_t, Q_{t-2}, Q_{t-1}, S_{t-1}, S_t$	(f9)	$Q_t, Q_{t-1}, S_{t-1}, S_t, T_t, P_t, P_{t-1}$	(f2)
$Q_t, Q_{t-1}, S_{t-1}, S_t, P_t$	(f10)	Q_t, S_{t-1}, S_t	(f3)
$Q_t, Q_{t-1}, S_{t-1}, S_t, T_t, T_{t-1}$	(f11	Q_t, S_t, T_t	(f4)
$Q_t, Q_{t-1}, S_{t-1}, S_t, T_t, T_{t-1}, P_{t-1}, P_t$	(f12)	Q_t, S_t, P_t	(f5)
$Q_t, Q_{t-1}, S_{t-1}, S_t, P_{t-1}, P_t$	(f13)	$Q_t, Q_{t-1}, S_{t-1}, S_t$	(f6)
		Q_t, S_t	(f7)

phase, 0.92 and 13461, respectively. Also, the weakest performance of the model was model number 3, which includes the amount of flow on the same day. The results of statistical error indicators are relatively unacceptable, which shows the dual effects of the behavior between suspended sediment and stream flow. It can be said that by considering the current flow rate and temperature on the same day as input (model no. 5), the performance of the model has improved to a great extent. The results show that the use of flow rate in the time step before and on the same day and the amount of sediment on the same day (model no. 6) has improved the performance of the model to a great extent. Also, the use of debit on the same day (pattern no. 7) has slightly improved the results. By using the flow rate values alone in the time step of the previous and the same day (pattern no. 8), the results have improved. In patterns 9 to 13, the results have improved to some extent. Figures 4 and 5 show the program's; output for the best pattern (pattern 1). The time series and scatter diagram of observed and simulated suspended sediment data for the best input model (model no. 1) for the SVR model are shown. According to this figure, it can be said that the SVR model has been able to

Table 5 Parameter values and SVR model's; performance for different input patterns during the test period

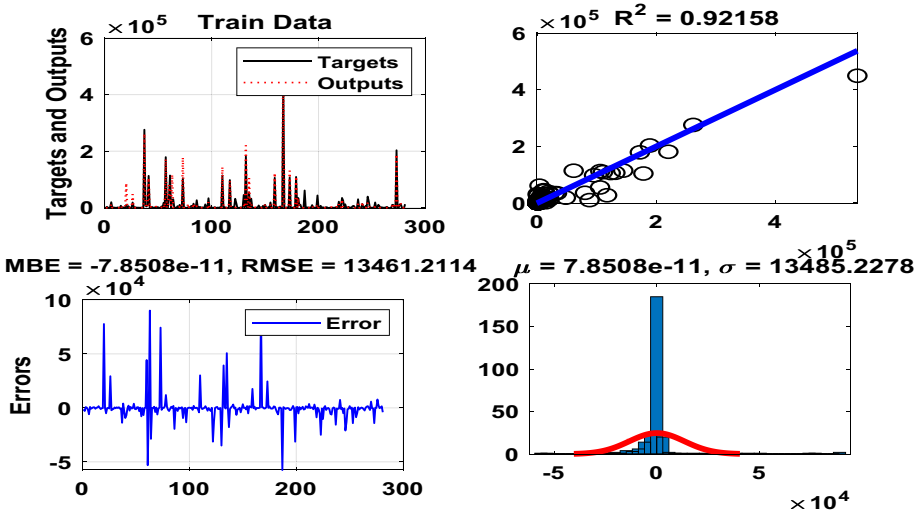| Statistical indicators of performance in the test phase			Statistical indicators of performance in the training phase			Model
R^2	σ	RMSE (kg/day)	R^2	σ	RMSE (kg/day)	
0.83	13763	13708	0.81	872448	24444	SVR1
0.71	19976	19893	0.90	14499	14473	SVR2
0.25	64868	65007	0.14	27094	27046	SVR3
0.90	35155	35158	0.92	13485	13461	SVR4
0.81	45512	45515	0.71	18463	18430	SVR5
0.72	16372	16329	0.90	15332	15305	SVR6
0.86	39355	39429	0.70	25589	25544	SVR7
0.78	31248	31129	0.83	17411	17380	SVR8
0.64	27386	27784	0.86	18601	18568	SVR9
0.78	41789	42300	0.86	17380	17349	SVR10
0.79	31300	31417	0.86	18409	18376	SVR11
0.80	17318	17251	0.81	244887	24444	SVR12
0.71	22925	22939	0.90	14492	14473	SVR13

Fig. 4 The output results of the model in the MATLAB environment during the training phase

show the non-linear and complex relationship between the input and output values. The main weakness of the model is in predicting the peak values of suspended sediment. In general, the results obtained in this section show that by removing the temperature values in the current time step, the model's; performance decreases significantly.
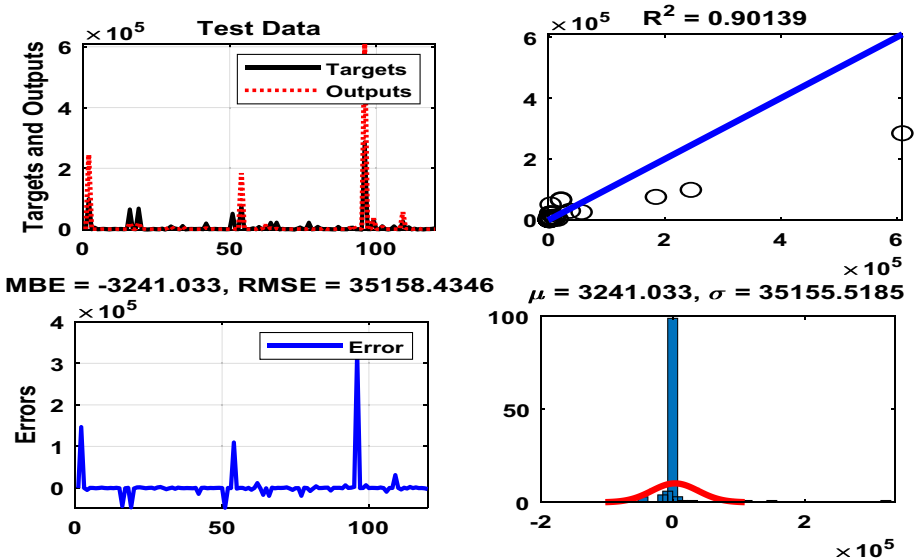

Fig. 5 The output results of the model in the MATLAB environment during the test phase

Table 6 Parameter values and ANFIS model performance for different input patterns in the test period

Statistical indicators of performance in the test phase			Statistical indicators of performance in the training phase			Model
R^2	σ	RMSE (kg/day)	R^2	σ	RMSE (kg/day)	
0.73	26767	26896	0.67	33147	33088	F1
0.79	39775	39844	0.82	20404	20606	F2
0.42	26028	25919	0.34	39304	39238	F3
0.80	39592	39719	0.64	30768	307	F4
0.77	35395	35719	0.62	19186	19153	F5
0.72	13591	13575	0.69	28100	28060	F6
0.80	50185	50259	0.61	22078	22039	F7
0.80	50353	50436	0.61	22074	22034	F8
0.76	30558	31425	0.79	22229	22207	F9
0.63	17607	17577	0.79	235	23384	F10
0.79	19061	18990	0.87	16075	16083	F11
0.89	19478	19496	0.71	19404	19376	F12
0.93	16556	16608	0.68	19847	19813	F13

3.1 The Results of the ANFIS Method

In this step, for each model, the model was executed and outputted 15 times. Finally, the average R^2 and RMSE of these 15 models were selected and shown in Table 6. In this table, the performance of the ANFIS model for 13 different input patterns is shown in the form of statistical error indices, as well as the optimal values of the model parameters (σ) for each model. In the next step, 13 models (scenarios) were compared, and the best one with a higher explanation coefficient (R^2) and lower RMSE was selected. By comparing the outputs of the model, model number 13 was chosen because it had a higher R^2 explanation coefficient and a lower RMSE. In the next step, different methods are used to divide the data in the ANFIS model; one of the standard methods in this regard is the network separation method. This method is based on choosing the type of membership function (triangular, trapezoidal, Gaussian, bell-shaped, etc.) and the number of membership functions for each input variable. In this research, all kinds of membership functions were evaluated using the optimal input model (model no. 13), and the values of the RMSE index corresponding to each function were reported in Table 7. It should be noted that the number of membership functions was obtained using trial and error for the lowest RMSE value. According to Table 7, it can be said that the best membership function is of the triangular type, the number of which is 2, 2, and 4 for the variables of flow rate, rainfall in the current time steps and

Table 7 The performance of the ANFIS model according to different membership functions for the optimal model

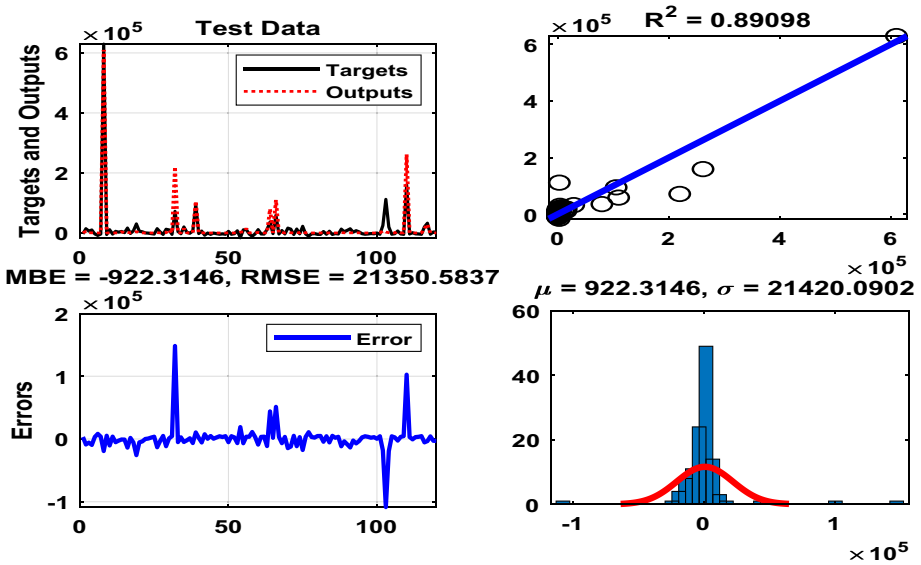| RMSE (kg/day)	Number of membership functions	Membership function type
16556	2,2,4	Triangular
17562	2,3,4	Gossi
17892	2,2,3	trapezoidal
16981	3,2,2	a bell

Fig. 6 The output results of the model in the MATLAB environment during the test phase

flow rate, sediment, and rainfall in a delay step, respectively. Russel and Campbell (1996) also stated that using triangular membership functions gives better results in practical terms. In Figs. 6 and 7, the model's; output is shown for the optimal pattern (pattern 1). According to these figures, it can be seen that the performance of the ANFIS model in predicting suspended sediment values is similar to the performance of the SVR model, However, the values of the statistical indicators indicate the superiority of the SVR model.
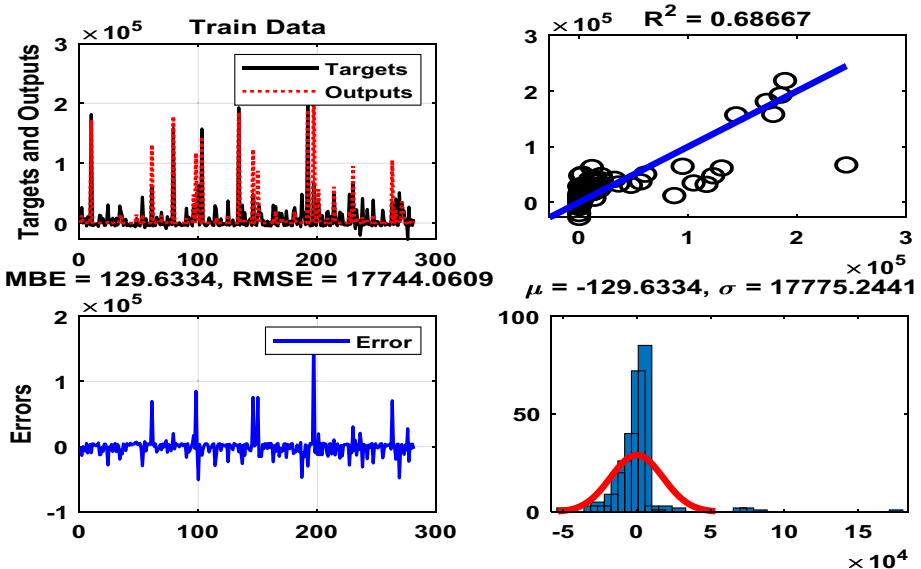
Fig. 7 The output results of the model in the MATLAB environment during the training phase

Table 8 Values of parameters and operators used in the GEP model at Jalair station

Genetic operators		General settings	
0.044	Mutation rate	30	Number of chromosomes
0.1	Inversion rate	7	head size
0.1	Consecutive insertion tranche rate	3	Number of genes per chromosome
0.1	The rate of his song is the root of consecutive insertion	100000	Number of production population
0.3	Single point combination rate	Sum (+)	Link function

3.2 The Results of the Gene Expression Programming (GEP) Model Method

As mentioned before, the first step in using the GEP model is to choose the appropriate fitting function. In this research, the results of the selection of the fitting function in the GEP model indicated that the use of the relative root mean square error (RRSE) fitting function has better results compared to other functions for modeling suspended sediment. Therefore, the RRSE function was chosen as the fitting function in the model. The values of the parameters and operators used in the GEP model at Hasan Abad station are presented in Table 8. In this research, the model's; performance for specific sets of input patterns is shown in Table 9. In this table, the performance of the GEP model for 13 different input patterns is shown in the form of statistical error indices, as well as the optimal values of the model parameters for each pattern. In the next step, 13 models (scenarios) were compared, and the best one with a higher explanation coefficient (R^2) and lower RMSE was selected. According to the table and the values of statistical error indicators, it can be seen that the best performance of the GEP9 model has been achieved for model number 8. The lowest value of the RMSE index was obtained for the GEP9 function.

Table 9 Values of parameters and performance of the GEP model for different input patterns in the training and test period

Statistical indicators of performance in the test phase					Statistical indicators of performance in the training phase					Model
RRSE	MAE	MSE	RMSE (kg/day)	R^2	RRSE	MAE	MSE	RMSE (kg/day)	R^2	
0.79	5.71	46.86	6.83	0.38	0.844	4.70	32.77	5.72	0.28	SVR1
0.74	5.26	41.50	6.44	0.45	0.823	4.42	31.17	5.58	0.32	SVR2
0.14	1243	0.00	3973	0.98	0.894	5040	0.00	0.00	0.87	SVR3
0.80	0.94	48.35	6.95	0.36	0.894	4.97	36.73	4.06	0.30	SVR4
0.18	2921	0.00	4840	0.97	0.350	5855	0.00	0.00	0.87	SVR5
0.20	1770	0.00	5624	0.96	0.564	6961	0.00	0.00	0.68	SVR6
0.18	2409	0.00	4560	0.97	0.372	6506	0.00	0.00	0.86	SVR7
0.13	982	0.00	3721	0.98	0.297	4926	0.00	0.00	0.91	SVR8
0.18	1461	0.00	5062	0.97	0.370	2109	0.00	0.00	0.86	SVR9
0.29	1935	0.00	7989	0.91	0.365	6762	0.00	0.00	0.86	SVR10
0.75	5.26	42.82	6.54	0.43	0.719	3.400	23.73	4.87	0.49	SVR11
0.75	5.22	42.70	6.53	0.44	0.695	3.406	22.16	4.708	0.51	SVR12
0.28	1940	0.00	7595	0.92	0.321	5690	0.00	0.00	0.89	SVR13

Table 10 The results of applying different mathematical functions on the value of the RMSE index

R^2	MAE	RMSE (kg/day)	Mathematical model	Function
0.99	0.01	0.04	$+ - * /$	F1
0.99	0.04	0.07	$+ - * / \ln x, e^x$	F2
0.99	0.02	0.04	$+ - * /, \sqrt{X}, \sqrt[3]{X}, x^3, x^2$	F3
0.99	0.01	0.04	$+ - * / \ln x, e^x, \sqrt{X}, \sqrt[3]{X}, x^3, x^2$	F4
0.99	0.01	0.03	$+ - * / \ln x, e^x, \sqrt{X}, \sqrt[3]{X}, x^3, x^2, \sin x,$ $\cos x$, Arctg x	F5
				Link function type
0.77	0.09	0. 20	$+$	plural
0.77	0.09	0.20	$-$	subtraction
0.90	0.02	0.05	$\times$	multiplication
0.91	0.03	0.09	$/$	Division

The next step is to choose the leading operators to build the parse tree. The mathematical functions used in this research and the model's; performance for a specific set of functions are shown in Table 10. This table shows the results of using different mathematical functions on R^2, RMSE, and MAE index values in Jalair Station. After choosing the best combination of mathematical functions, the next step involves finding the appropriate link function. Among the link functions, including addition, subtraction, multiplication, and division, the division link function had a better performance than other functions, and the results presented in Table 10 confirm this issue. Figure 8 shows the time series and dispersion of observed and simulated data with the GEP model during the test period. According to this figure, it can be seen that the GEP model is acceptable and meaningful for predicting suspended sediment values, and the model has been able to perform well in predicting suspended sediment peak values. Figure 9 shows the output flowchart of the model. According to the flowchart, the output equation of the model is very complicated.

The comparison of the results of three ANFIS, GEP, and SVR models shows the superiority of the GEP model in predicting the amount of suspended sediment according
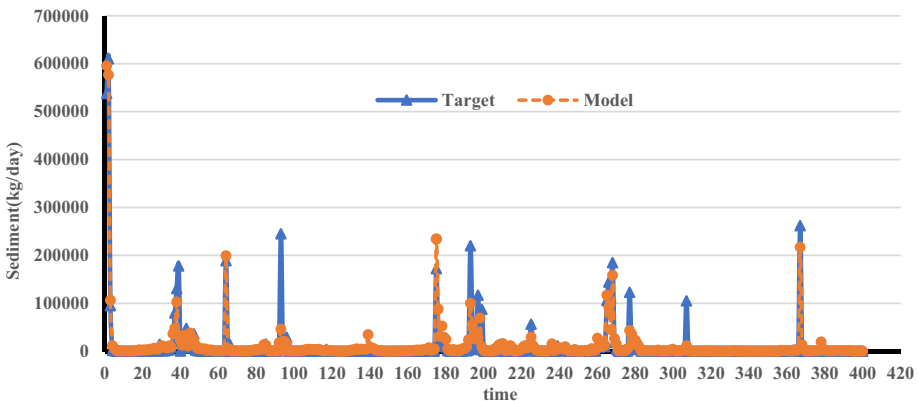

Fig. 8 Time series diagram of observational and calculated sediment values from the GEP model
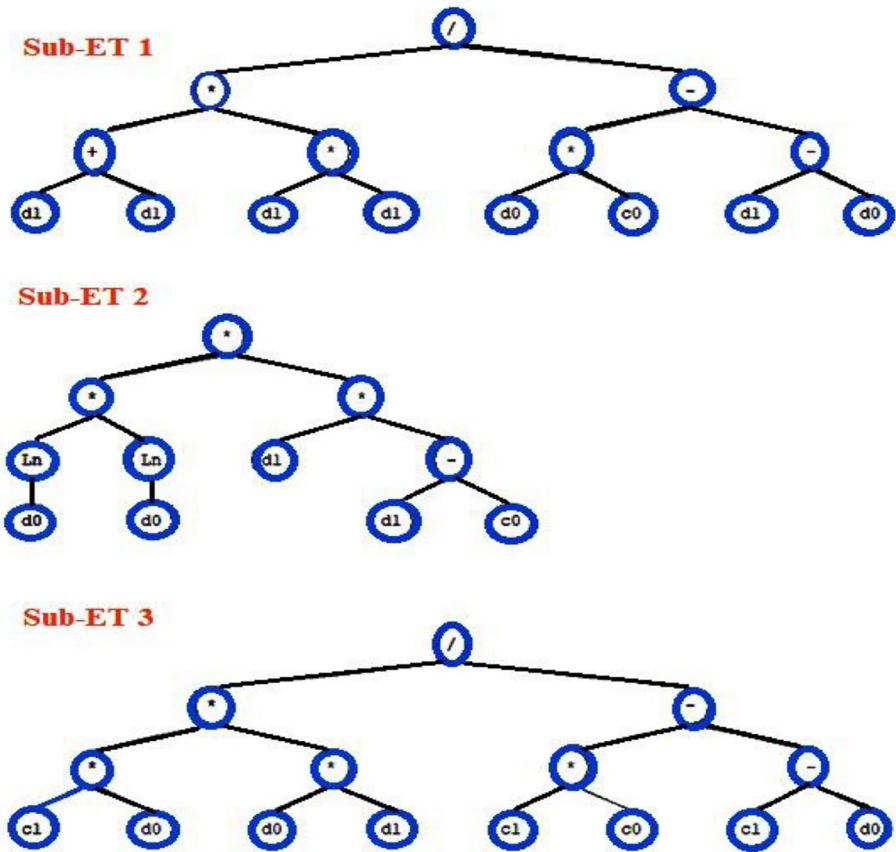
Fig. 9 Flow chart of the Gene Expression Algorithm

to the input model 13. These results are consistent with the research results of Sheikh Alipour et al. (2014) and Kisi and Shiri (2012). In the next step, the best-selected pattern of (ANFIS), (SVM), and (GEP) models were used as the input of the GMDH model. First, input pattern 13, the best pattern for (ANFIS) models, was introduced as the input of the GMDH model. In the training and test phase, the values of R^2 statistical indices were equal to 0.95 and 0.85, and the RMSE error value was equal to 27216 and 18639, respectively. The best-selected pattern model (SVM) pattern number 4 was used as input to the GMDH model. In the training and test phase, R^2 statistical indices were 0.97 and 0.82, respectively, and the RMSE error value was 37082 and 23673, respectively. Then, input pattern 8, the best pattern for the GEP model, was introduced as GMDH input. In the training and test phase, the values of R^2 statistical indicators were equal to 0.99 and 0.80, and the RMSE error value was equal to 18638 and 24170, respectively. Figures 10 and 11 show the output results of the GMDH model in the MATLAB environment during the training and test phase at Jalair Station. The results showed the acceptable performance of the GMDH model with the highest R^2 determination coefficient of 0.97, 0.95, and 0.99 and the lowest root mean square error of 37082, 27216, and 18638 (kg/day) in the test phase. A comparison of the performance of the
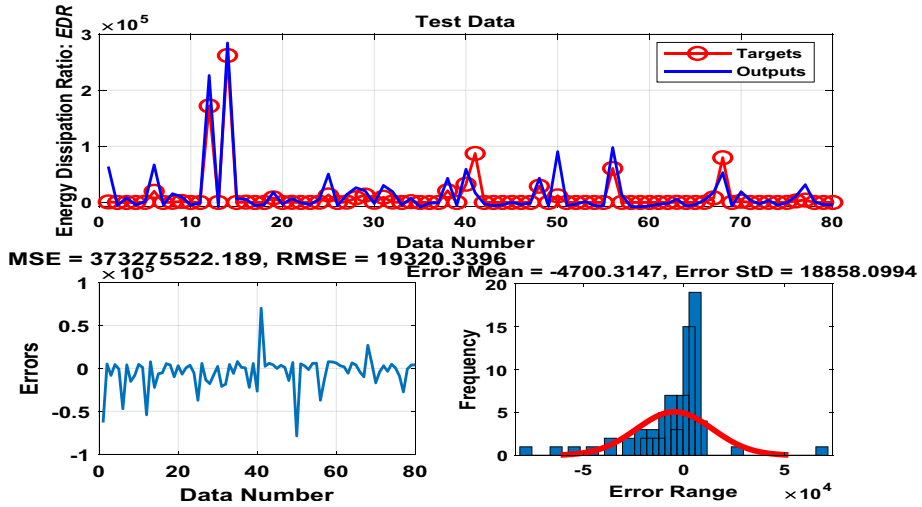
Fig. 10 The output results of the GMDH model in the MATLAB environment during the training phase

models used to estimate the suspended load for the best input model in Table 11 was shown. Figure 12 of the Taylor diagram for visual inspection. According to the obtained results, it can be seen that the performance of the GMDH model is better compared to other models. GEP, SVR, and ANFIS models are ranked second, third, and fourth. The results showed that all four investigated data mining methods provide far better results than the sediment gauge curve. According to the obtained results, it can be said that the GMDH model, as a powerful and high-speed model, can be used to model suspended sediment in the Jalair catchment. Since the peak points are essential in determining the amount of storage capacity or water passage of various structures and are essential information needed in the design of all structures, this model has been able to predict the sediment peak values well.

The result obtained from this research shows that all four soft calculation methods used in this study can estimate suspended load. Machine learning methods have higher accuracy than the sediment gauge curve method. This can be due to the regression property of the sediment rating curve. This means that most of the data used for modeling from the SRC method are related to low flow rates, and because the most enormous amount of sediment is transported at high flow rates, this model cannot introduce high sediment periods. Also, converting the results of this model from a logarithmic space to an arithmetic space will cause an underestimation of the suspended sediment

Table 11 Comparison of the performance of the models used to estimate the suspended load for the best input model

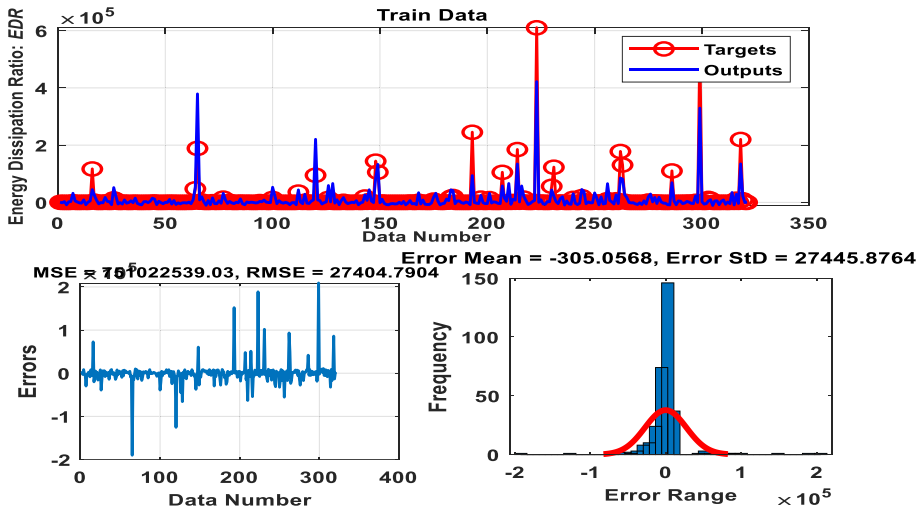| NS	MBE	RMSE (kg/day)	R^2	Optimal Pattern	Model
0.85	3241	35158	0.90	F4	SVR
0.69	2006	16556	0.93	F13	ANFIS
0.99	239	3721	0.98	F8	GEP
0.16	1150	8928	0.63	SRC	
0.77	2621	37082	0.96	F4	GMDH
0.75	2750	27216	0.95	F13	GMDH
0.79	2506	18638	0.97	F8	GMDH

Fig. 11 The output results of the GMDH model in the MATLAB environment during the training phase

load. In comparison, intelligent models such as ANFIS, GEP, SVR, and GMDH will have an accurate estimate of the SSL value even when the data is not of good quality and quantity. According to the obtained results, the reason for the superiority of the results of some models over others in this research can be stated as follows: the support vector machine model is more accurate and efficient than the artificial neural network. The support vector machine model determines the best decision boundary for separating the data, while the artificial neural network stops learning when it reaches the first separating decision boundary. In this case, the obtained decision boundary may not be suitable. The execution time of support vector machines depends on the number of available categories and requires less time. At the same time, the execution time of the artificial neural network depends on the number of training inputs, the number of hidden layer neurons, the training function, the learning coefficient, and the momentum, which are obtained by trial and error. In terms of the speed of operation of the models (speed of code execution and calculations) in estimation, the data control group model is in the first place, the neural-fuzzy adaptive inference system is in the second place, and the support vector machine model is in the third place, and the gene expression programming is in the next places have. The method of gene expression programming is vital due to the presentation of a mathematical relationship for the model and the possibility of using that relationship for future data. From this point of view, it can be preferable to the other three models for modeling. The relations produced between the input and output variables in the models are very complex, which is one of the weak points.
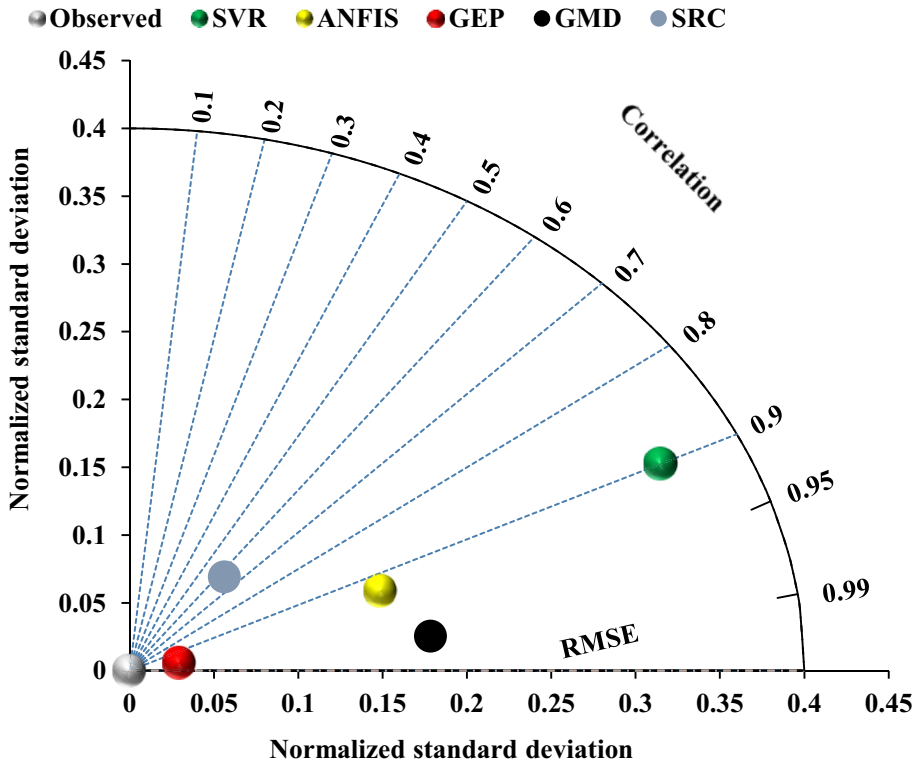
Fig. 12 Comparison of the performance of the models used to estimate the suspended load with the visual method of the Taylor curve method

4 Conclusion

Modeling the suspended sediment of rivers is very important and can influence the management and exploitation of water bodies and river morphology. In this study, the performance of ANFIS, GMDH, GEP, and SVR models was investigated in predicting the f suspended sediments in the study river. In this regard, sample data of flow discharge and suspended load, rainfall, and temperature of Jalair Station located in the catchment area of QaraChai River in the Markazi Province of Iran were used over 40 years. Various input patterns, including flow rates and suspended sediment, temperature, and rainfall, were used to model suspended sediment in the current and previous time steps. The obtained results indicated the acceptable performance of the methods used in predicting suspended sediment amounts. Comparing the results of ANFIS, GEP, GMDH, and SVR models indicates the superiority of the GEP model in predicting suspended sediment amounts. The results showed the acceptable performance of the GEP model with the highest coefficient of determination R^2 equal to 0.97 and the lowest root mean square error equal to 3721 kg per day, respectively. According to the obtained results, the GEP model can be used as a powerful model to model the suspended sediment at the Jalair Station. ANFIS, GMDH, and SVR models are ranked second to fourth. The research showed that all four data-mining methods have far better efficiency and accuracy in estimating the suspended river sediment load

than the sediment rating curve. Data mining-based methods can be used as an alternative to estimate the suspended load of the river.

Due to climate change and droughts, industrial development, colonized land use changes, and changes in the morphology of watersheds, the obtained results cannot be used forever at any time, and the conditions must be updated when using the models. Another weakness of the models is that with the increased number of developed layers, the accuracy of the produced answers increases, but the produced relationships between the input and output variables become very complicated. In most studies, one or two input parameters of flow rate and rainfall have been used. One of the strengths of this research is the use of several input parameters (rainfall, temperature, flow rate, sediment flow rate). The patterns were selected based on the results of others' research; weather conditions, river regimes, and the mountainous nature of the region by trial and error. When the variables are entered into the models by trial and error, it is possible that a variable that has little effect on the correct estimation of the output variable is used in the modeling process, and the influential variables are removed. Therefore, this type of modeling makes it difficult to justify and explain the results of the models (Wu et al. 2014). The gamma test method, as a data preprocessing method, is suggested to select suitable combinations of input variables for the models. It is also suggested that the effectiveness of the inputs of this research be investigated using other innovative models (such as neurophase models, decision trees, etc.) and comparing their results with the current research as well as its application in other catchments. In order to complete the research, it is better to use other variables as inputs to the models, in addition to the hydrological and climatic variables of the basin (rainfall and temperature). Also, the use of meta-heuristic algorithms (for example, genetic algorithm) in setting the parameters of SVR models can increase the accuracy of suspended sediment estimation modeling. This improvement in predicting suspended sediment is valuable for planning and managing water resources.

Availability of Data and Materials The data used in the text of the article can be provided at any time and by anyone.

Declarations

Ethical Approval This article does not contain any studies with human participants or animals performed by any authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

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

References

Adnan RM, Yaseen ZM, Heddam S, Shahid S, Sadeghi-Niaraki A, Kisi O (2022) Predictability performance enhancement for suspended sediment in rivers: inspection of the newly developed hybrid adaptive neuro-fuzzy system model. Int J Sedim Res 37(10):383–398. https://doi.org/10.1016/j.ijsrc.2021.10.001

Alp M, Cigizoglu HK (2005) Suspended sediment load simulation by two artificial neural network methods using hydrometeorological data. Environ Model Softw 22:2–13

Asadi M, Fathzadeh A (2017) Investigating the effectiveness of models based on computing intelligence in estimating the suspended load of the river (case study: Gilan province). J Rangeland Watershed Manag Nat Resour Iran 1(71):45–60. (In Persian)

Aytek A, Kisi O (2007) A genetic programming approach to suspended sediment modeling. J Hydrol 351:288–298

Azamathulla H, Cuan Y, Ghani AAB, Chang CHK (2013) Suspended sediment load prediction of river systems: GEP approach. Arab J Geosci 3:3469–3480

Beiranvand N, Sepahvand A, Haghi Zadeh A (2023) Suspended sediment load modeling by machine learning algorithms in low and high discharge periods (Case study: Kashkan watershed). Water Soil Manag Model 3(2). https://doi.org/10.22098/mmws.2022.11262.1115

Doroudi S, Sharafati A, Mohajeri SH (2021) Estimating daily suspended sediment load using a novel hybrid support vector regression model incorporated with an observer-teacher learner-based optimization method. Complexity 5540284:1–13. https://doi.org/10.1155/2021/5540284

Duan WL, He B, Takara K, Luo PP, Nover D, Hu MC (2015) Modeling suspended sediment sources and transport in the Ishikari River basin, Japan, Using SPARROW. Hydrol Earth Syst Sci 19:1293–1306. https://doi.org/10.5194/hess-19-1293-2015

Eder AP, Strauss T, Krueger B, Quinton B (2010) A Comparative calculation of suspended sediment loads concerning hysteresis effects (in the Petzenkirchen catchment), Austria. J Hydrol 389:168–176. https://doi.org/10.1016/j.jhydrol.2010.05.043

Firat M, Gungor M (2009) Monthly total sediment forecasting using adaptive neuro fuzzy inference system. Stoch Environ Res Risk Assess 24:259–270

Ghani AAB, Azamathulla HMD, Chang CHK, Zakaria NA, AbuHasan Z (2010) Prediction of total bed material load for rivers in Malaysia: a case study of Langat, Muda and Kurau Rivers. Environ Fluid Mech 11:307–318

Hesavi M, Shafaee-bajestan M (2010) Estimate of sediment bed load in karoon river ahwaz station. Int River Eng Conf, Ahwaz, Iran 1–9

Ivakhnenko AG (1968) The group method of data handling-a rival of the method of stochastic approximation. Soviet Automatic Control c/c of Avtomatika 1:43–55

Keshtegar B, Piri J, Hussain WU, Ikram K, Yaseen M, Kisi O, Adnan RM, Adnan M, Waseem M (2023) Prediction of sediment yields using a data-driven radial M5 tree model. J Water 15:1437. https://doi.org/10.3390/w15071437

Kisi O (2005) Suspended sediment estimation using neuro-fuzzy and neural network approaches. Hydrol Sci J 50:683–696

Kisi O, Ozkan C (2017) A new approach for modeling sediment-discharge relationship: local weighted linear regression. Water Resour Manag 31:1–23

Kisi O, Ozkan C, Bahriye A (2012) Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm. J Hydrol 428–429(4):94–103. https://doi.org/10.1016/j.jhydrol.2012.01.026

Kisi O, Shiri J (2012) River suspended sediment estimation by climate variables implication: comparative study among soft computing techniques. Comput Geosci 43:73–82

Kisi O, Yuksel I, Dogan E (2008) Modeling daily suspended sediment of rivers in Turkey using several data-driven techniques. Hydrol Sci J 53(6):1270–1285

Melesse AM, Ahmad S, McClain ME, Wang X, Lim YH (2011) Suspended sediment load prediction of river systems: an artificial neural network approach. Agric Water Manag 98:855–866

Moradinejad A, Davodmaqami D, Moradi M (2018) Investigating the effectiveness of methods for estimating the suspended sediment load of Qara Chai River. Environ Water Eng (4)5:328–338. https://doi.org/10.22034/jewe.2020.211925.1341

Nikpour MR, Sanikhani H (2017) Modeling of river suspended sediments using soft calculations (Darah-Rood River). Irrig Water Eng Sci Res Quart 30:29–44

Nourani V, Gokcekus H, Gelete G (2020) Estimating suspended sediment load using artificial intelligence-based ensemble model. Complexity 1–19. https://doi.org/10.1155/2021/6633760

Onderka M, Krein A, Wrede S (2012) Dynamics of storm-driven suspended sediments in a headwater catchment described by multivariable modeling. J Soils Sediments 12:620–635

Piraei R, Afzali SH, Niazkar M (2023) Assessment of XGBoost to estimate total sediment loads in rivers. Water Resour Manag 37:5289–5306. https://doi.org/10.1007/s11269-023-03606-w

Rahul AK, Shivhare N, Kumar S, Dwivedi SB, Dikshit PKS (2021) Modeling of daily suspended sediment concentration using FFBPNN and SVM algorithms. J Soft Comput Civ Eng 5(2):120–134. https://doi.org/10.22115/scce.2021.283137.1305

Rajaee T, Mirbagheri SA, Kermani MZ, Nourani V (2009) Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Sci Total Environ 407(17):4916–4927

Russel SO, Campbell PF (1996) Reservoir operating rules with fuzzy programming. J Water Resour Plan Manag 122(3):165–170

Sahoo BB, Sankalp S, Kisi O (2023) A novel smoothing-based deep learning time-series approach for daily suspended sediment load prediction. Water Resour Manag 37:4271–4292. https://doi.org/10.1007/s11269-023-03552-7

Sheikh Alipour Z, Hasanpour F, Azimi A (2014) Comparison of artificial intelligence methods in estimating suspended sediment load (case study: Sistan River). Water Soil Conserv Res 2(7):41–60

Wu W, Dandy G, Maier H (2014) Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modeling. Environ Model Softw 54:108–127