

## Modeling soil erosion by data-driven methods using limited input variables

Shahla Yavari, Saman Maroufpoor and Jalal Shiri

### ABSTRACT

Soil is one of the main elements of natural resources. Accurate estimation of soil erosion is very important in optimum soil resources development and management. Analyzing soil erosion by water on cultivated lands is an important task due to the numerous problems caused by erosion. In this study, the performance of three different data-driven approaches, e.g. multilayer perceptron artificial neural network (ANN), grid partitioning (GP), and subtractive neuro-fuzzy (NF) models were evaluated for estimating soil erosion. Land use, slope, soil and upland erosion amount were used as input parameters of the applied models and the erosion values obtained by MPSIAC method were considered as the benchmark for evaluating the ANN and NF models. The applied models were assessed using the coefficient of determination ( $R^2$ ), the root mean square error (RMSE), the BIAS, and the variance accounted for (VAF) indices. The results showed that the subtractive NF model presented the most accurate results with the minimum RMSE value (3.775) and GP, NF and ANN models were ranked successively.

**Key words** | erosion, MPSIAC, neural networks, neuro-fuzzy

**Shahla Yavari** (corresponding author)  
Geology Department,  
Kharazmi University,  
Tehran,  
Iran  
E-mail: [shahlayavari3@gmail.com](mailto:shahlayavari3@gmail.com)

**Saman Maroufpoor**  
**Jalal Shiri**  
Department of Water Engineering,  
Faculty of Agriculture,  
University of Tabriz,  
Tabriz,  
Iran

### INTRODUCTION

The planning of measures for protection and conservation of soil in watersheds is an important issue in soil conservation, hydrology and optimum watershed management. Study of soil erosion by water on cultivated lands is crucial due to the numerous problems caused by the erosion as well as the significant environmental and economic consequences (Tarolli & Sofia 2016). Soil erosion is a serious environmental problem that negatively influences agricultural productivity, water quality, aquatic ecology, and river morphology (Peter *et al.* 2010). Land degradation is defined as the loss in potential utility or productivity of land resources in a country that is usually subjected to serious soil erosion. It is a major worldwide environmental problem, especially in arid and semi-arid regions.

Erosion adversely affects climate, soil fertility, vegetation cover, economy and human welfare (Polyakov & Lal 2008). Soil loss may be reduced in cropped soils by

optimum soil management and tillage practices (Boix-Fayos *et al.* 2005; Cerdan *et al.* 2010; Vanwallegem *et al.* 2011). For erosion control, a sustainable land management perspective is required. Watershed management practices are variable, taking into consideration land use and vegetation cover with the intention of rehabilitation of degraded lands and protection of soil and water systems (Alemayehu *et al.* 2009). However, watershed management approaches need to be adapted to the local situation (Darghouth *et al.* 2008). There are numerous methods for estimating watershed sediment yield. The PSIAC (Pacific Southwest Inter-Agency Committee) model (PSIAC 1968) is the commonly used method for evaluating soil erosion (Clark 2001).

The PSIAC model was developed primarily for application in arid and semi-arid areas in the southwestern USA, and is believed to be appropriate for the same

environmental conditions in other regions, e.g. Iran (Bagherzadeh 1993; Sadeghi 1993). Usually, the evaluation of the erosion processes is carried out by using the conceptual models, which present good results. However, those models are difficult to develop and the calibration of the model parameters is subjective. Alternatively, experimental models as well as data-driven techniques (e.g. artificial neural network (ANN) or neuro-fuzzy (NF) techniques) may be used, which connect inputs and output by means of a mathematical function without an explicit relationship with the catchment specifications (Pandey *et al.* 2016).

ANN is a computing method that mimics the human brain and nervous system. It has a mathematical structure able to approximate arbitrarily intricate nonlinear processes that relate the inputs and output of any system. ANN has been applied successfully for modeling intricate nonlinear input/output time series relationships, classification, pattern recognition and other problems in a wide diversity of fields (Kisi 2005). The high degrees of empiricism and approximation in the analysis of hydrologic systems are highly suitable for the application of ANN (Hsu *et al.* 1995). ANNs are used for simulating hydrologic variables such as stream flow, temperature, snow melting, rainfall-runoff and suspended materials. Among others, ANNs have been successfully employed in estimation of maximum flood (Bodri & Cermak 2000), estimation of flow rate (Dibike & Solomatine 2001), modeling precipitation and runoff (Nagy *et al.* 2002), reservoir operation (Chang *et al.* 2005), hydrologic time series modeling and sediment transport prediction (Firat & Gungor 2004; Agarwal *et al.* 2006; Singh *et al.* 2012), nutrient concentrations in surface runoff (Kim & Gilley 2008), direct runoff estimation (Dhamge *et al.* 2012), and prediction of event-based storm-water runoff quantity and quality (He *et al.* 2011). Moreover, the adaptive NF inference system has been applied to simulate/predict different factors in hydrologic sciences (e.g. Besalatpour *et al.* 2013). Only limited researches have been carried out on application of ANN and NF techniques for predicting soil erosion rate, including, for example, Metternicht & Gonzalez (2005), Kim & Gilley (2008), and Demirel & Tüzün (2011). Nonetheless, there is not any comparative study on assessing the NF and ANN models' capabilities for simulating soil erosion in different watersheds. Also, it is of great importance to

establish models (e.g. ANN and NF) which can simulate soil erosion with limited input parameters. The objective of this study is to evaluate the capabilities of NF and ANN (based on limited input parameters) in modeling soil erosion, using the standard MPSIAC erosion values as benchmark.

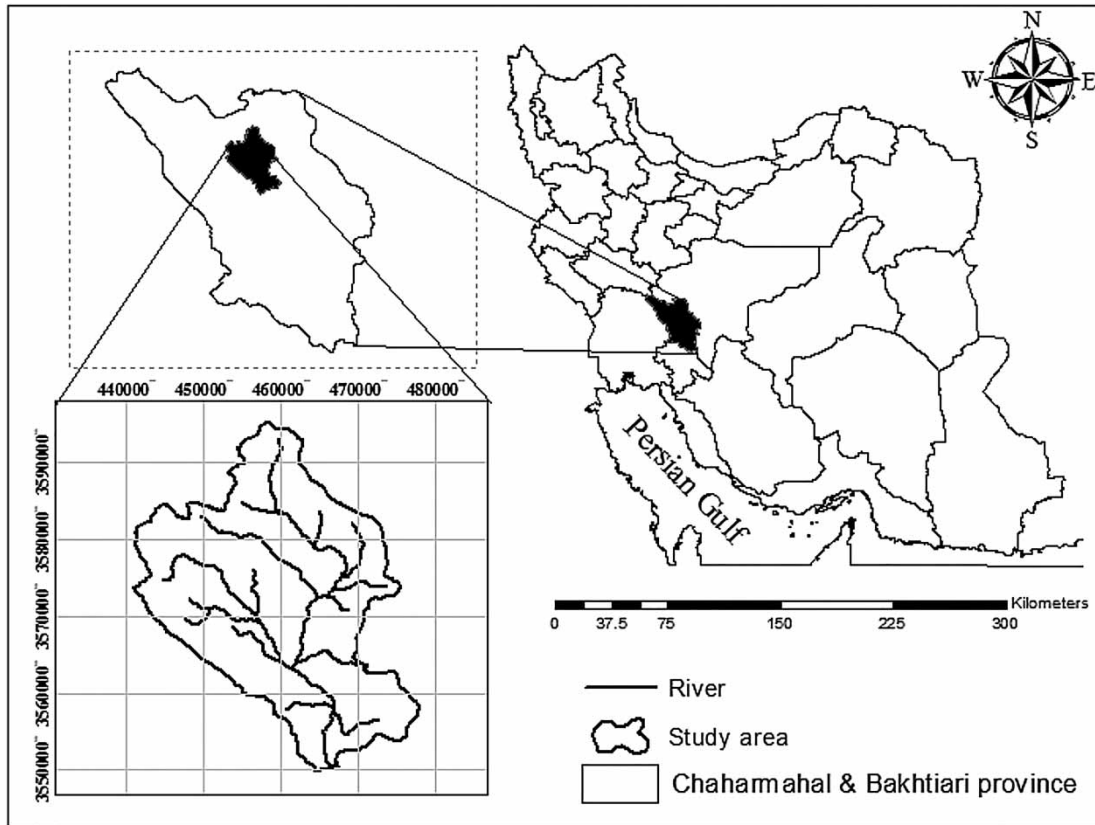
## MATERIALS AND METHODS

### Study area

Observed data from the Jooneghan watershed, located in Chaharmahal and Bakhtiari province in the west of Iran, were used here for developing and validating the applied models. The watershed is situated between latitudes 32°29' to 32°40' N and longitudes 50°47' to 50°20' E. The watershed covers 903.92 km<sup>2</sup>. Figure 1 shows the location of the studied watershed. The highest elevation in the watershed is 3,580 m. The average annual rainfall in the watershed is 512.21 mm, and about 80% of the annual rainfall occurs between January and February. The mean monthly temperature varies between -0.8°C and 22.2°C. Range condition over the entire watershed is poor and has contributed to excessive runoff and soil erosion. The dominant geology in the watershed is units of Quaternary sediments in the basin outcrops, shale, marl and limestone with shale and marl, sandstone (high- and mid-Asmari) and nummolities of white limestone, clastic conglomerate deposits with medium to low degree of cementation.

### PSIAC model

The PSIAC method estimates total annual sediment yield comprising the sheet and rill erosions (PSIAC 1968). This model was first applied over the watershed of Walnut Gulch in south-east Arizona, United States. Later, considering the modification applied by Johnson & Gebhardt (1982), it was called modified PSIAC (MPSIAC hereafter). The successful applications of this model for estimating the sediment yield of watersheds in semi-arid areas of Iran were reported in several previous studies (Tangestani 2006; Khaledian *et al.* 2012). The method is based on a review of a few representative points within a given



**Figure 1** | Geographical position of the studied area.

sub-catchment, which are then used to project average values for the entire watershed area. The procedure considers nine factors for erosion estimation: surface geology, soils, climate, runoff, topography, ground cover, land use, channel erosion, and upland erosion. The procedure was developed for sub-catchments in the western United States greater than 30 km<sup>2</sup>; however, it has also been applied to smaller basins (Noori *et al.* 2016). Compared with other experimental methods, the MPSIAC model considers the greatest number of factors, so the results are more realistic (Tangestani 2006; Daneshvar & Bagherzadeh 2012). Each factor is subdivided to different categorical classes, and a weighting value is assigned to each class using the model tables based on the degree of impact of each factor class (PSIAC 1968). Calibration in MPSIAC is a process of parameter adjustment (automatic or manual), until catchment and model behavior show a sufficiently (to be specified by the hydrologist) high degree of similarity.

Finally, in the MPSIAC model, the erosion severity and the annual sediment yield are estimated based on the total sum of the nine aforementioned factors, which are expressed by  $R$ . In order to control the accuracy of the interpolations and extrapolations of erosion-factor weights, Equation (1) is applied. This equation evaluates the relationship between the rate of sediment yield in each catchment area unit ( $Q_s$ ) (m<sup>3</sup>/km<sup>2</sup>/year), and the total weights of causal factors ( $R$ ),

$$Q_s = 38.77 \times e^{0.0353 R} \quad (1)$$

### Topography

Topography is one the principal factors affecting the soil erosion, which has the following sub-criteria: slope, vector, size and shape of the basin. The erosion rate directly changes with any change in the length, steepness and the

shape of the slope, while apart from the size, the shape of the basin is also important in erosion formation. In addition, five different classes of slope theme expressed in percentages are constructed in ArcGIS 10.2. These classes are defined as flat and gentle <11%, medium 11–24%, steep 24–38%, very steep 38–57%, and extreme >57%.

### Climate

Climatic factors include rainfall and temperature. Duration and density of the rain directly affects the erosion. A high temperature increases fragmentation of organic substances, which will result in a decline in the plant cover and increase in the erosion rate.

### Runoff

The runoff factor in the watershed was calculated as 0.29 of total average runoff and the peak special discharge. The Soil Conservation Service Curve Number (SCS-CN) model was used to estimate runoff in this area. This model has a long, fruitful application history and is generally referred to as 'blue collar' hydrology (Hawkins et al. 2010).

### Channel erosion

This factor indicates the rate of erosion from river and drainage channels. The slope steepness, type of bedrock, and the potential energy of floods are the major factors affecting the channel erosion.

### Upland erosion

Upland erosion was obtained based on the method suggested by Bureau of Land Management (BLM) (Johnson & Gebhardt 1982).

### Land use and land cover

The surface of the land has a layer that protects it from erosion. As this layer weakens, the risk of erosion increases. The following are determined as the sub-criteria for this factor: (a) plant cover – the land may be covered by forest or an agricultural plant; (b) land use – land can be used for a dense agricultural application, which can harm the soil.

Figure 2 displays the lithology and land use map of the studied region. The study area in this research presents

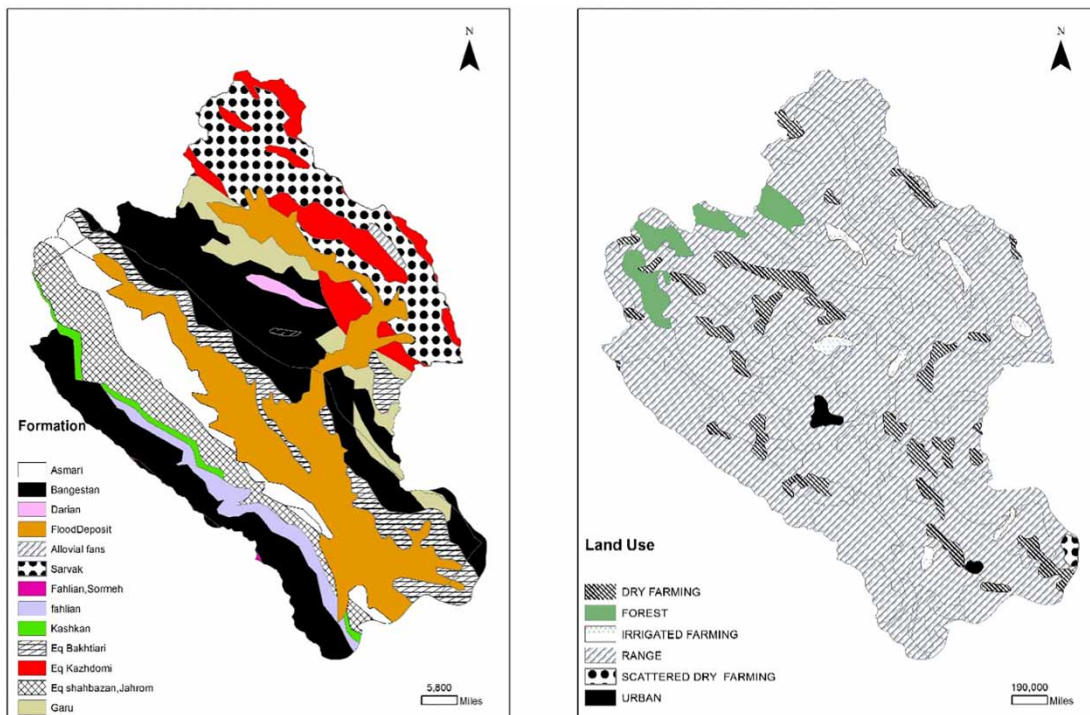


Figure 2 | Lithology map (left) and land use map (right).

various agricultural and industrial activities and infrastructures (see Figure 3). The high traditional agricultural vocation, and use of wrong farming techniques and heavy grazing, is marked in the figure. The industrial activities in the region include the extraction of aggregates in the riverbed in the western area.

### Geology map

In terms of geomorphology, as well as a glimpse of land in the area, two types may be identified: a mountain in the north, east and west basin lands, and plains in the center and south of the watershed. The area has a high diversity of opinion of geology and the lithological units (Table 1).

Using the percentage cover, the area, and the range of annual sediment yield, the mean sediment yield of the studied area was computed using the PSIAC model (Table 2). The dominant erosion potential categories are low to moderate degrees (80.9% of total area), while the areas with very high erosion potential cover only 3.3% of the sub-catchment area. Results of the model also show a sediment yield range of  $954.67 \text{ m}^3/\text{km}^2/\text{year}$  for the very

low erosion potential category, and  $862,748.42 \text{ m}^3/\text{km}^2/\text{year}$  in the regions with very high erosion potential (Figure 4).

It is noted that the MPSIAC method is based on a review of a few representative points within a given sub-catchment, which are then used to project average values for the entire sub-catchment area (Tangestani 2006). So, the obtained erosion values belong to the average of the erosion values of a polygon. Moreover, the region was divided into polygons based on nine factors affecting erosion in the MPSIAC model. To begin with, all factors were combined and the final MPSIAC score was calculated on the map. Then, the borders between homogeneous areas that have equal MPSIAC points were removed in order to produce final polygons. The dissolve tool in GIS is designed for performing this action.

### ANN-MLP model

The ANN model has a similar structure to the human neural network system, where it imitates the structure of the human brain and its operational programs and conducts predictions based on repeated trainings. The structure of an ANN model



Figure 3 | Industrial activity (left) and traditional farming (right).

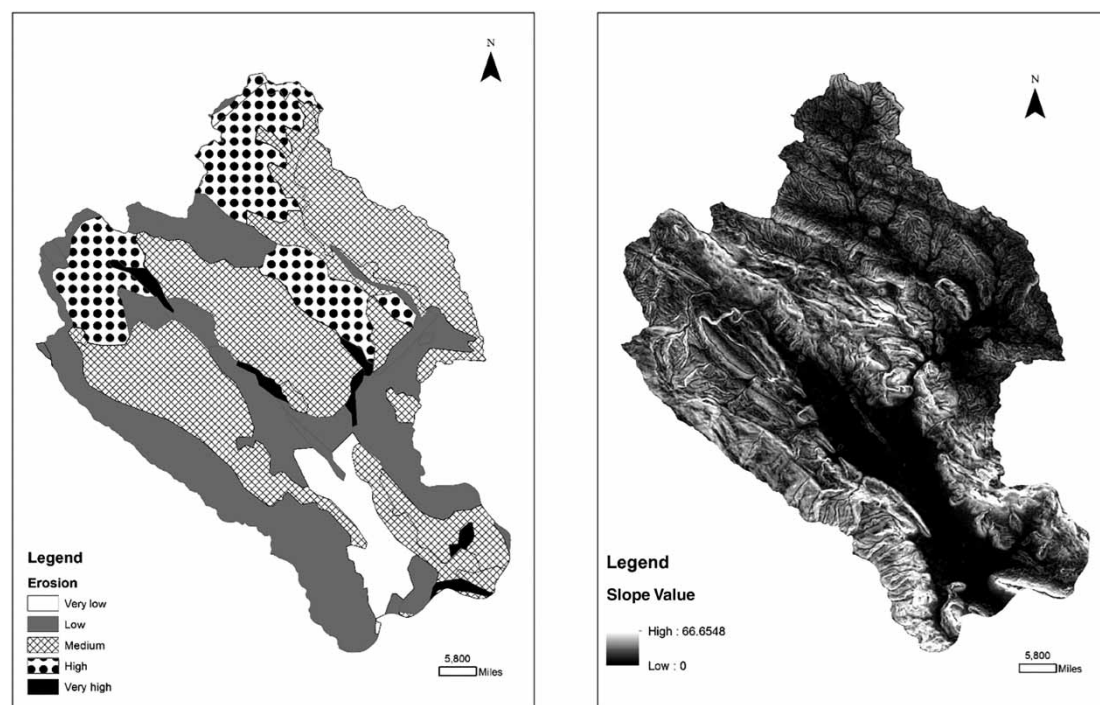
**Table 1** | Geological units of the study area

Row	Age	Lithology	Formation
1	OM(1,2,3)	Gray marl and limestone rocks with layers and layers of sand	Asmari
2	K	Limestone	Bangestan
3	K5	Limestone with orbitolinids	Darian
4	Q	Quaternary	Flood deposits
5	QT	Quaternary	Deposits of alluvial fans
6	K1	Thin-bedded limestone with shale and marl	Sarvak
7	JK	Lime and shale	Surmeh
8	K8	Lime thin layer of cream to brown	Fahlian
9	E	Sandstone and conglomerate and mudstone and chert with radiolarites	Kashkan
10	P1	Conglomerate and sandstone	Equivalent Bakhtiari
11	K2	Red conglomerate and sandstone	Equivalent Kazhdumi
12	K7	Marl and limestone with Ammonites, Orbitolina	Garu

**Table 2** | Sediment yield of study area

Sediment yield ( $m^3/km^2$ )	Mean sediment yield	Area (ha)
954.67	862,748.42	903.71

is flexible, so, generally, its input parameters are selected from the available data. They are composed of neurons, which are arranged in groups called layers and connected through weights. In the input layer, the neurons in this first layer propagate the weighted data and bias randomly selected through

**Figure 4** | Erosion map (left) and slope map (right).

the hidden layers. Once the net sum at a hidden node is determined, an output response is provided at the node using a transfer function. Each neuron receives many inputs from other neurons through weighted connections. These weighted inputs are further added up and produce the argument for a transfer function such as a linear, logistic or hyperbolic tangent function which in turn produces the final output of the neuron (Talebizadeh et al. 2009). Further theoretical information about ANNs can be found in, for example, Bishop (1995) and Haykin (1999). As ANN does not require detailed information about the physical governing rules of the phenomena, it might be effectively employed for modeling complex hydrological processes. In this study, the multilayer perceptron (MLP) algorithm was applied. The MLP model is a flexible type of ANN composed of one input layer, one or more hidden layers, and one output layer (Rai et al. 2005). The MLP is a network formed by simple neurons called perceptrons. The perceptron calculates a single output from multiple real-valued inputs by forming compounds of linear relationships according to input weights and even nonlinear transfer functions. Using only one hidden layer is recommended because using more layers worsens the problem of local minima (Rai et al. 2005).

For all training algorithms, the tangent sigmoid transfer function was used in the hidden layers, and purelin transfer function in the output layer. In this study, different back-propagation algorithms including Levenberg-Marquardt (lm), gradient descent (gd), gradient descent with adaptive learning rate (gda), gradient descent with momentum and adaptive learning rate back-propagation (gdx), and scaled conjugate gradient (scg) have been utilized for erosion prediction. The optimal numbers of neurons were determined by trial and error procedure. At each training process, 100 networks were examined and the optimum structure of each case (transfer functions) was chosen. The minimum and maximum values of weight decay in the hidden layer were found to be 0.0001 and 0.002.

### NF model

NF systems, which are based on rule-based fuzzy systems, use the capability of neural networks' learning algorithm for adapting their rule-base parameters (Jang 1993). NF uses neural network learning algorithms and fuzzy reasoning to

map an input to an output space. The fuzzy decision rules are implemented as membership functions (MFs) and the model has the advantages of both neural networks and fuzzy control systems. There are numbers of methods expressed for partitioning the input space. More details about NF can be found in Jang (1993) and Jang et al. (1997). There are different methods to optimize the NF parameters, e.g. grid partitioning (GP) and sub-clustering (SC) methods. In this study, both the GP and SC methods were utilized. The MF may take different types, e.g. the difference between two sigmoidal (dsig), product of two sigmoidal (psig) and generalized bell (gbellmf), which were evaluated in the present study. For each MF, different values of functions including 2, 3 and 4 were tried, and finally the best function type and its numbers were selected based on the lowest root mean square error (RMSE) (Russell & Campbell 1996). The sub-model parameter sensitivity analysis on the basis of changes in the range of 0 to 1 was conducted.

### Performance assessment indices

Three different data-driven methods, e.g. ANN, NF-GP, and NF-SC, were applied for estimating erosion. The erosion magnitude obtained by the standard MPSIAC model was used as target for the modeling process. The study region was divided into 30 hydrological units and the values of land use, soil, slope, and current erosion data (models' inputs) were calculated using GIS. Seventy percent of the total data were used for training the models randomly, and the remaining 30% were reserved for testing. Random selection of the train-test set would decrease the risk of models' overfitting, as discussed by Roshangar et al. (2014). The models were evaluated using the RMSE (root mean square error), determination coefficient ( $R^2$ ), VAF (variance accounted for), and BIAS statistics as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - x_i)^2}{N}} \quad (2)$$

$$R^2 = \frac{(\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}))^2}{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2} \quad (3)$$

$$VAF = 1 - \frac{Var(x_i - y_i)}{Var(x_i)} \quad (4)$$

$$BIAS = \frac{\sum_{i=1}^N (x_i - y_i)}{N} \quad (5)$$

where  $N$  is the number of the data set,  $y_i$  and  $x_i$  denote the erosion rate values produced by various models and the standard MPSIAC model, respectively,  $\bar{x}$  stands for the average erosion rate values produced by the benchmark MPSIAC model,  $\bar{y}$  presents the average simulated values, and  $Var$  denotes the variance magnitudes.

## APPLICATION AND RESULTS

The first step in establishing the NF- and ANN-based soil erosion models with different influential factors is the selection of the independent variables. Table 3 sums up the

correlation values between the soil erosion and some affecting parameters. From the table, it is seen that the slope and plant are the most influential parameters on soil erosion followed by the land, lithology, current erosion value and soil. An adequate value of the linear cross correlation for an accurate simulation must be higher than 0.6, as stated by Bechrakis & Sparis (2004). However, to avoid system complexity and instability, selection of the minimum input parameters to produce an accurate estimate of soil erosion was attempted. So, based on multi-collinearity between the input parameters, the model inputs were selected as land use, soil, slope, and current erosion data.

The comparison of different ANN training algorithms is presented in Table 4. In this table, 4-2-1 indicates an ANN model comprising 4 input, 2 hidden and 1 output nodes. The number of the hidden nodes was determined iteratively. For each ANN model, the hidden node number increased

**Table 3** | Correlation matrix of the variables

Variables	Value erosion	Lithology	Soil	Current erosion	Plant	Slope	Land
Value erosion	1	0.919	0.782	0.915	0.988	0.990	0.980
Lithology		1	0.602	0.726	0.899	0.913	0.921
Soil			1	0.887	0.776	0.763	0.758
Current erosion				1	0.900	0.896	0.871
Plant					1	0.991	0.992
Slope						1	0.991
Land							1

**Table 4** | Statistical parameters of different ANN training algorithms during the train and test period

Training algorithm	Model structure	Iteration number	Training period				Test period			
			RMSE	R <sup>2</sup>	BIAS	VAF	RMSE	R <sup>2</sup>	BIAS	VAF
Levenberg-Marquardt	4-9-1	1,000	0.001	0.989	-0.001	7.42 × 10 <sup>-9</sup>	6.048	0.741	1.131	0.428
Gradient descent	4-10-1	50,000	0.387	0.980	0.001	0.0008	6.450	0.546	-1.980	0.457
Gradient descent with adaptive learning rate	4-1-1	50,000	1.143	0.992	-0.029	0.008	4.510	0.920	3.328	0.112
Gradient descent with momentum	4-10-1	50,000	0.363	0.990	0.002	0.0007	7.100	0.422	-1.594	0.579
Gradient descent with momentum and adaptive learning rate	4-1-1	50,000	<b>1.119</b>	0.992	-0.0002	0.007	<b>4.262</b>	0.932	-1.804	0.180
Scaled conjugate gradient	4-1-1	1,000	1.110	0.993	0.002	0.007	5.724	0.933	-2.985	0.289

Figures in bold indicate the superior models.



from 1 to 10 and the optimum were chosen based on *RMSE* criterion. Training of the lm, gd, gda, gdm, gdx and scg algorithms was stopped after 1,000, 50,000, 50,000, 50,000,

50,000 and 1,000 iterations, respectively, because the increase in accuracy was too small after these epochs. The tangent sigmoid and linear activation functions were applied

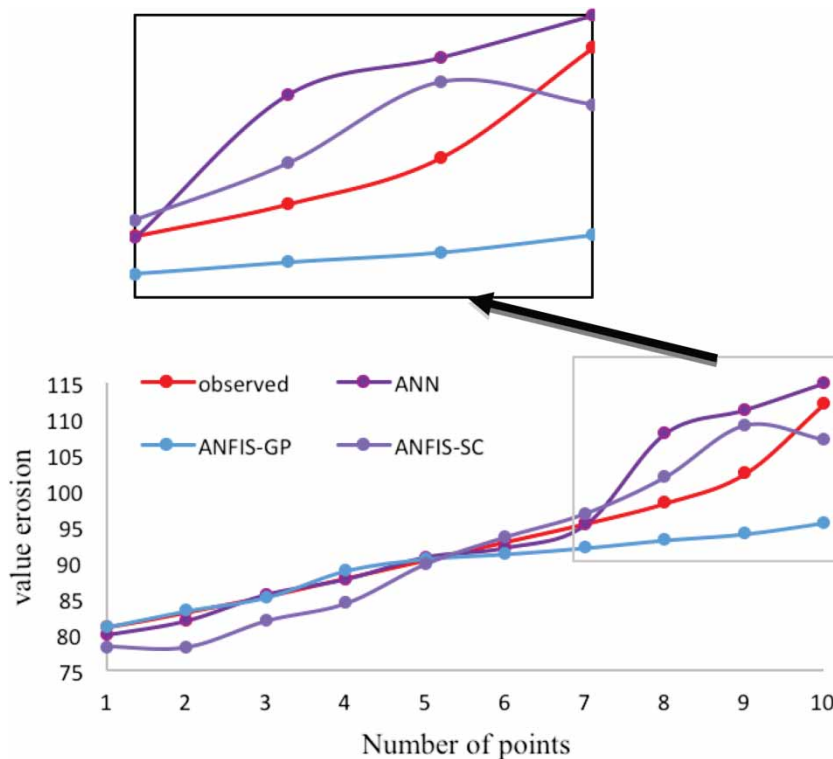
**Table 5** | Statistical parameters of the ANFIS-GP structures during the train and test periods

Type MFs	Model structure	Training period				Test period			
		<i>RMSE</i>	$R^2$	<i>BIAS</i>	<i>VAF</i>	<i>RMSE</i>	$R^2$	<i>BIAS</i>	<i>VAF</i>
Psig	4-2-3-4	0.0008	0.998	0.0005	$3.27e-09$	10.708	0.858	7.322	0.740
Dsig	2-3-3-2	0.072	0.997	$6.44e-05$	$2.97e-05$	13.560	0.467	9.524	1.130
Generalized bell (gbellmf)	2-2-2-3	<b>0.030</b>	<b>0.998</b>	<b><math>7.13e-05</math></b>	<b><math>5e-06</math></b>	<b>6.259</b>	<b>0.837</b>	<b>3.440</b>	<b>0.331</b>

Figures in bold indicate superiority of the Generalized bell function

**Table 6** | Error statistics of the optimal ANFIS-GP, ANFIS-SC, and ANN models

Model	Training period				Test period			
	<i>RMSE</i>	$R^2$	<i>BIAS</i>	<i>VAF</i>	<i>RMSE</i>	$R^2$	<i>BIAS</i>	<i>VAF</i>
ANN	1.119	0.992	-0.0002	0.007	4.262	0.932	-1.804	0.180
ANFIS-GP	0.030	0.998	$7.13e-05$	$5e-06$	6.259	0.837	3.440	0.331
ANFIS-SC	0.0006	0.998	$6.18e-08$	$2.77e-11$	3.775	0.898	0.816	0.165



**Figure 5** | Test results of the optimal ANN, ANFIS-GP and ANFIS-SC models.

for the hidden and output nodes, respectively. It is clear from Table 4 that gradient descent with momentum and adaptive learning rate back-propagation algorithm present more accurate results than the other algorithms in the test period.

Different NF-GP models possessing different types of MFs are compared in Table 5. In this table, 4-2-2-2 indicates an NF-GP model with 4, 2, 2 and 2 MFs for the land use, soil, upland erosion and slope inputs. The NF-GP method proposes independent blocks of each antecedent variable via defining the MFs of all antecedent variables. Fuzzy MFs could take different forms, and the best number of MFs is selected by trial and error. In choosing the number of MFs, large numbers of MFs or parameters should be avoided to save time and computational costs (Kisi & Shiri 2012). So, two or three numbers of MFs were used in the applied NF models. It is observed from the table that

the NF-GP model with generalized bell MFs has the best accuracy based on  $RMSE$  and  $R^2$  values.

Train and test results of the optimal ANN and NF models are given in Table 6. It is obvious from the table that the NF-SC model has the lowest  $RMSE$  and the highest  $R^2$  for both training and test stages. Figure 5 shows the erosion values for different applied models in comparison to the target values. The graph clearly shows the NF-SC's superiority to the NF-GP and ANN models. Figure 6 illustrates the scatterplots of the observed vs. simulated soil erosion using ANN, NF-GP and NF-SC methods for the test stage. From this figure, it is clear that all applied models can simulate soil erosion with high accuracies. Analyzing the regression line equation ( $y = ax + b$ ) shows that the  $a$  and  $b$  values are closer to 1 and 0, respectively, which demonstrates the capability of the model in simulating soil erosion. The observed vs. predicted erosion values of the best models are presented in Figure 7

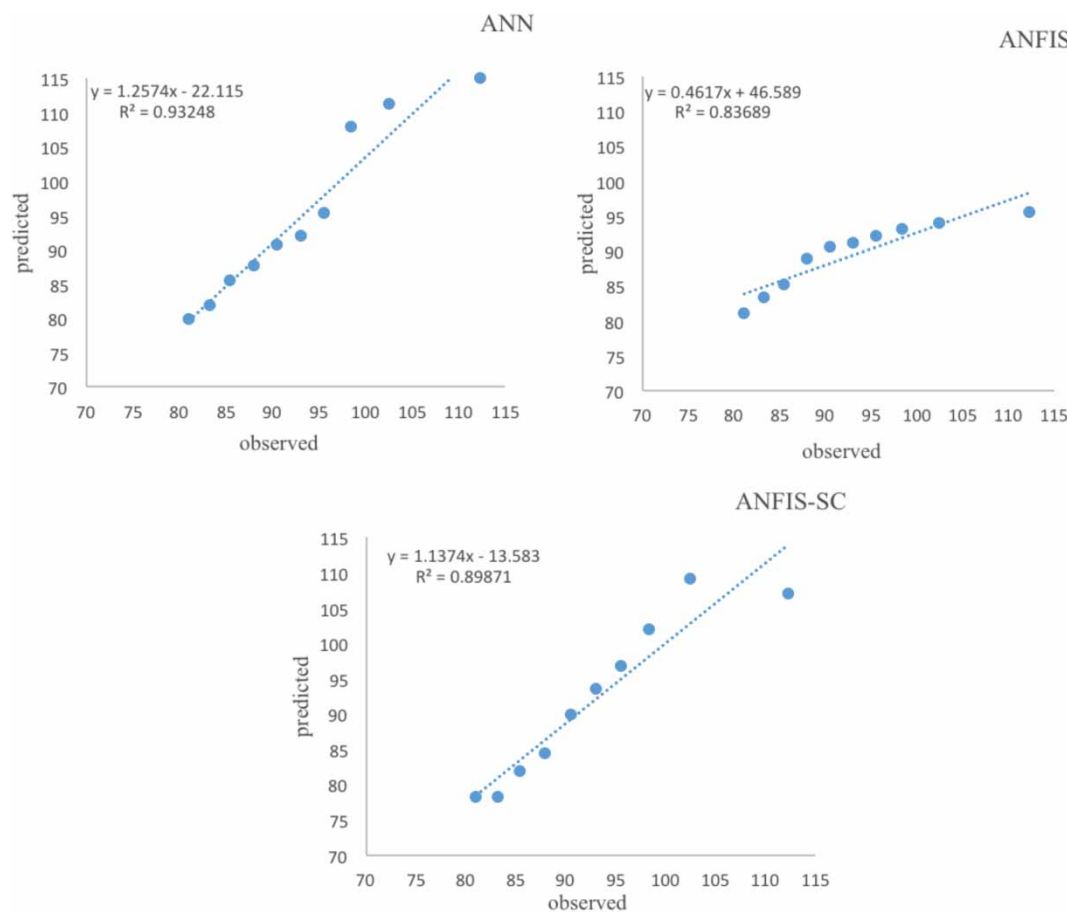
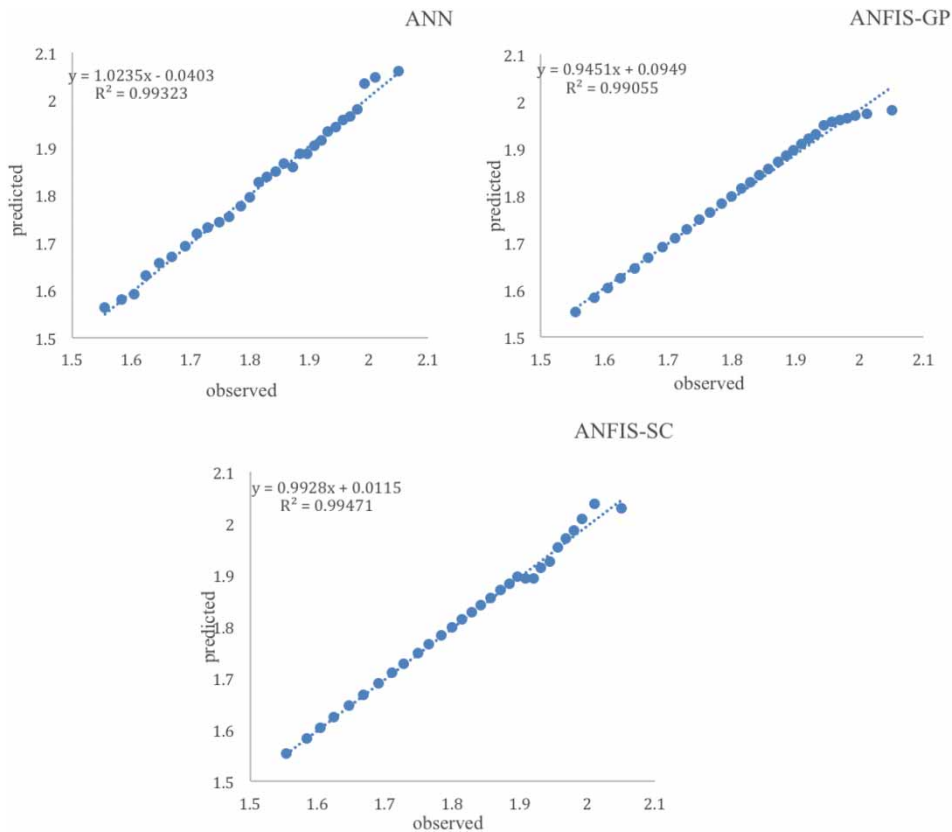


Figure 6 | Predicted vs. observed erosion rate values during the test stage.



**Figure 7** | Relationships between the predicted and measured logarithm erosion rate values (using all available patterns).

(plotted as double logarithmic for better representation) for the whole data (comprising train and test patterns). It is clearly seen from the scatterplots that the estimates of the NF-SC model are closer to the exact fit line than those of the NF-GP and ANN models, especially for the peak values. As can be seen from the figures, there are some scatters between the observed and simulated erosion values in the scatterplots for larger erosion amounts, while the models show good fit with the corresponding observed values in the rest of the points. Although this might be considered as a potential weakness of the applied models in reproducing larger erosion values by taking into account the input variables, a possible reason for such over and under estimations might be the complex behavior of the erosion at higher magnitudes, which needs the involvement of further (possibly not considered here) input variables for simulation.

In order to assess the sensitivity of the applied model to each input variable, the NF-SC and ANN models were established using single-input configurations, and the

corresponding results are summarized in Table 7 for the best ANN and NF models. From the table it is seen that the single-input ANN and NF-SC models that comprise the land slope as the sole input variable produce the most accurate results, while the models relying on soil as an input variable produce the highest RMSE values, which depict the lowest performance accuracy. This might be explained through comparing the correlation values presented in Table 3, where the slope shows the highest correlation values of 0.990 (positive; increasing effect), while the soil shows the lowest correlation values of 0.782 (positive; increasing effect) with erosion. Although both parameters show the positive correlations with increasing effect on soil erosion, the magnitude of the linear correlation is the highest for slope. Nevertheless, it should be noted that the interrelations between the independent and target parameters are usually nonlinear, and taking into consideration the linear relations might produce partially valid conclusions. However, the outcomes produced by

**Table 7** | Error statistics of the single input models for the test data

Input	ANN				ANFIS-SC			
	RMSE	R <sup>2</sup>	BIAS	VAF (%)	RMSE	R <sup>2</sup>	BIAS	VAF (%)
Lithology	9.587	0.594	7.486	56.25	6.991	0.592	0.9453	41.84
Soil	23.947	0.607	23.199	57.30	24.165	0.605	23.462	59.394
Current erosion	19.109	0.607	15.086	66.74	20.163	0.838	19.532	69.654
Plant	13.943	0.418	10.326	9.23	8.788	0.970	-6.328	54.925
Slope	3.428	0.911	0.9778	86.91	3.367	0.915	0.8107	87.05
Land	5.477	0.650	-0.256	63.73	12.124	0.938	-7.0179	18.473

the ANN and NF models confirmed the relations presented by the correlation analysis. Moreover, the higher performance accuracy of the NF model might be linked to its capability in using both the fuzzy inference system and neural network algorithm, which make it easy to simulate complex phenomena such as soil erosion.

Summarizing, it could be stated that, when relying on suitable input configuration, both the NF and ANN models display ability for mapping the interrelations between soil erosion and its influential parameters. By using these models, one can simulate the erosion magnitudes using limited input variables, which would be of great interest for practical issues. Further studies are needed for strengthening these conclusions using data from other catchments and using different models.

## CONCLUSIONS

In this study, the ability of three data-driven methods, ANN, NF-GP and NF-SC, were investigated to predict the erosion rate using geographical input data. Land use, soil, slope, and current erosion were used as input parameters for training and testing the applied models. The erosion values produced by MPSIAC model were used as benchmark patterns. Different training algorithms were used for ANN models and gradient descent with momentum and adaptive learning rate back-propagation algorithm was found to be better than the other algorithms. Various MFs were also tried in NF-GP models and generalized bell MFs gave the best accuracy. The NF-SC model generally performed better than the other models. According to results, data-driven approaches

such as ANN and NF can be a good alternative for the standard PSIAC model. Summarizing, the data-driven models can perform as well as MPSIAC even if they use only part of the information used as input in the benchmark model. This alternative is not only economically beneficial for studies but can also be used in areas with data scarcity. In the present study, a 70% training, 30% testing hold-out strategy was applied for developing and testing the models, which is common in similar studies. Nevertheless, the conclusions obtained through this strategy might be limited and need the robust k-fold testing strategy. Moreover, similar studies might be carried out using data from other regions with similar erosion characteristics to make generalized soft computing-based erosion models. These could be a subject for future studies.

## REFERENCES

- Agarwal, A., Mishra, S. K., Ram, S. & Singh, J. K. 2006 [Simulation of runoff and sediment yield using artificial neural networks](#). *Biosyst. Eng.* **94**, 597–613.
- Alemayehu, F., Taha, N., Nyssen, J., Girma, A. & Zenebe, A. 2009 [The impacts of watershed management on land use and land cover dynamics in Eastern Tigray \(Ethiopia\)](#). *Resour. Conserv. Recycl.* **53**, 192–198.
- Bagherzadeh, M. 1993 *A Study on the Efficiency of Erosion Potential and Sediment Yield Models Using Remote Sensing and Geographic Information Systems*. Unpublished MSc Thesis, Natural Resources College, University of Tarbiat Moedares, Tehran, Iran.
- Bechrakis, D. A. & Sparis, P. D. 2004 [Correlation of wind speed between neighboring measuring stations](#). *IEEE Trans. Energy Convers.* **19**, 400–406.
- Besaltpour, A. A., Ayoubi, S., Hajabbasi, M. A., Mosaddeghi, M. R. & Schulin, R. 2013 [Estimating wet soil aggregate](#)

- stability from easily available properties in a highly mountainous watershed. *Journal of Catena* **111**, 72–79.
- Bishop, C. M. 1995 *Neural Networks for Pattern Recognition*. Oxford University Press, Oxford, UK, 504 pp.
- Bođri, L. & Cermak, V. 2000 Prediction of extreme precipitation using a neural network: application to summer flood in Moravia. *Adv. Eng.* **31**, 311–321.
- Boix-Fayos, C., Martínez-Mena, M., Calvo-Cases, A., Castillo, V. & Albadalejo, J. 2005 Concise review of inter-rill erosion studies in SE Spain (Alicante and Murcia): erosion rates and progress of knowledge from the 1980s. *Land Degrad. Dev.* **16**, 517–528.
- Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., Saby, N., Gobin, A., Vacca, A., Quinton, J., Auerwald, K., Klik, A., Kwaad, F. J. P. M., Raclot, D., Ionita, I., Rejman, J., Rousseva, S., Muxart, T., Roxo, M. J. & Dostal, T. 2010 Rates and spatial variations of soil erosion in Europe: a study based on erosion plot data. *Geomorphology* **122**, 167–177.
- Chang, F. J., Chen, L. & Chang, L. C. 2005 Optimizing the reservoir operating rule curves by genetic algorithms. *Hydrological Processes* **19** (11), 2277–2289.
- Clark, K. B. 2001 An estimate of sediment yield for two small sub catchments in a geographic information system. PhD Thesis, University of New Mexico.
- Daneshvar, M. & Bagherzadeh, A. 2012 Evaluation of sediment yield in PSIAC and MPSIAC models by using GIS at Toroq Watershed, Northeast of Iran. *Frontiers of Earth Science* **6**, 83–94.
- Darghouth, S., Ward, C., Gambarelli, G., Styger, E. & Roux, J. 2008 *Watershed Management Approaches, Policies, and Operations: Lessons for Scaling up*. Water Sector Board Discussion Paper Series, Paper No. 11, The World Bank, Washington, DC.
- Demirel, T. & Tüzün, S. 2011 Multi Criteria Evaluation of the Methods for Preventing Soil Erosion Using Fuzzy ANP: The Case of Turkey. In: *Proceedings of the World Congress on Engineering* (Vol. 2).
- Dhamge, N. R., Atmapoojya, S. L. & Kadu, M. S. 2012 Genetic algorithm driven ANN model for runoff estimation. *Procedia Technology* **6**, 501–508.
- Dibike, Y. B. & Solomatine, D. P. 2001 River flow forecasting using artificial neural networks. *Phys. Chem. Earth (B)* **26**, 1–7.
- Firat, M. & Gungor, M. 2004 Askı Maddesi Konsantrasyonu ve Miktarının Yapay Sinir Ağları ile Belirlenmesi. *IMO Teknik Dergi* **15**, 3267–3282 (in Turkish).
- Hawkins, R. H., Ward, T. J., Woodward, D. E. & Van Mullem, J. A. 2010 Continuing evolution of rainfall-runoff and the curve number precedent. In: *2nd Joint Federal Interagency Conference*, Las Vegas, NV.
- Haykin, S. 1999 *Neural Networks: A Comprehensive Foundation*. Prentice-Hall, Upper Saddle River, NJ, 842 pp.
- He, J., Valeo, C., Chu, A. & Neumann, N. 2011 Prediction of event-based stormwater runoff quantity and quality by ANNs developed using PMI-based input selection. *Journal of Hydrology* **400**, 10–23.
- Hsu, K.-L., Gupta, H. V. & Sorooshian, S. 1995 Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.* **31**, 2517–2530.
- Jang, J. R. 1993 ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man, Cybernet* **23**, 665–685.
- Jang, J. S. R., Sun, C. T. & Mizutani, E. 1997 *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Prentice-Hall, New Jersey.
- Johnson, C. & Gebhardt, K. 1982 Predicting sediment yields from sagebrush rangelands. In: *Proceedings of the Workshop of Erosion and Sediment Yield on Rangelands*, Tuscon, Arizona, 7–9 March 1981.
- Khaledian, Y., Kiani, F. & Ebrahimi, S. 2012 The effect of land use change on soil and water quality in northern Iran. *Journal of Mountain Science* **9**, 798–816.
- Kim, M. & Gilley, J. E. 2008 Artificial Neural Network estimation of soil erosion and nutrient concentrations in runoff from land application areas. *Computers and Electronics in Agriculture* **64**, 268–275.
- Kisi, Ö. 2005 Suspended sediment estimation using neuro-fuzzy and neural network approaches. *Hydrol. Sci. J.* **50**, 683–696.
- Kisi, O. & Shiri, J. 2012 River suspended sediment estimation by climatic variables implication: comparative study among soft computing techniques. *Computers & Geosciences* **43**, 73–82.
- Metternicht, G. & Gonzalez, S. 2005 FUERO: foundations of a fuzzy exploratory model for soil erosion hazard prediction. *Environmental Modelling & Software* **20** (6), 715–728.
- Nagy, H. M., Watanabe, K. A. N. D. & Hirano, M. 2002 Prediction of sediment load concentration in rivers using artificial neural network model. *Journal of Hydraulic Engineering* **128** (6), 588–595.
- Noori, H., Siadatmousavi, S. M. & Mojaradi, B. 2016 Assessment of sediment yield using RS and GIS at two sub-basins of Dez Watershed, Iran. *International Soil and Water Conservation Research* **4** (3), 199–206.
- Pandey, A., Himanshu, S. K., Mishra, S.K. & Singh, V. P. 2016 Physically based soil erosion and sediment yield models revisited. *Catena* **147**, 595–620.
- Peter, H. B. C., Chandler, J. H. & Armstrong, A. 2010 Applying close range digital photogrammetry in soil erosion studies. *Photogramm. Rec.* **25**, 240–265.
- Polyakov, V. O. & Lal, R. 2008 Soil organic matter and CO<sub>2</sub> emission as affected by water erosion on field run-off plots. *Geoderma* **143**, 216–222.
- PSIAC 1968 *Report of the Water Management Subcommittee on Factors Affecting Sediment Yield in the Pacific Southwest Area and Selection and Evaluation of Measures for Reduction of Erosion and Sediment Yield*. ASCE 98 Report No. HY12.
- Rai, P., Majumdar, G. C., DasGupta, S. & De, S. 2005 Prediction of the viscosity of clarified fruit juice using artificial neural network: a combined effect of concentration and temperature. *J. Food Eng.* **68**, 527–533.

- Roshangar, K., Vojoudi, F. & Shiri, J. 2014 Modeling river total bed material load discharge using artificial intelligence approaches (based on conceptual inputs). *Journal of Hydrology* **514**, 114–122.
- Russell, S. O. & Campbell, P. F. 1996 Reservoir operating rules with fuzzy programming. *Journal of Water Resources Planning and Management* **122**, 165–170.
- Sadeghi, H. 1993 Comparison of some erosion potential and sediment yield assessment models in Ozon-Dareh sub-catchment. In: *Proceedings of the National Conference on Land Use Planning*, Tehran, Iran.
- Singh, R. D., Lohani, A. K. & Kumar, R. 2012 Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *Journal of Hydrology* **442**, 23–35.
- Talebizadeh, M., Morid, S., Ayyoubzadeh, S. A. & Ghasemzadeh, M. 2009 Uncertainty analysis in sediment load modeling using ANN and SWAT model. *Water Resour. Manage.* **24**, 1747–1761.
- Tangestani, M. H. 2006 Comparison of EPM and PSIAC models in GIS for erosion and sediment yield assessment in a semi-arid environment: Afzar Catchment, Fars Province, Iran. *Journal of Asian Earth Sciences* **27**, 585–597.
- Tarolli, P. & Sofia, G. 2016 Human topographic signatures and derived geomorphic processes across landscapes. *Geomorphology* **255**, 140–161.
- Vanwallegem, T., Amate, J. I., de Molina, M. G., Fernández, D. S. & Gómez, J. A. 2011 Quantifying the effect of historical soil management on soil erosion rates in Mediterranean olive orchards. *Agric. Ecosyst. Environ.* **142**, 341–351.

First received 7 March 2017; accepted in revised form 8 May 2017. Available online 11 August 2017