INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print) 2222-5234 (Online)
http://www.innspub.net
Vol. 3, No. 9, p. 116-122, 2013

RESEARCH PAPER OPEN ACCESS

Modeling sediment yield using artificial neural network and multiple linear regression methods

Lida Eisazadeh[1*], Reza Sokouti[2], Mehdi Homaee[3], Ebrahim Pazira[4]

[1]*Department of Soil Science, Science and Research Branch, Islamic Azad University, Tehran, Iran*

[2]*Research centers of agricultural and natural resources of West Azerbaijan, Urmia, Iran*

[3]*Department of Soil Science, Tarbiat Modares University, Tehran, Iran*

[4]*Department of Soil Science, Science and Research Branch, Islamic Azad University, Tehran, Iran*

Key words: Sediment estimation, regression, artificial neural network, sediment-retention reservoirs.

doi: http://dx.doi.org/10.12692/ijb/3.9.116-122 Article published on September 15, 2013

Abstract

Estimating sediment yield in upstream sub basins of reservoirs is an important issue for designing and operation of water resources structures. In classical methods of predicting sediment yield (e. g. regression models) internal uncertainties are not explicitly taken into consideration. However this model cannot improve understanding the internal relationships between the data extracted and cannot determines the impact of each factor of sediment yield. The use of artificial neural networks modeling for prediction and forecasting variables in sedimentation are easier, cheaper and they begin to solve nonlinear problems. In this study, 25 sub basin of reservoir in West Azerbaijan province, Iran, were selected for estimating sediment yield by using multiple linear regression (MLR) and artificial neural network (ANN) methods. Therefore, 160 data sets of sediment yield have been used in selected sub basins of reservoirs. In ANN method, different combinations of inputs and different kinds of functions were designed with the best model by error back propagation algorithm. Also, in the MLR method, a model established by using different parameters of climatic and geomorphological factors. Some statistics including RMSE and R^2 were used to evaluate the performance of applied models. The results indicated the proposed ANN model could well predict the sediment yield with R^2 = 0. 86 and RMSE= 0.09 in comparison to the MLR model which it's R^2 and RMSE are 0.64 and 1.41 respectively. In particular, the ANN model had the capability of discovering non-linear relationships of sedimentation using geomorphologic parameters with reasonable precision.

* **Corresponding Author:** Lida Eisazadeh ✉ lidaeisazadeh@gmail.com

Introduction

Today, the soil conservation is one of the development fundamentals in the world, while the main problems resulted from the soil erosion and more specifically the water and soil loss by extra runoff is not considered efficiently. On the other hand, awareness of the amount and severity of erosion and sediment yield for efficient and optimized strategies of preventing the soil losses have been attended by researchers to be able to predict the sediment location and erosion based on them. In order to implement the sediment yield estimation programs, and also the precise calculating and designing the dam's capacity in building the reservoir dams, the simulation and evaluation of the amount of produced sediment behind the sediment-retention dams is necessary. Different approaches have been used so far to estimate the sediment amount in small basins and sediment-retention dams. Using the simulators, physical models, mathematical models, statistical distributions and computer based models are of those cases used for predicting and modeling the hydrologic phenomena. The empirical and regression approaches are common extensively for having the simple structure, simple mathematical calculations and the ability of working with the limited input data set. Thus, the linear or non-linear common regression models can present the sediment loads with relative precision of the simple empirical-calculative models' accuracy (Morgan, 1996).

During the recent decade, the non-linear mathematical model has been added to the prediction tools with the name of artificial neural network. This model is used much more in engineering sciences. In this issues, selecting the models closing to the reality existed in the system is so difficult. The artificial neural network is one of the simulating models which can cover the existed realities with a relative appropriate accuracy and its prediction could be more close to the reality. The artificial neural network is like the human brain which with analyzing the conditions, limitations and the parameters working in a process, makes the optimized decision and repeats it in a way that the coefficients of each parameter has been balanced and accommodated with the subject's reality in a close way (Tayfur and Guldal, 2006).

Sarangi and Bhattacharya (2005) have shown the superiority of the artificial neural network over the regression approaches in sediment and erosion prediction in a watershed area in India with the coefficient of determination 0.98. Zu *et al.* (2007) with studying on the Langchuang Jing River in China estimated the suspension load successfully by using the artificial neural network and presented the designed climatic network advantage over the designed network based on the daily statistics of the flow. Cigizoglu and Alp (2007) proved the superiority of the artificial neural network over the regression approach based on the speed and accuracy with comparing the regression approach and neural networks in estimating the sediment of the river located in the United States. In a research by Kerem *et al.* (2006) used a specific separation algorithm for dividing the data series by the homogenous amounts in order to enhance the neural network models results and minimizing the error estimation model and obtained acceptable results. Agrowal *et al.* (2006) calibrated the runoff and sediment provided model using the neural network in a watershed area and stated that the artificial neural network has a high capability for modeling the runoff-sediment processes. Mirbagheri and Rajaei (2005) estimated the suspend load of the Zohreh River for estimating the suspend load of the river using the Perceptron multi-layer network and radius-based network. They compared the obtained result from the model by the sediment rating curve approach and evaluated the neural network model successful specially in predicting the suspended load in the upstream discharges. Yusefvand *et al.* (2005) have considered the regression modeling based on the dry and rainy months obtaining the minimum mean square errors as a useful approach based on the sediment transferring measurement of the Gharesu River in Kermanshah.

Only a few studies (Tayfur, 2002; Nagy, 2006 and Cigizoglu 2006) focused on sediment yield modeling and sediment concentration. The aim of this study is the modeling sediment yield using artificial neural network and multiple linear regression Methods in the 25 sub basins of reservoirs in West Azerbaijan, northwest of Iran

Material and methods

Study area

In this research, 25 sub basins upstream of reservoirs have been selected in West Azarbaijan province, Iran. All reservoirs have been extended in different locations of the Province and these extensions have had lithological, climatic and physiographical diversities. For estimating amount sediment yield in selected reservoirs all the factors including the height of the structure from the bed bottom, the length of the structure' crest, and the deposition length along the channel were surveyed. Figure1 has shown the distribution of dams in West Azerbaijan province, Iran.

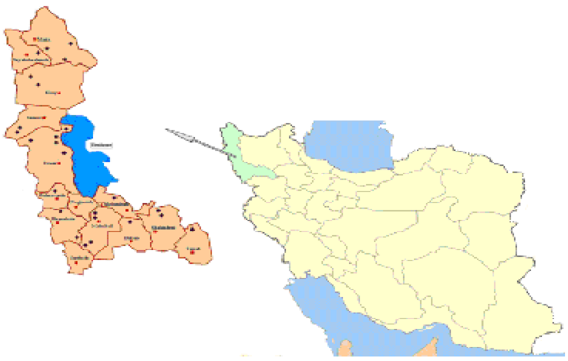

Fig. 1. Distribution of dams in West Azerbaijan province, Iran.

Since working with Auger was so difficult and impossible practically because of the grain sediment, for being sure of measurement accuracy of sediment yield and the structures aspects, digging profile was done for 160 spots randomly and the accuracy of the work was controlled. Finally, the amount of the sediment yield was calculated of each structure. The characteristic of each of these areas are illustrated individually in Table 1.

Multiple linear regression method

The main purpose of MLR in its general form is to find the relationship between the dependent and independent variables. In the current research the following relation has been used.

$$Y_i = \sum_{i=1}^{i=6} c_i x_i + c_7 \qquad \text{(Eq. 1)}$$

In which X_1 to X_6 are the drainage area, basin perimeter, basin length, channel slope, mean annual precipitation, and lithology factor respectively. C_i are the regression coefficients of these variables, also y_i is the especial sediment yield (m3/ha/yr). For modeling, MATLAB mathematical tool has been used for obtaining the C coefficient with the MLR approach.

Artificial neural network

ANNs are massively parallel systems composed of many processing elements connected by links of variable weights. One of the first models which made use of the McCulloch and Pitts (1943) model of a neuron was a neural network called the perceptron (Rosenblatt 1958). The neurons used in the perceptron have a simple summation input function and a hard-limited threshold activation function or linear threshold activation function. The input values are in general real numbers and the outputs are binary. To overcome the linear separability limitation of the perceptron's, MLPs were introduced. An MLP consists of an input layer, at least one intermediate or "hidden" layer, and one output layer. An example of a multilayered network is given in figure 2. The neurons from each layer being fully connected to the neurons from the next layer. The MLP was put into practice only when learning algorithms was developed for them. At every learning cycle the training algorithm consists of two passes: (1) a forward pass, when inputs are supplied and propagated through the intermediate layers to the output layer; and (2) a backward pass, when an error is calculated at the outputs and propagated backward for calculating the weights' changes. Calculating the error and changing the connection weights can be done either in a batch mode or in an individual mode. In this research, MLP network with the error back propagation algorithm was used. At the first step, available physiographic data were normalized in the range of 0 to 1 because the sigmoid activation function used for training the network had lower and upper limits of 0 and 1, respectively. In modeling the artificial neural network, data has been divided into two training and testing groups. So, 70% and 30% of the data dedicated to training and testing respectively. The equation 2 was adopted to normalize the data set.

$$X_{i\,norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \qquad \text{(Eq. 2)}$$

Where, x_i is the original values of different selected inputs, $x_{i\ norm}$ is the normalized value; x_{max} and x_{min} are the maximum and minimum values respectively.

The performances evaluation criteria in multiple linear regression and ANN methods were the root mean square errors (RMSE) and the coefficient of determination (R^2) expressed between estimated and observed annual sediment yield as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Y_i - Y_e)^2} \qquad \text{(Eq. 3)}$$

$$R^2 = \qquad \text{(Eq. 4)}$$

Where y_i and y_e are the observed (actual) and estimated values of y respectively, $\bar{y}$ is the mean of the observed values of y and N is the number of observations.

Results and discussion

Multiple linear Regression model

By using correlational matrix technique, the relationship between annual sediment yield and physiographic parameters were determined. The obtained results indicated that two variables of drainage area and lithology factor in MPSIAC model were selected as sensitive parameters for generation of sediment yield; these parameters were included in the development of the multiple linear regression model. Figure 5 presents a scatter plot between observed and computed sediment load using linear regression method, also C coefficients have been shown in Table 2.

As it is seen in the figure the MLR model does not have a high accuracy which its cause can be known as the dominant non-linear feature of the natural phenomena including the sedimentation in dams' reservoirs. The amount of RMSE and R^2 in this method is equal to 1.41and 0.64 respectively.

The artificial neural network

Identifying the architecture of the used ANN for modeling annual sediment yield process is primary and important aspect of the modeling. . In this study some parts of network such as number of hidden layers (one layer), activation function (tansig) and number of output layer neurons (just one), were supposed as the constant and on the other hand, some other parts, such as number of the neurons in input and hidden layers and number of training epochs, are calculated as dynamic parameters which must be optimized through a trial-error process. Number of hidden layer's neurons varies from 5 to 20.

Table 1 presents R^2 and RMSE for training and R^2 for validation stages of different ANN structures. Considering these results 7.20.1 structure with the input data of drainage Area, basin perimeter, channel slope, Mean annual precipitation, Lithology factor, and basin's slope are the suitable one in comparison with other used ANN structure, that leads to considerable calibration and verification R^2 and RMSE. To prove the efficiency of selected ANN model, the obtained results was compared with the results of multiple linear regression model. The comparison of the observed and computed sediment load in the best structure in the stage of artificial neural network testing (48 removed data) is shown in figure (4). Also, Table 3 illustrates different structures of the artificial neural network. As it is seen, the results of the neural network have a high accuracy and can be used as a useful tool for prediction.

Table 1. The Sub basins physiographic features of West Azerbaijan Province, Iran.

rows	micro catchments	Drainage Area(ha)	basin perimeter(m)	Basin Length(m)	channel slope	Mean annual precipitation	Lithology factor
1	Ghaziabad(1)	644	12486	4008	11.3	376.4	5
2	Ghaziabad(2)	202	7202	2284	12	376.4	4
3	Emam kandi	170	6630	2284	20.9	346.7	6
4	Rabat	689.2	1330	5690	7.02	632	6
5	Kulij	758	11377	3025	6.3	350	4
6	Ashtarmol	353	11895	4951	10.9	640	5
7	Silveh	79	5570	3350	11.1	344	3
8	Khure	114.7	5082	1923	10.9	418	3
9	West of reyhanloo	132	5479	1970	13.7	370.2	8
10	south of reyhanloo	143	6698	2415	4.9	370.2	9
11	West of gharaaghaj	197	8638	3726	12.3	355	4
12	north of gharaaghaj	124.2	4785	1024	13.7	355	6.5
13	Guleh guleh(1)	128.4	7665	2880	10.4	352	3.5
14	Guleh guleh(2)	123.3	7735	3069	9.4	352	5
5	Alucheloo(1)	93.1	5085	1921	10.4	335	7
16	Alucheloo(2)	48	3691	1114	13.9	335	7
17	Alucheloo(3)	63.5	3921	888	5.6	632	5.5
18	khalian	817.5	13211	5077	7.3	360	3.5
19	taramesh	64.4	4232	2081	14.4	389.6	6
20	Fishel gharaghuyn	177.2	6719	2707	14.4	380	6.5
21	South of zaviebala	81.94	4722	1560	12.9	324	4
22	North of zaviebala	29	3034	1238	13.7	324	8.5
23	Ghalat mangoor	570	10680	4057	16.2	564	9
24	Gardane ghushchi	55.9	3570	1630	6.1	301	5
25	halabi	1214	17456	6015	12.6	376.4	5

Table 2. C coefficients calculated by multiple linear regressions.

Coefficients	C_1	C_2	C_3	C_4	C_5	C_6
Value	-0.0039	0.0004	0.0001	-0.2736	0.0267	-0.0045

Table 3. Different structures of the artificial neural network.

Neural Network Structure	Training		Validation
	R^2	RMSE	R^2
7-5-1	0.62	0.14	0.35
7-10-1	0.68	0.13	0.49
7-15-1	0.80	0.10	0.68
7-20-1	0.86	0.09	0.73

Using available data of sub basins up streams of reservoirs, the network architecture that yielded the best results in terms of determination coefficient and root error mean square on the training and verification steps, determined thorough trial and error steps.

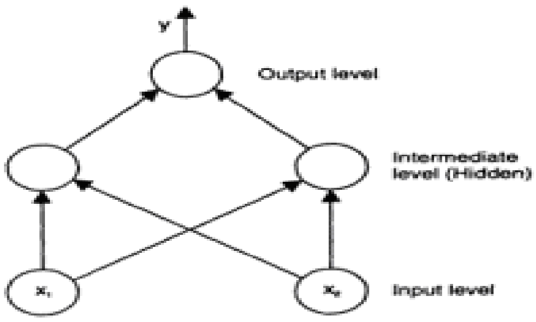

Fig. 2. Two layer neural networks in MLP model.

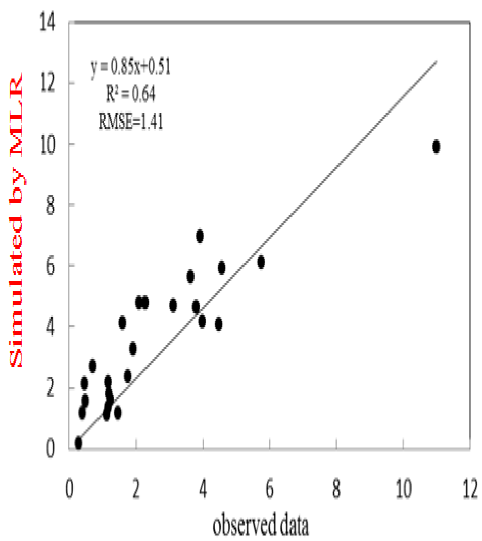

Fig. 3. Distribution the estimated values of the sediment deposited in the regression model.

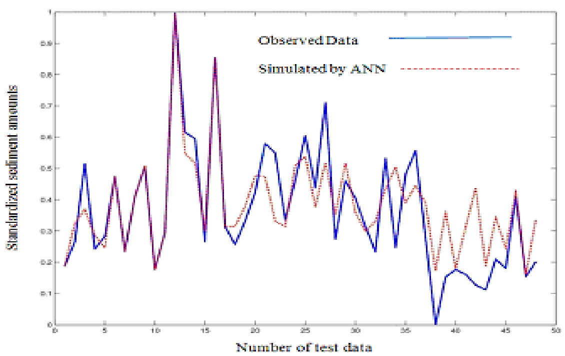

Fig. 4. Distribution of observed and simulated sediment yield of ANN with the optimal structure 7-20-1.

Conclusion

In this study, the artificial neural network and multiple linear regression models were applied to estimate annual sediment yield in the 25 sub basins up streams of reservoirs in West Azarbaijan province, Iran. The obtained results from comparing the two approaches indicated that the artificial neural network has a higher accuracy in comparison with the multiple linear regression method for estimating the annual sediment load. The cause of this fact is the extensive and parallel processing system and considering the non-linear transformations. In the multiple linear regression approaches using the correlational matrix two variables of drainage area and lithology factor in MPSIAC model have been selected for estimating the annual sediment and finally the obtained relation with the R^2 and RMSE of 0.64 and 1.41 respectively. The obtained results from the artificial neural network implementing with error back propagation algorithm indicated that the best amount of estimated sediment with 7 inputs and 20 neurons in the hidden layer with the the RMSE amounts equal to 0.09 and $R^2 = 0.86$ were obtained. Also, the obtained results from applying the different functions indicated that the moving functions tansig as the threshold function of the artificial neural network has been more appropriate than other functions. Since the neural network has less sensitivity to the input, using this network is superior to other regression models. The results of this study are consistent with the results of Agrowal *et al* (2006), Kerem (2006) and Sarnegi *et al* (2005) studies. The mentioned researchers have concluded that the artificial neural network approach has a higher accuracy than the linear regression model and also stated that using the effective geomorphological parameters in sediment yield in the ANN as the model input causes the accuracy increasing in estimating the sediment in sub basins.

Thus, regarding this issue that the sediment yield phenomenon is a non-linear process, ANN is more reliable than the conventional regression method to monitor sediment load over the river. However the ANN modeling techniques thorough this research can be replicated over other sub basin and watershed systems to account for hydrological parameters. Efforts should be made to associate morphological parameters thorough different mathematical functions to develop geomorphological association functions leading to more accurate prediction of sediment losses.

Acknowledgement

The author would like to express their sincere thanks to Dr Fahmi and Dr Massih Abadi for important role in the development and implementation of this ongoing research.

References

Agrowal A, Singh RD, Mishra S K, Bhung PK. 2006. ANN based sediment yield models for Vamsadhara River Basin (India). Water SA **31,** 95-100.

Alborzi M. 2010. Introduction of neural networks. Sharif university press. 207 p.

Cigizoglu HK, Alp M. 2007. Generalized regression neural network in modeling river sediment yield. Advance in engineering software **37,** 63-68. http://dx.doi.org/10.1080/10286600500126256

Cigizoglu HK, Kisi O. 2006. Methods to improve the neural network performance in suspended sediment estimation", J. of Hydrology **317,** 221-238. http://dx.doi.org/10.1016/j.jhydrol.2005.05.019

Kerem H. 2006. Generalized regression neural network in modeling river sediment yield. Advances in Engineering Software **37,** 63–68.

Mirbagheri SA, Rajaee T. 2005. Using artificial neural network to estimate bed load sediment in Zohre River. Proceeding of first international congress of civil engineering. Sharif University, Tehran. (In Persian).

Morgan RPC. 1996. Soil erosion and conservation. Second Edition. 198 p.

Nagy HM, Watanabe K, Hirano M. 2002. Prediction of sediment load concentration in rivers using artificial neural network model. Journal of Hydraulic Engineering ASCE **128(6),** 588–595. http://dx.doi.org/10.1061/(ASCE)0733-9429(2002)128:6(588)

Sarangi A, Bhattacharya AK. 2005. Comparison of artificial neural network and regression models for sediment loss prediction from Banha watershed in India. Agricultural Water Management **78,** 195-208. http://dx.doi.org/10.1016/j.agwat.2005.02.001

Tayfur G, Guldal V. 2006. Artificial neural networks for estimating daily total suspended sediment in natural streams, Nordic Hydrology **37,** 69-79.

Youssefvand F, Ghelmaei S H, Ghamarnia H, Zia tabarahmadi MK. 2005. The effect of time measuring classification of suspended sediment of rivers. 5th Hydraulic conference. Kerman, Iran. (In Persian).

Zhu YM, Lu XX, Zhou Y. 2007. Suspended sediment flux modeling with artificial neural network: An example of the Longchuanjiang River in the Upper Yangtze Catchment, China. Geomorphology **84,** 111-125.