Sigma J Eng & Nat Sci 38 (2), 2020, 703-714

Sigma Journal of Engineering and Natural Sciences
Sigma Mühendislik ve Fen Bilimleri Dergisi

Research Article

SUSPENDED SEDIMENT LOAD PREDICTION IN RIVERS BY USING HEURISTIC REGRESSION AND HYBRID ARTIFICIAL INTELLIGENCE MODELS

Banu YILMAZ[1], Egemen ARAS*[2], Murat KANKAL[3], Sinan NACAR[4]

[1]*Karadeniz Technical University, Dept. of Civil Engineering, TRABZON;* ORCID: 0000-0002-4671-6384
[2]*Bursa Technical University, Department of Civil Engineering, BURSA;* ORCID: 0000-0002-7553-9313
[3]*Bursa Uludağ University, Department of Civil Engineering, BURSA;* ORCID: 0000-0003-0897-4742
[4]*Karadeniz Technical University, Dept. of Civil Engineering, TRABZON;* ORCID: 0000-0003-2497-5032

Received: 08.10.2019 Revised: 05.03.2020 Accepted: 24.04.2020

ABSTRACT

Accurate prediction of amount of sediment load in rivers is extremely important for river hydraulics. The solution of the problem has been become complicated since the explanation of hydraulic phenomenon between the flow and the sediment on the river is dependent many parameters. The usage of different regression methods and artificial intelligence techniques allows the development of predictions as the traditional methods do not give enough accurate results. In this study, data of the flow and suspended sediment load (SSL) obtained from Karşıköy Gauging Station, located on Çoruh River in the north-eastern of Turkey, modelled with different regression methods (multiple regression, multivariate adaptive regression splines) and artificial neural network (ANN) (ANN-back propagation, ANN teaching-learning-based optimization algorithm and ANN-artificial bee colony). When the results were evaluated, it was seen that the models of ANN method were close to each other and gave better results than the regression models. It is concluded that these models of ANN method can be used successfully in estimating the SSL.
Keywords: Artificial intelligence, Çoruh river basin, regression analysis, river hydraulics, suspended sediment load.

1. INTRODUCTION

The sediment from various sources and transported by rivers reduces the capacity of water storage structures. Moreover, it is caused decrease of the value of the land and the infiltration rate of the soil, augmentation of the risk of floods and damage by raising the stream bed, blockage of intake of water structures, decrease of the capacity of irrigation and drainage channels and increase of in maintenance costs, reduce of the amount of dissolved oxygen in water and restrict life in water, prevention of the proliferation of fish by covering fish eggs, increase of the potable water treatment costs , pollution of by transporting various pollutants from agriculture, industry and other sectors, reduce of the recreation of natural and artificial lakes and degradation of environmental aesthetic.

* Corresponding Author: e-mail: egemen.aras @btu.edu.tr, tel: (224) 300 38 47

Although direct measurements suspended sediment load (SSL) in gauging stations the most reliable way to obtain the amount of SSL, this method is a quite time-consuming and costly [1,2]. In Turkey, sediment rating curve expressed by an equation based on regression and which correlates the flow with SSL transported by river was used. The sediment load corresponding to the daily average flow is determined from these rating curves. However, the estimates using these curves have been unsuccessful in Turkey, especially in the planning of reservoir [3]. Sedimentation movement is affected by many random variables. It is quite difficult to explain the non-linear relationship between these natural phenomena and variables. In recent years, the development of soft computing methods has enabled dam designers to use artificial intelligence and heuristic regression techniques to better estimate the amount of SSL. These methods have recently replaced traditional time series techniques, including sediment rating curves (SRC), multiple linear regressions and autoregressive models [4]. For the estimation of SSL, artificial neural networks (ANN) are the primary methods used in recent years. The success of this method in the prediction of SSL in rivers has been demonstrated by many studies [5-12]. In addition, estimation methods such as support vector machines and adaptive neural fuzzy logic modelling [13-16] to investigate the relationship between the flow and SSL have been developed in recent years and successful results were obtained on the studied rivers.

When the recent studies are examined, it is seen that many researchers dwell on genetic algorithm, artificial bee colonies (ABC), teaching-learning-based optimization (TLBO) algorithm and so on meta-heuristic methods. These studies were applied to all kinds of water structures, SSL estimation, water quality, construction, electricity and energy problems and remarkable results were taken [17- 25].

In this study, SSL were estimated by using the data collected from the Karşıköy Gauging Station on the Çoruh River. Two regression and three ANN models were used for the estimation and the results were compared with each other and the observed values. In this way, it is tried to determine the model that best estimates the value of SSL.

2. MATERIALS AND METHODS

2.1. Study Area and Available Data

Even though Çoruh River, the longest river in the Eastern Black Sea Region, is undeveloped economically and socially, it has high economic potential for Turkey owing to the large amounts of hydropower potential. On the Çoruh River Basin, there are a total of 9 large dams, of which 6 (Muratlı, Artvin, Borçka, Deriner, Güllübağ ve Arkun) are in operation, 3 (Yusufeli, Bayram ve Bağlık) of which are under construction. Dam reservoirs in the basin have risks with regards to filling in a short time and decrease of economic lifetime due to severe erosion. Furthermore, the basin is one of the regions where natural disasters such as avalanche and flood are experienced most (Figure 1).

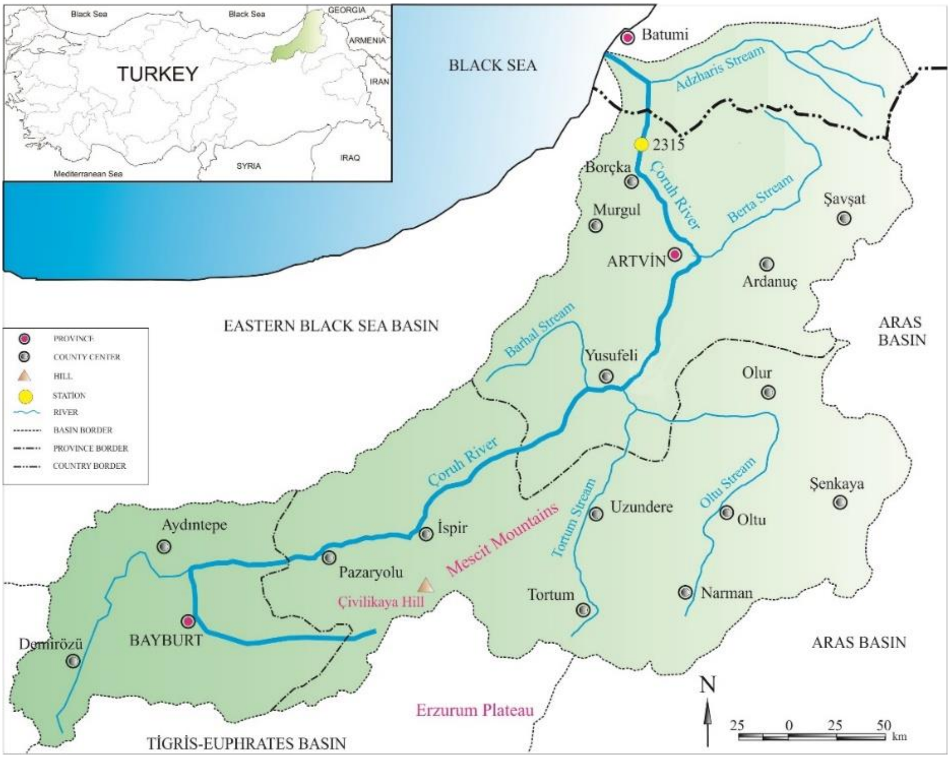


Figure 1. Çoruh river basin and No. 2315 (Karşıköy) gauging station

The data used in the study were taken from the flow observation annual published by General Directorate of Electrical Power Resources Survey and Development Administration (EIE) in 2005. On the Çoruh River, 2315 Karşıköy station (1967-2001) between 41° 42 '38 "E and 41 ° 27'07" N coordinates, 2316 İspir Koprusu station (1984-2001) between 40 ° 57' 40 "E and 40 ° 27'32 "N coordinates, 2320 Laleli Station (1969-2005) between 40° 36' 19" E and 40° 23' 31 "N coordinates, 2322 Altınsu station (1984-2001) between 41° 53' 36" E and 41° 09'47 "N coordinates, 2337 Çamlıkaya station (1999-2005) between at 41° 12' 22 "E and 40° 37'46" N coordinates. The years in parentheses are the observation times of the station.

One of the selection criteria for the station is that the station is located in the main branch rather than the side branches of the river due to the presence of large dams on the main branch. In addition, since ANN and regression-based methods were used in the study, the station which is one of the longest common measurement time of flow and suspended sediment observations was chosen since the data length was thought to affect especially the ANN performance. In the study, Karşıköy station, which has the value of 390 data pairs which measured between 20.06.1967 and 16.07.2002, was used.

Altitude of the station is 57 m and its drainage area is 17906.4 km^2. SSL was measured between 1967 and 2002 at the station. The average flow is 207 m^3/s. SSL was composed of 45.1% sand and 54.9% clay and silt. The average annual SSL measured at the station is 6,851,598 tons and SSL yield in a year is 383 tons/km^2 [26]. Table 1 shows the basic statistical information such as mean (x_{mean}), standard deviation (S_x), skewness (C_{sx}) and coefficient of variation (C_v) of all data belonging to Karşıköy Station and the range of the largest and smallest observation values of these data. The table also demonstrates the ratio of the maximum values of the SSL to the mean values of these. The x_{max}/x_{mean} value is too large to indicate of the complexity of the transportation of SSL. Besides, it is seen that there is a positively skewed distribution in the flow and SSL data when the station values are considered. The amount of skewness is higher in SSL data.

Table 1. The basic statistics of data belonging to Karşıköy Station

Basic Statistics	x_{mean}	S_x	C_x	C_v (S_x/x_{mean})	x_{mak}	x_{min}	x_{mak}/x_{mean}
Q (m^3/s)	209.8	195.7	1.97	0.93	1210.7	37.6	5.77
Q_s (ton/day)	26380.6	55577.8	4.07	2.11	400075.3	148.7	15.17

2.2. Methods

Selection of input variables of the used methods is one of the most significant stages in modelling. Q_{t-1}, Q_t, and S_{t-1} were used as independent variables in previous studies [4, 27] related to the Çoruh River and used similar data. In respect thereof, Q_{t-1} represents the previous day's flow record, while S_{t-1} represents the previous SSL measurement (usually the previous month's SSL measurement). The time lag in this study were limited to the previous measurement, since it has been demonstrated in previous studies [28] that using measurements earlier than these values did not change the accuracy of the models.

There were 390 measured data between 1967 and 2002 belonging to No.2315 Karşıköy Station. Data anomalies which may be caused by measuring errors, missing or incorrect measurements or extreme values were removed from this data. 362 flow and SSL data were used in the models and error values have been calculated based on these data.

Output values obtained from model results were evaluated according to different type of error criteria. Two well-known error criteria were used in order to make a better comparison between the models and evaluate the results more accurately. Correlation of the output values with the actual data and the accuracy of the estimation results were evaluated by the root mean square error (RMSE) and the mean absolute error (MAE). The formulas for these criteria are as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\mathrm{SSL}_{observed} - \mathrm{SSL}_{predicted})^2} \tag{1}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|\mathrm{SSL}_{observed} - \mathrm{SSL}_{predicted}\right| \tag{2}$$

Where, the value of N, $\mathrm{SSL}_{observed}$ and $\mathrm{SSL}_{predicted}$ indicates the number of observations, the actual values and the predicted values, respectively.

2.2.1. Regression Methods

In the study, three different nonlinear regression models were used, namely power, exponential and quadratic function. The equation of quadratic function which produces the best result in these functions is as follows.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1x_2 + b_5x_1x_3 + b_6x_2x_3 + b_7x_1x_1 + b_8x_2x_2 + b_9x_3x_3 \tag{3}$$

In the equations, x_1, x_2, x_3 are the independent variables, $b_0, b_1, b_2, b_3 \ldots b_9$ are the coefficients of regression and y is the calculated value of the independent variables.

Another regression method used in the study was the multivariate adaptive regression splines (MARS) developed by Friedman in 1991. This is a form of nonparametric regression analysis. In most of the applied fields, nonparametric regression methods are used to represent the phenomenon that has no linearity between the variables. The main advantage of this model is that it can explain the complex and non-linear relationship between the independent and the dependent variables [29]. Details of the method can be found in the study of Nacar et al. [30].

2.2.2. Artificial Neural Networks Methods

Three algorithms were used for ANN training. The first one of them is the back-propagation algorithm which is one of the most popular algorithms for most of the engineering problems by virtue of its simplicity and success in the outlook of application. This algorithm is called the back-propagation algorithm because it tries to reduce the errors backwards which means in the direction from output to input. The back-propagation learning algorithm is used to recalculate the weights in each layer based on the current error performance found at the output point of network. Any function has not been used in the input layer while creating the architecture of the network. In the study, single hidden layer architecture has been formed. The tangent sigmoid and linear transfer functions were used in the hidden layer and the output layer, respectively. The process of the method can be found in the publication of Bayram et al. [31].

Besides the back-propagation algorithm, the training of ANN has also been done with the meta-heuristic algorithms ABC and TLBO which have been frequently used in many studies in recent years. The ABC and TLBO algorithms were used to overcome the drawbacks resulting from the back-propagation algorithm and to achieve smaller error values by obtaining more acceptable parameters in the training process of ANN including weight and threshold values. The ABC algorithm contains three control parameters. These are the number of food sources (SN), equal to the number of employed or onlooker bees, colony size (NP), limit value and maximum cycle number (MCN). Control parameters in the TLBO algorithm are the maximum number of iterations (NMI) and the population size (SP). In both algorithms, the training process was applied to the input vector several times and the network has been updated each iteration until it reached a certain stop criterion. The parameters in the algorithm are defined the same for all different models of ANN and are shown in Table 2.

Table 2. Control parameters of ANN-ABC and ANN-TLBO

ANN-ABC Parameters						ANN-TLBO			
NP		SN (NP/2)		MCN	Limit Value	NMI	SP		
50	100	25	50	2000	[-1,1]	2000	50	100	200

3. RESULT AND DISCUSSION

In the MARS and multiple regression analysis (MRA), 80% of the data of Karşıköy Station was used as a training set and 20% of these were utilized as a validation set. 3 different functions (power, exponential and quadratic) were applied in MRA. The quadratic function was applied to the test set as it gave the lowest RMSE values. The equation of quadratic function is as follows.

$$Q_{S_t} = 2955,971 + (-74,507)*Q_{t-1} + (58,677)*Q_{S_{t-1}} + (0,602)*Q_t + (0,027)*Q_t*Q_{t-1} + (4,118*10^{-4})*Q_t*Q_{S_{t-1}} + (-3,166*10^{-4})*Q_{t-1}*Q_{S_{t-1}} + (0,04)*Q_{t-1}*Q_{t-1} + (0,157)*Q_t*Q_t + (-5,569*10^{-6})*Q_{S_{t-1}}*Q_{S_{t-1}} \quad (4)$$

Scattering and time series diagrams belonging to MRA and MARS were shown in Figure 2.

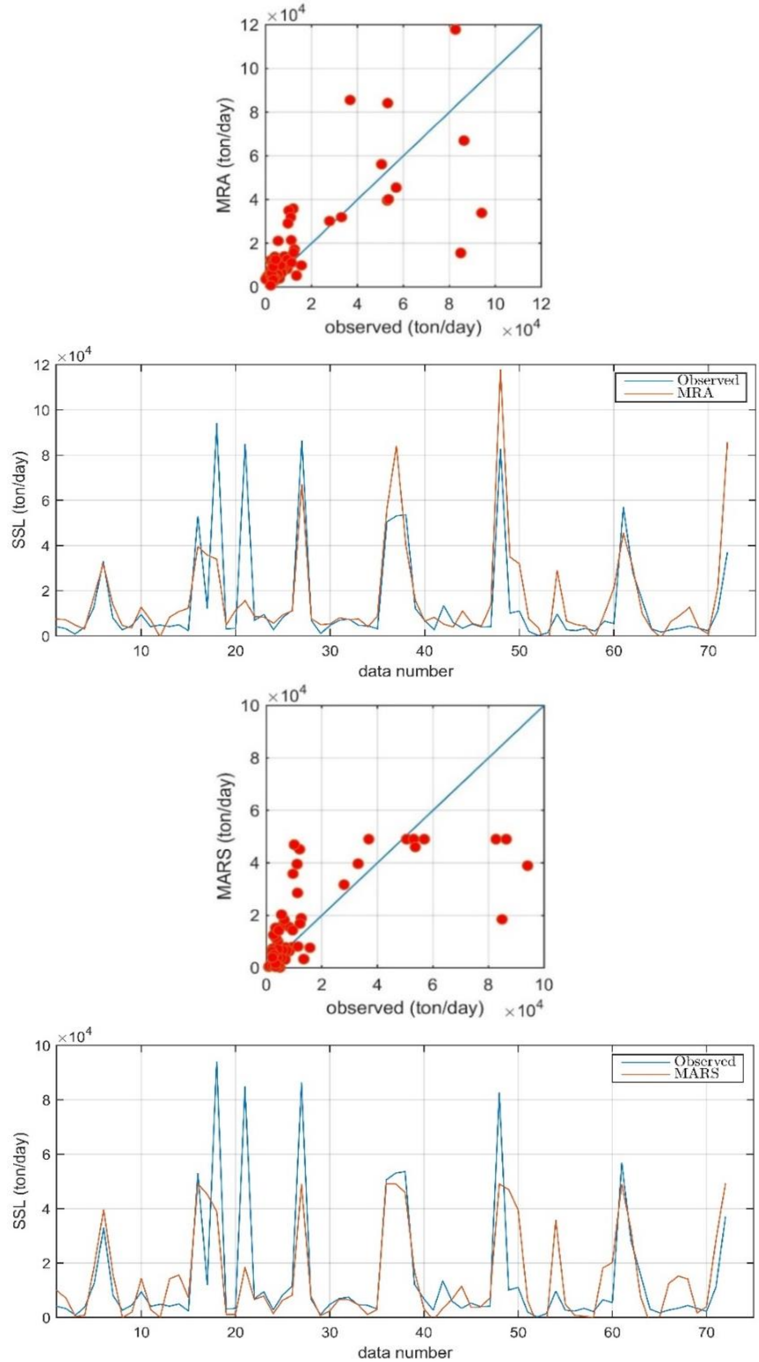


Figure 2. Scatter diagrams and time series of prediction efficiency in the validation period for the regression methods

It was observed that the error values were lower in MARS compared to MRA. The equation of the MARS model and the basic function in the equation are given below.

$$Q_{S_t} = 49068.9 - 155.54 * BF2; \tag{5}$$

$$BF2 = \max(0, 377.784 - Q_t) \tag{6}$$

In ANN models, data were divided into three sets as training, testing and validation. The validation set was selected the same as in the regression models (20%), the remaining 80% data part was divided into 60% training and 20% test and weights of ANN models were calculated with these data.

36 different combinations were created by changing the nodes number of hidden layers (5, 10, 15, 20), learning (0.1, 0.5, 1) and momentum (0.1, 0.5, 1) coefficients in the ANN-back propagation (BP) models. In the ANN-BP model, the best performance was obtained from the ANN architecture where the nodes of hidden layer with 15, the learning coefficient with 0.5 and the momentum coefficient with 0.1. When the results of ANN-BP models were examined, it was seen that the model was more successful than the regression models. The RMSE and MAE values of all models were given in Figure 3.

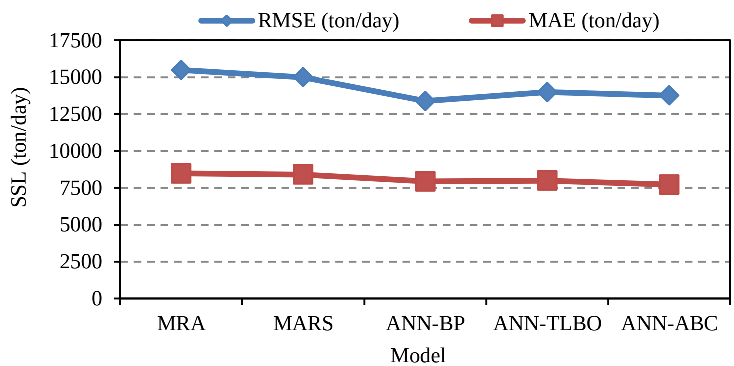
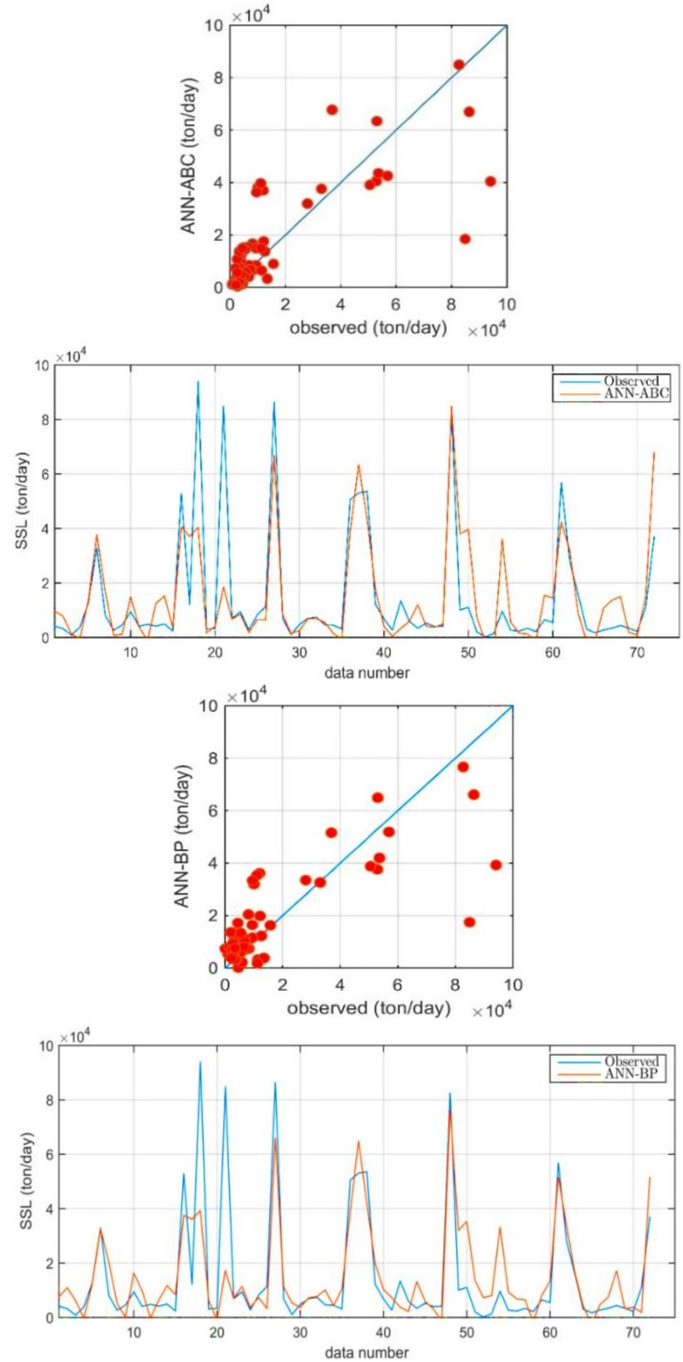


Figure 3. RMSE and MAE values belonging to Karşıköy gauging station

In ANN-ABC model, 16 different analyses were performed by differentiating between the numbers of hidden layer nodes, colony size and limit values. The best results were obtained in the analysis that has 10 hidden layer nodes, 50 colony size and 100 limit values. Scatter diagrams and time series of prediction efficiency in the validation period for the ANN model were shown in Figure 4.


ANN-ABC (ton/day)
observed (ton/day)
Observed
ANN-ABC
SSL (ton/day)
data number
ANN-BP (ton/day)
observed (ton/day)
Observed
ANN-BP
SSL (ton/day)
data number

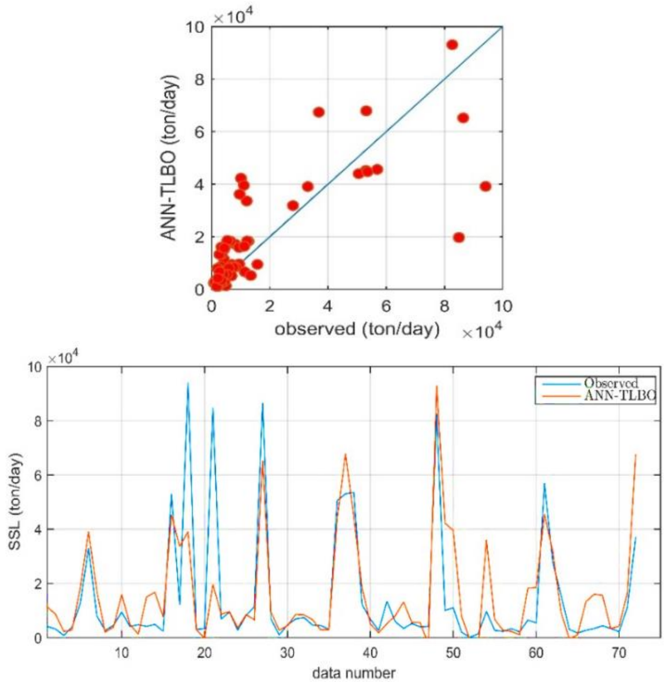


Figure 4. Scatter diagrams and time series of prediction efficiency in the validation period for the ANN Model

The RMSE value obtained from the ANN-ABC model was higher than that of the ANN-BP model, while the MAE value was lower. It is seen time series graph showing all models together in Figure 5. Although it is seen that the models gave similar results to each other, it obvious that the ANN-BP-and ANN-ABC models produce better results when the error values were examined.

As can be seen in Figure 5, ANN based models gave better results in terms of capturing peak values in prediction of suspended sediment load in general. However, almost all models were insufficient in estimating peak data at some points. The probable cause for underestimation of peak values has been attributed to the local variations in the function being mapped due to varying skewness in the data series, and theoretical considerations of the network functioning confirm this [32].

In particular in this basin area with severe sediment erosion, it is also possible to work with these methods in other gauging stations, which have a length of data that can be used in prediction models. Only models that are related to discharge can be used in the sections up to the parts with a significant side branch combination on both the upstream and downstream parts of the main branch and where the slope does not change much.

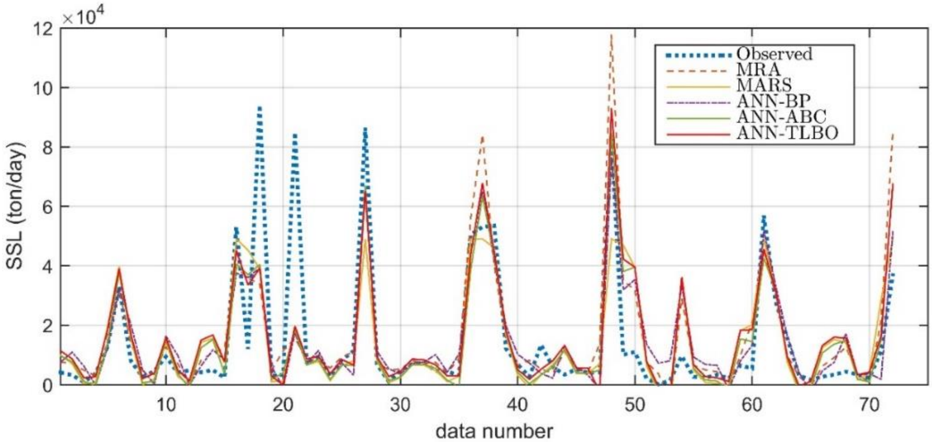


Figure 5. Time series of validation set graph for all models of Karşıköy Station

4. CONCLUSION

In this study, the ability of two regression models with three function and three artificial intelligence models were investigated to estimate the SSL. Current day flow, previous day flow and previous SSL records were used as input data in the models.

In the regression method, MARS model gave better results than MRA; however, accuracy of ANN models was better than regression models. Even though ANN models gave values close to each other, RMSE values of ANN-BP were lower 2.8% from ANN-ABC model and %4.4 from ANN-TLBO model. In the ANN-BP model, the lowest error value was obtained from the ANN architecture where the nodes of hidden layer with 15, the learning coefficient with 0.5 and the momentum coefficient with 0.1.

As a result of this study, it has been observed that artificial intelligence methods gave highly good results in the estimation of SSL. In addition to that, it is considered that it would be useful to collect more data in existing stations and make modelling with a wider data group so that peak values can be better estimated.

Acknowledgments

This study is dedicated to the memory of the late Assoc. Prof. Dr. Murat İhsan KÖMÜRCÜ, who passed away in February 2013.

REFERENCES

[1] Olive, L.J., and Rieger, W.A., (1988) An examination of the role of sampling strategies in the study of the suspended sediment transport. Sediment Budgets, Proceedings of the Porto alegre Symposium. IAHS Publ. No. 174, 259-267.

[2] Öztürk, F., (2002) Determination of Surface Flow and Sediment Amount by AGNPS Model (In Turkish), Scientific Research Project Final Report, Ankara University Scientific Research Projects, Ankara.

[3] Kisi, Ö., Karahan M.E. ve Şen, Z., (2003) Fuzzy logic modeling of suspension material amount in rivers (In Turkish), İTÜ Journal, 2, 34, 43-54.

[4] Yilmaz, B., Aras, E., Nacar, S., Kankal, M., (2018) Estimating suspended sediment load with multivariate adaptive regression spline, teaching-learning based optimization, and artificial bee colony models. Science of The Total Environment, 639, 826-840.

[5] Jain, S. K., (2001) Development of integrated sediment rating curves using ANNs. Journal of Hydraulic Engineering, 127(1):30-37.

[6] Tayfur, G., (2002) Artificial neural networks for sheet sediment transport. Hydrological Sciences Journal, 47(6): 879-892.

[7] Cigizoglu, H. K., (2004) Estimation and forecasting of daily suspended sediment data by multi-layer perceptrons, Advances in Water Resources 27(2):185-195.

[8] Tayfur, G., and Guldal, V., (2006) Artificial neural networks for estimating daily total suspended sediment in natural streams. Hydrology Research 37(1):69-79.

[9] Alp, M., and Cığızoğlu H. K., 2007. Suspended sediment load estimation by two artificial neural network methods using hydrometeorological data. Environmental Modelling & Software, 22, 2-13.

[10] Zhu, Y. M., Lu, X. X., 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(1):111-125.

[11] Ardıclıoglu, M., Kişi, Ö., Haktanır, T., (2007) Suspended sediment prediction using two different feed-forward back-propagation algorithms. Canadian Journal of Civil Engineering 34(1):120-125.

[12] Kisi, Ö., (2008) Constructing neural network sediment estimation models using a data-driven algorithm. Mathematics and Computers in Simulation 79(1):94-103.

[13] Cobaner, M., Unal, B., and Kisi, O., (2009) Suspended sediment concentration estimation by an adaptive neuro-fuzzy and neural network approaches using hydro-meteorological data. Journal of hydrology, 367(1-2), 52-61.

[14] Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M., Nourani, V., (2009) Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. Science of The Total Environment 407(17):4916-4927.

[15] Kisi, O., and Shiri, J., (2012) River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. Computers & Geosciences 43:73-82.

[16] Lafdani, E. K., Nia, A. M., Ahmadi, A., (2013) Daily suspended sediment load prediction using artificial neural networks and support vector machines. Journal of Hydrology 478:50-62.

[17] Altunkaynak, A., (2009) Sediment load prediction by genetic algorithms. Advances in Engineering Software, 40(9), 928-934.

[18] Karaboga, N., (2009) A new design method based on artificial bee colony algorithm for digital IIR filters. Journal of the Franklin Institute 346(4):328-348

[19] Zhang C, Ouyang D, Ning, J (2010) An artificial bee colony approach for clustering. Expert Systems with Applications 37(7):4761-4767

[20] Pour, O. M. R., Shui, L. T., and Dehghani, A. A., (2011) Genetic algorithm model for the relation between flow discharge and suspended sediment load (Gorgan river in Iran). Electronic Journal of Geotechnical Engineering, 16, 539-553.

[21] Uzlu, E., Akpınar, A., Özturk, H. T., Nacar, S., Kankal, M., (2014a) Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey. Energy 69:638-647.

[22] Uzlu, E., Kömürcü, M, İ., Kankal, M., Dede, T., and Öztürk, H. T., (2014b) Prediction of berm geometry using a set of laboratory tests combined with teaching–learning-based optimization and artificial bee colony algorithms. Applied Ocean Research 48:103-113.

[23] Bayram, A., Uzlu, E., Kankal, M., & Dede, T. (2015) Modeling stream dissolved oxygen concentration using teaching–learning based optimization algorithm. Environmental Earth Sciences, 73(10), 6565-6576.
[24] Kankal, M., Uzlu, E., (2017) Neural network approach with teaching–learning-based optimization for modeling and forecasting long-term electric energy demand in Turkey. Neural Computing and Applications 1-11.
[25] Baki, O.T., and Aras, E., (2018) Estimation of BOD in Wastewater Treatment Plant by Using Different ANN Algorithms. Membrane Water Treatment, An International Journal, 9(6), 455-462.
[26] EİE, (2005) General Directorate of Electrical Works, Suspended sediment observation and transport in Turkey's rivers (In Turkish), EİE Press, Ankara.
[27] Buyukyildiz, M., and Kumcu, S. Y., (2017) An estimation of the suspended sediment load using adaptive network based fuzzy inference system, support vector machine and artificial neural network models. Water resources management, 31(4), 1343-1359.
[28] Nacar, S., Hınıs, M. A., & Kankal, M. (2018a) Forecasting Daily Streamflow Discharges Using Various Neural Network Models and Training Algorithms. KSCE Journal of Civil Engineering, 22(9), 3676-3685.
[29] Kisi, O., and Parmar, K. S., (2016) Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. Journal of Hydrology, 534, 104-112.
[30] Nacar, S., Kankal, M., and Hinis, M. A., (2018b) Estimation of Daily Stream Flows with Multivariable Adaptive Regression Curves -Haldizen Creek Case (In Turkish), Gumushane University Journal of the Institute of Science and Technology, 8(1), 38-47.
[31] Bayram, A., Kankal, M., and Önsoy, H., (2012) Estimation of suspended sediment concentration from turbidity measurements using artificial neural networks. Environmental monitoring and assessment, 184(7), 4355-4365.
[32] Sudheer, K. P., Nayak, P. C., & Ramasastri, K. S. (2003). Improving peak flow estimates in artificial neural network river flow models. *Hydrological Processes*, *17*(3), 677-686.