How to cite: Üneş, F., Taşar, B., Varçin, H., Gemici, E. (2022) River Sediment Amounts Prediction with Regression and Support Vector Machine Methods. *2022 "Air and Water – Components of the Environment" Conference Proceedings*, Cluj-Napoca, Romania, p. 98-107, DOI: 10.24193/AWC2022_10.

RIVER SEDIMENT AMOUNTS PREDICTION WITH REGRESSION AND SUPPORT VECTOR MACHINE METHODS

Fatih ÜNEŞ[1] *, Bestami TAŞAR [2], Hakan VARÇİN [3]
Ercan GEMİCİ [4]

DOI: 10.24193/AWC2022_10

ABSTRACT. Accurate estimation of the amount of sediment in rivers; determination of pollution, river transport, determination of dam life, etc. matters are very important. In this study, sediment estimation in the river was made using Interaction Regression (IR), Pure-Quadratic Regression (PQR) and Support Vector machine (SVM) methods. The observation station on the Patapsco River near Catonsville was chosen as the study area. Prediction model was developed by using daily flow and turbidity data between 2015-2018 as input parameters. Models were compared to each other according to three statistical criteria, namely, root mean square errors (RMSE), mean absolute relative error (MAE) and determination coefficient (R^2). These criteria were used to evaluate the performance of the models. When the model results were compared with each other, it was seen that the IR model gave results consistent with the actual measurement results.

Keywords (4-6): Estimation, turbidity, regression, support vector machine, streamflow.

1. INTRODUCTION

Accurate estimation of the amount of sediment carried in rivers is very important in terms of engineering and water structures planning in the planning and projecting of structures built on streams. The amount of sediment in the rivers reduces the life of the facilities built on the river, and also damages the river transport and agricultural areas. Particularly, the sediment accumulating in water storage facilities such as dam reservoirs reduces the reservoir capacity and causes the reservoir to become unable to function over time. In addition, estimation of the amount of sediment transport is very important in determining the amount of scour or accumulation that may occur on the feet of other structures such as viaducts and bridges in the river for flood control, and in terms of taking the necessary precautions.

[1]* Iskenderun Technical University, Iskenderun/Hatay – TURKEY, e-mail: fatih.unes@iste.edu.tr
[2] Iskenderun Techn. University, Iskenderun/Hatay – TURKEY, e-mail: bestami.tasar@iste.edu.tr
[3] Iskenderun Technical University, Iskenderun/Hatay – TURKEY, e-mail: hakan.varcin@iste.edu.tr
[4] Bartin University, Bartin – TURKEY, e-mail: egemici@bartin.edu.tr

Rajaee et al. (2009), using data from Little Black River and Salt River stations in the USA, determined daily sediments concentration using Fuzzy Logic (FL), Artificial Neural Networks (ANN), Multiple Linear Regression (MLR) and Sediment Rating Curve (SRC) methods. They showed that the results estimated by ANN are better than the classical methods. Mirbagheri et al. (2010) evaluated the performance of Sediment Rating Curve (SRC), ANN and fuzzy rule-based models in estimation of sediments concentration in rivers by using coefficient of determination, and compared the results. Fuzzy rule-based model showed better results for sediments concentration estimation they showed. Vafakhah (2013) used precipitation and runoff data from the Kojor basin near the Caspian Sea in Iran for sediments estimation. In their study, they studied 776 days of data for the years 2007-2010. Olyaie et al. (2015) compared ANN, Adaptive Neural Fuzzy Inference System (ANFIS), SRC and Combined Wavelet Artificial Neural Networks (WANN) methods for sediments prediction in a river in the USA. They observed that the WANN and ANFIS methods gave the best results. Kisi and Zounemat (2016) conducted a study to estimate the amount of sediments in 2 stations on the Muddy river in the USA. In the study, they used daily flow rate and sediments concentration data. Shameei and Kaedi (2016), for the estimation of the amount of sediments measured at Rio Valenciano and Quebrada Blanca stations in the USA. They investigated the performance of Linear Genetic Programming (LGP) and Neuro Fuzzy (NF) methods and found that both methods gave appropriate results. Cherif, et al. (2017) estimated the sediment load in rivers during the storm period over a 22-year period in Wadi El Hammam, Northwest Algeria, using the SRC method. Riahi-Madvar and Seifi (2018) applied ANN and ANFIS models to estimate sediment load using different combinations of input parameters. Rahman and Chakrabarty (2020), investigated the success of the ANN method in predicting sediment transport and evaluating morphological changes in an alluvial river. Zounemat-Kermani, et al. (2020) examined machine learning models including ANFIS, support vector regression and hybrid genetic algorithm models (GAANFIS and GA-SVR) for suspended sediment and bed load estimation. Rajaee and Jafari (2020), reviewed the literature on artificial intelligence models for river sediment concentration estimation. They showed that artificial intelligence models can effectively predict the sediment concentration in rivers. In addition, artificial intelligence methods are widely used in studies in many different hydrology fields (Demirci et al., 2016, Üneş et al., 2018, Demirci et al., 2018, Üneş et al., 2019, Üneş et al., 2020).
The aim of this study is to investigate the sediment concentration (SC) changes estimation based on Interaction Regression (IR), Pure-Quadratic Regression (PQR) and Support Vector machine (SVM) models performance.

2. DATA AND METHODS

Within the scope of this study, 915 daily meteorological data between 2015-2018 of the station 01589025 located at 39°15'04.5" North latitude and 76°45'49.6"East longitude on the Patapsco River near Catonsville, USA were used. In the study, 78%

of all data were education; 22% is reserved for testing. 715 days of data were used for training and 200 days of measurement data were used for testing. In these model applications, the amount of sediment (SC) was estimated using the average daily flow (Q), turbidity (T), obtained from the United States Geological Survey (USGS). The location of the studied station of the Patapsco River is shown in Figure 1.

Fig. 1. The location of the studied station- Patapsco River(USGS)

In this paper, Interaction Regression (IR), Pure-Quadratic Regression (PQR) and Support Vector machine (SVM) methods were used to obtain SC predictions.

2.1. Interaction Regressions (IR)

In Multi-variate regression methods, the effect of independent variables on dependent variable was expressed by the regression coefficients. These coefficients in the regression equation expresses the degree of dependence of the independent variable to the dependent variable.

Interaction multivariate regressions (IR) result model includes constant, linear, and interaction terms. The general form of the IR, method was given in equations 1.

$$Y_i = (B_0 + B_1X_1 + B_2X_2 + \ldots + B_nX_n + B_{n+1} * X_1 * X_2 + \cdots) + \varepsilon_i \qquad (1)$$

In the equation, "Y" refers to the dependent variable, "X" independent variable, "B" regression coefficients and "ε" error component.

2.2. Pure-Quadratic Regression (PQR)

Pure-quadratic multivariate regressions (PQR) model includes constant, linear, and squared terms. The general form of the PQR method was given in equations 2.

$$Y_i = (B_0 + B_1 X_1 + B_2 X_2 + \cdots + B_n X_n + B_{n+1} * X_1 * X_2 + B_{n+1} X_1^2 + B_{n+2} X_2^2 + .. + B_m X_n^2) + \varepsilon_i \quad (2)$$

In the equation, "Y" refers to the dependent variable, "X" independent variable, "B" regression coefficients and "ε" error component.

2.3. Support Vector Machines Method

The SVM has become a relatively novel and promising estimator in data-driven research fields, of which basic concept and theory have been introduced by Vapnik (1998). The generalization ability of the SVM is considered to be better than ANN, in the sense that it is based on the structural risk minimization rather than the empirical risk minimization of ANN. The main process of SVM model building consists of selecting support vectors which support the model structure and determining their weights. Fig. 2 shows the general SVM schematic representation.

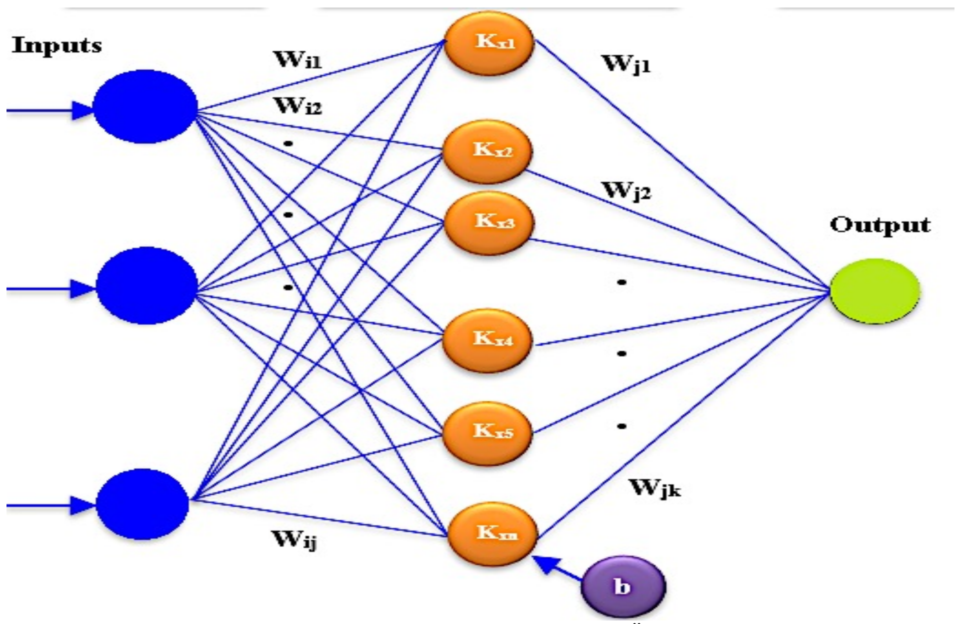

Fig. 2. Schematic structure of SVM model (Üneş et al., 2021)

SVM provides to define how to draw this boundary between variables group. In SVM, the Kernel method greatly increases machine learning in nonlinear data. The process of an SVM estimator (y) can be expressed as :

$$y = (K_{xi} \cdot W_{jk}) + b \quad (3)$$

where the Kernel function is K_{xi}, b is bias term of SVM network and W_{jk} is called as the weight vector. K_x and W show Lagrange multipliers. K_{xi} is a nonlinear function that maps the input vectors into a high-dimensional feature space. The inner product of the inputs is calculated by using kernel functions. Lagrange multipliers show the weights the non-linear Radial-Base functions used in this study. Details about SVM can be found in Haykin (1999).

3. RESULTS AND DISCUSSIONS

In this study, IR, PQR and SVM methods were compared according to the following statistical criteria. In the models, average daily flow (Q) and turbidity (T), were used for the sediment concentration (SC) modeling. In this study, 705 of the daily Q, T and SC were used for training and the 200 daily data were used for testing. In the modeling, Statistical criteria such as Determination coefficient (R^2), Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were calculated and the results were interpreted for each model.

The R^2 measures the strength of the correlation between the predicted and real values. MAE and RMSE measure the accuracy by continuously calculating the mean size of the errors in the estimation without taking into account the aspects of the variables. MAE and RMSE are used to diagnose the possibility of errors. MAE and RMSE calculations were determined according to below equations:

$$\mathrm{MAE} = \frac{1}{n} \sum_{j=1}^{n} \left| SC_{\mathrm{MEASURE}} - SC_{\mathrm{predicted}} \right| \tag{4}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(SC_{\mathrm{MEASURE}} - SC_{\mathrm{predicted}} \right)^2} \tag{5}$$

In equation 5 above, "SC" refers to the daily measured sediment concentration *(mg/L)* values. MAE, RMSE and R^2 statistics are calculated for comparison of methods used. R^2, PQR and SVM results are given in Table 1.

Table 1. Statistical results of IR, PQR and SVM models

MODELS	INPUTS	MAE (mg/L)	RMSE (mg/L)	R^2
PQR	T,Q	41,71	128,51	0,89
IR	T,Q	25,40	65,22	0,98
SVM	T,Q	64,42	231,64	0,91

T: Turbidity (fnu), Q: Streamflow (m^3/s).

The most appropriate result among the models where data is used, as shown in Table 1, is given by IR model. Distribution and scatter graphs of PQR model are shown in Figure 3 and 4 below, respectively.

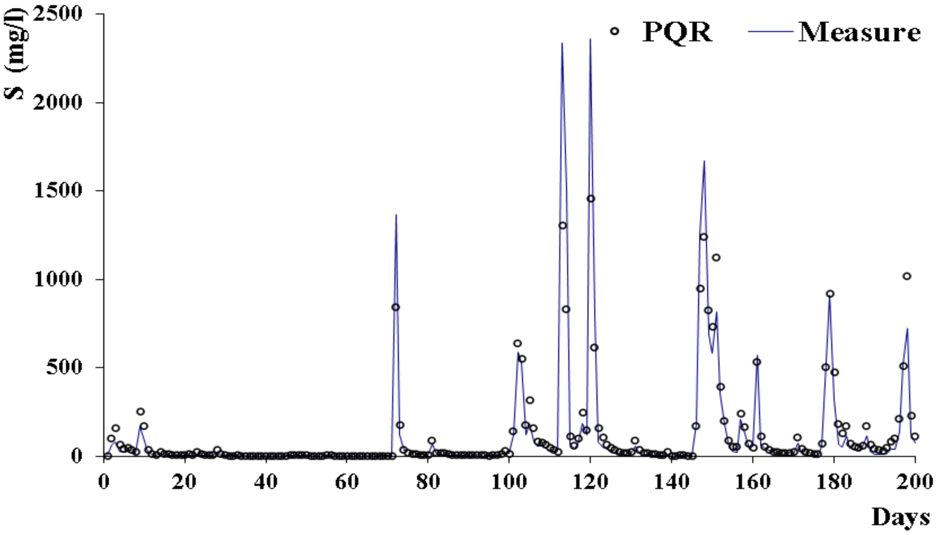

Fig. 3. PQR model distribution charts for Sediment test data

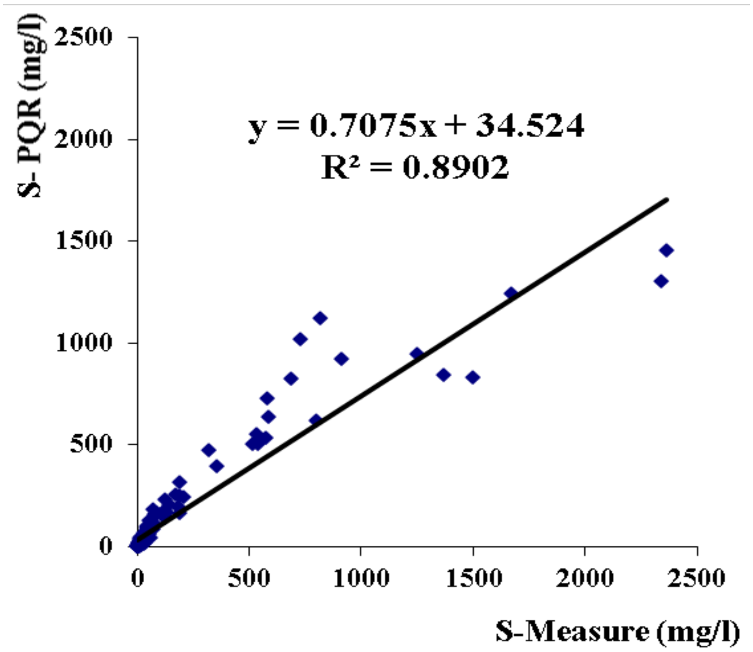

Fig. 4. PQR model scatter charts for Sediment test data

Figure 3 and 4. show the performance of PQR model. Determination coefficient for PQR model is 0,890. Distribution and scatter graphs of IR model are shown in Figure 5 and 6 below, respectively.

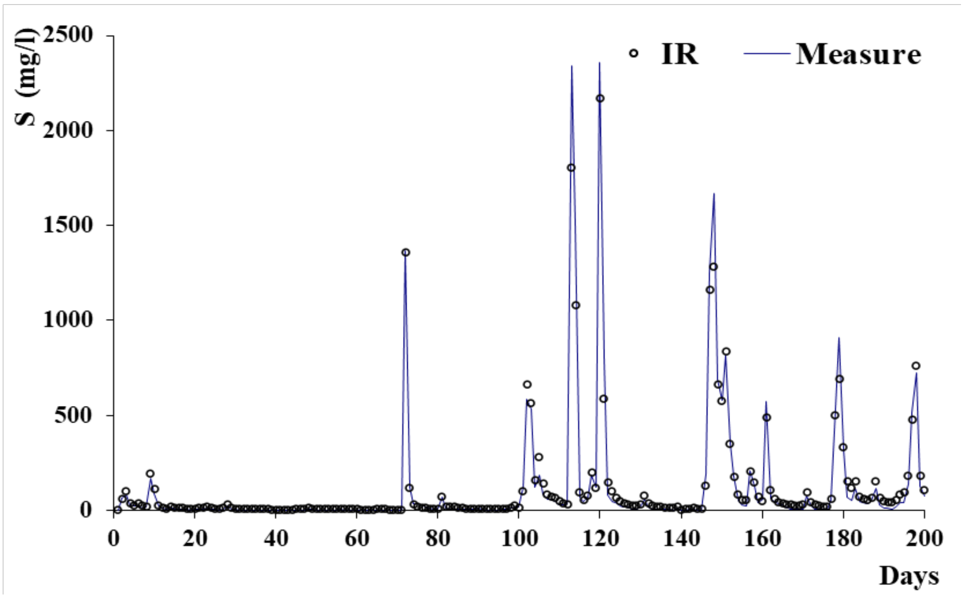

Fig. 5. IR model distribution charts for Sediment test data

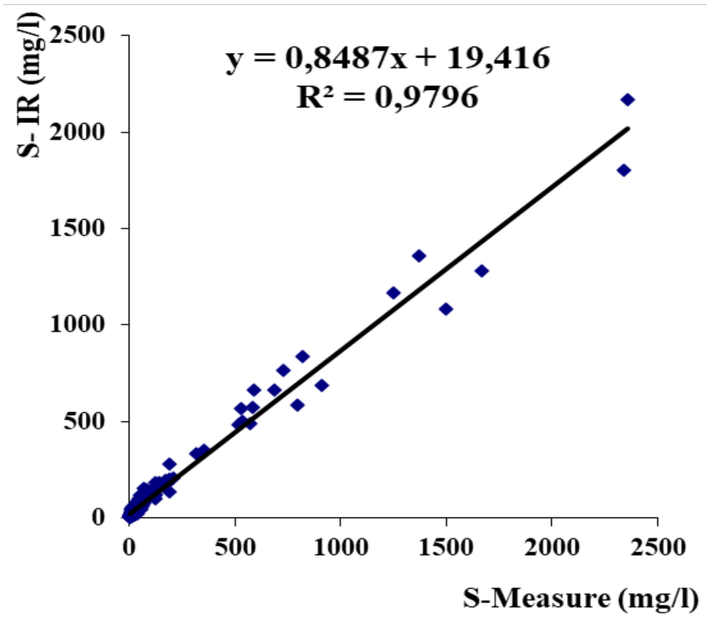

Fig. 6. IR model scatter charts for Sediment test data

Figure 5 and 6. show the performance of PQR model. Determination coefficient for PQR model is 0,980. Distribution and scatter graphs of SVM model are shown in Figure 7and 8 below, respectively.

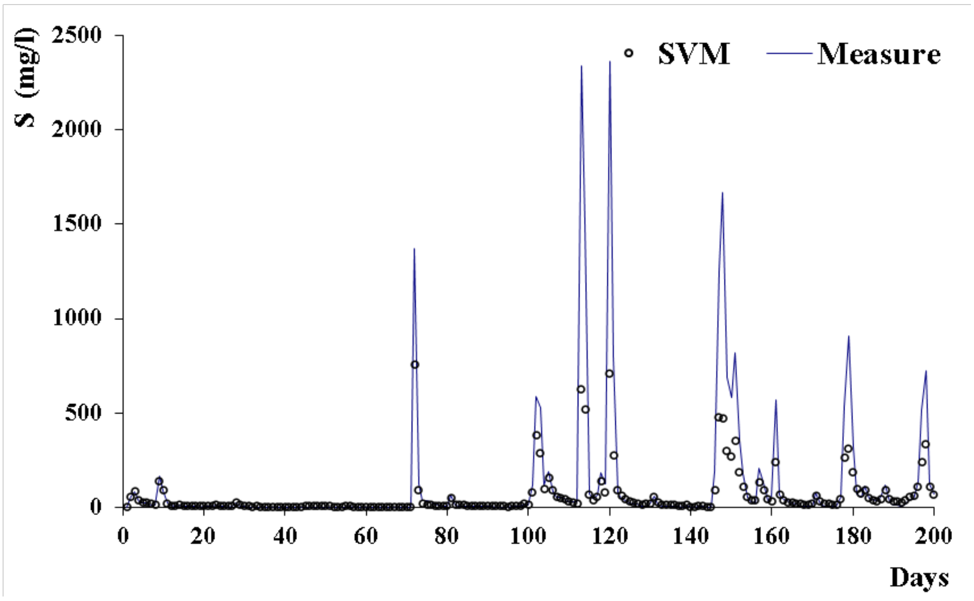

Fig. 7. SVM model distribution charts for Sediment test data

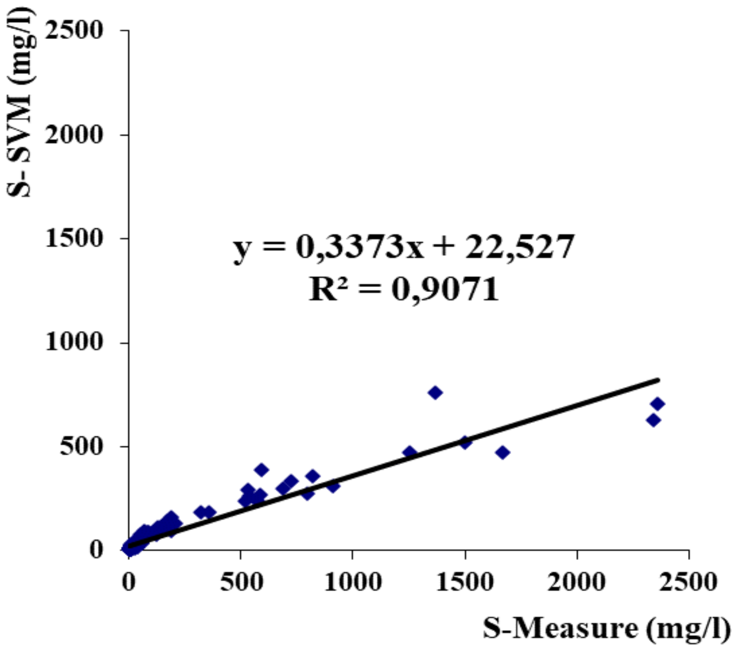

Fig. 8. SVM model scatter charts for Sediment test data

Figure 7 and 8. show the performance of SVM model. Correlation coefficient for SVM model is 0,907.

According to Table 1 and distribution-scatter charts, it is observed that IR model has good results for the test data. The good results can be expressed by a high correlation (R^2) and a low error amount (RMSE, MAE). Accordingly, the best estimation is given by the IR model with the highest value of determination (R^2 = 0,980) and the lowest error value - RMSE (65,22 mg/L) and MAE (25,40 mg/L). As a result of this study, the use IR methods for modeling the relationship between Sediment can be presented as an alternative to traditional methods. Among all models, PQR and SVM methods showed poorer performance in SC estimation.

4. CONCLUSIONS

In this study, Interaction Regression (IR), Pure-Quadratic Regression (PQR) and Support Vector machine (SVM) methods were used to obtain the Sediment Concentration (SC) estimation. This study investigated the abilities of new developed SVM traditional regression (IR and PQR) methods to provide SC estimation for the Patapsco River in USA.

Average daily streamflow (Q) and turbidity (T), were used for the SC modeling. As a result, the low amount of error (MAE, RMSE) ratios and high correlation (R^2) provided the desired performance in IR regression method were that determine SC.

IR method has been found to be a model that can be applied in the estimation of the SC occurring with different streamflow and turbidity conditions in the studies which water planning is required and in determining the water level changes. As a final result, it is understood that regression model can be used for hydrological modelling which is necessary for water resources management and planning future requirements.

REFERENCES

1. Ara Rahman, S., & Chakrabarty, D. (2020). Sediment transport modelling in an alluvial river with artificial neural network. *Journal of Hydrology*, *588*, 125056. https://doi.org/10.1016/j.jhydrol.2020.125056
2. Demirci, M., Unes, F., & Kaya, Y. Z. (2018). *Modeling of Dam Reservoir Volume Using Adaptive Neuro Fuzzy Method*. Aerul si Apa. Componente ale Mediului; Cluj-Napoca 145–152. https://doi.org/10.24193/AWC2018_18
3. Demirci, M., Unes, F., & Akoz, M. S. (2016). Determination of Nearshore Sandbar Crest Depth Using Neural Network Approach. *International Journal of Advanced Engineering Research and Science*, *3*(12), 133–140. https://doi.org/10.22161/ijaers/3.12.27
4. Kisi, O., & Zounemat-Kermani, M. (2016). Suspended Sediment Modeling Using Neuro-Fuzzy Embedded Fuzzy c-Means Clustering Technique. *Water Resources Management*, *30*(11), 3979–3994. https://doi.org/10.1007/s11269-016-1405-8
5. Madani Cherif, H., Khanchoul, K., Bouanani, A., & Terfous, A. (2017). Prediction of sediment yield at storm period in Northwest Algeria. *Arabian Journal of Geosciences*, *10*(9), 198. https://doi.org/10.1007/s12517-017-2983-3
6. Mirbagheri, S. A., Nourani, V., Rajaee, T., & Alikhani, A. (2010). Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. *Hydrological Sciences Journal*, *55*(7), 1175–1189. https://doi.org/10.1080/02626667.2010.508871
7. Olyaie, E., Banejad, H., Chau, K.-W., & Melesse, A. M. (2015). A comparison of various

artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environmental Monitoring and Assessment*, *187*(4), 189. https://doi.org/10.1007/s10661-015-4381-1

8. Rajaee, T., & Jafari, H. (2020). Two decades on the artificial intelligence models advancement for modeling river sediment concentration: State-of-the-art. *Journal of Hydrology*, *588*, 125011. https://doi.org/10.1016/j.jhydrol.2020.125011
9. Rajaee, T., Mirbagheri, S. A., Zounemat-Kermani, M., & Nourani, V. (2009). Daily suspended sediment concentration simulation using ANN and neuro-fuzzy models. *Science of The Total Environment*, *407*(17), 4916–4927. https://doi.org/10.1016/J.SCITOTENV.2009.05.016
10. Riahi-Madvar, H., & Seifi, A. (2018). Uncertainty analysis in bed load transport prediction of gravel bed rivers by ANN and ANFIS. *Arabian Journal of Geosciences*, *11*(21), 688. https://doi.org/10.1007/s12517-018-3968-6
11. Shamaei, E., & Kaedi, M. (2016). Suspended sediment concentration estimation by stacking the genetic programming and neuro-fuzzy predictions. *Applied Soft Computing*, *45*, 187–196. https://doi.org/10.1016/j.asoc.2016.03.009
12. Üneş, F., Demirci, M., Taşar, B., Kaya, Y. Z., & Varçin, H. (2019). Modeling of dam reservoir volume using generalized regression neural network, support vector machines and m5 decision tree models. *Applied Ecology and Environmental Research*, *17*(3), 7043–7055. https://doi.org/10.15666/AEER/1703_70437055
13. Üneş, F., Karaeminoğullari, A. B., & Taşar, B. (2020). Forecasting of River Sediment Amount using Machine Model. *International Journal of Environment, Agriculture and Biotechnology*, *5*(1), 9–15. https://doi.org/10.22161/ijeab.51.2
14. Üneş, F., Taşar, B., Demirci, M., Zelenakova, M., Ziya Kaya, Y., & Varçin, H. (2021). Daily Suspended Sediment Prediction Using Seasonal Time Series and Artificial Intelligence Techniques. *Rocznik Ochrona Środowiska*, *23*, 117–137. https://doi.org/10.54740/ros.2021.008
15. Üneş, F., Taşar, B., Kaya, Y. Z., & Demirci, M. (2018). Üneş, F., Doğan, S., Taşar, B., Kaya, Y., & Demirci, M. (2018). The evaluation and comparison of daily reference evapotranspiration with ANN and empirical methods. *Natural and Engineering Sciences,* 3(3), 54-64.
16. USGS.gov. (n.d.). Science for a Changing World [WWW Document]. https://www.usgs.gov/
17. Vafakhah, M. (2013). Comparison of cokriging and adaptive neuro-fuzzy inference system models for suspended sediment load forecasting. *Arabian Journal of Geosciences*, *6*(8), 3003–3018. https://doi.org/10.1007/s12517-012-0550-5
18. Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S., & Yaseen, Z. M. (2020). On the complexities of sediment load modeling using integrative machine learning: Application of the great river of Loíza in Puerto Rico. *Journal of Hydrology*, *585*, 124759. https://doi.org/10.1016/j.jhydrol.2020.124759