Rocznik Ochrona Środowiska

Volume 23 Year 2021 ISSN 1506-218X pp. 117-137

https://doi.org/10.54740/ros.2021.008 open access

Received: 11 February 2021 Accepted: 28 April 2021 Published: 06 December 2021

Daily Suspended Sediment Prediction Using Seasonal Time Series and Artificial Intelligence Techniques

Fatih Üneş
Civil Engineering Department, Iskenderun Technical University, Turkey

*Bestami Taşar**
Civil Engineering Department, Iskenderun Technical University, Turkey

Mustafa Demirci
Civil Engineering Department, Iskenderun Technical University, Turkey

Martina Zelenakova
Environmental Engineering Institute, Kosice Technical University, Slovakia

Yunus Ziya Kaya
Civil Engineering Department, Osmaniye Korkut Ata University, Turkey

Hakan Varçin
Civil Engineering Department, Iskenderun Technical University, Turkey

**corresponding author's e-mail: bestami.tasar@iste.edu.tr*

Abstract: Estimating the amount of suspended sediment in rivers correctly is important due to the adverse impacts encountered during the design and maintenance of hydraulic structures such as dams, regulators, water channels and bridges. The sediment concentration and discharge currents have usually complex relationship, especially on long term scales, which can lead to high uncertainties in load estimates for certain components. In this paper, with several data-driven methods, including two types of perceptron support vector machines with radial basis function kernel (SVM-RBF), and poly kernel learning algorithms (SVM-PK), Library SVM (LibSVM), adaptive neuro-fuzzy (NF) and statistical approaches such as sediment rating curves (SRC), multi linear regression (MLR) are used for forecasting daily suspended sediment concentration from daily temperature of water and streamflow in the river. Daily data are measured at Augusta station by the US Geological Survey. 15 different input combinations (1 to 15) were used for SVM-PK, SVM-RBF, LibSVM, NF and MLR model studies. All approaches are compared to each other according to three statistical criteria; mean absolute errors (MAE), root mean square errors (RMSE) and correlation coefficient (R). Of the applied linear and nonlinear methods, LibSVM and NF have good results, but LibSVM generates a slightly better fit under whole daily sediment values.

Keywords: Prediction, Neuro-Fuzzy, Sediment Rating Curves,
Support Vector Machines, Suspended Sediment

1. Introduction

Estimation of suspended sediment amount in streams and rivers correctly is a substantial value in the design and maintenance of hydraulic structures such as dams, bridges, etc. In particular, the sediment that accumulates in the water storage structures such as dam reservoir reduces the reservoir capacity. Decrease of the reservoir capacity causes shortening of the economic life of facilities. In order to prevent or even delay these damages, a passive storage called dead storage is determined in the dam reservoir. It is designed to remain under the water intake structure. The service life of a dam, namely its useful life, depends on the amount of storage. Therefore, it is important to accurately forecast the type and amount of sediments in dam projects.

In rivers, suspended sediments are also transported with water during stream-flow movement. These sediments are consisted of either erosion in river basin or by abrasion in the stream bed. Throughout river, scouring and accumulation in stream bed occur as a result of sediment movements. As a result of this scouring and accumulation, the shape of the river bed and morphological structure is expected to change. For the solution of these problems, suspended sediment estimations are needed.

Determination of suspended sediments by measurements is the most accurate method. However, this method takes time and is costly. In addition, there is no measurement of the amount of sediments in many observation stations, although water flow is measured. It is especially difficult to measure the amount of sediments in the stations in case of flooding.

Artificial intelligence techniques have been widely used to solve complex problems in recent years. Examples of these are; artificial neural networks (ANN) (Saplıoglu & Cimen 2010, Turhan & Cağatay 2016, Demirci et al. 2017, Unes et al. 2018a, 2018b, Turhan et al. 2019) and adaptive network-based fuzzy inference system (ANFIS) (Jang 1993, Ghavidel & Montaseri 2017, Ebtehaj & Bonakdari 2017, Demirci et al. 2018, Catal & Saplioglu 2018, Ehteram et al. 2021).

In the past, many researchers also applied artificial intelligence methods and obtained different results in order to explain the sediment amount problem and provide correct solutions. Kisi (2005) estimated the concentration of sediment in the stream using ANN. Mirbagheri et al. (2010) evaluated the applicability of the sediment rating curves (SRC), and fuzzy rule-based (NF) models in estimating the concentration of sediment in the rivers using the coefficient of determination and demonstrated that the NF model gives better results for predicting the sediment concentration. Fırat and Güngör (2010) used ANN and NF methods for sediment estimation. According to the results, they demonstrated that the NF approach provides high performance. Wieprecht et al. (2013) used an ANFIS to estimate bed load and total bed material load in the Rhine River. They used two-thirds of the available data sets (bed load and total bed material) for the training phase and the remaining one for the testing phase. They stated that the ANFIS modeling approach could be a good alternative for estimating bed load and total bed material load. Demirci and Baltaci (2013) investigated the viability of the SRC, multi linear regression (MLR), and fuzzy logic (FL) methods in estimating sediment concentration. FL model has shown good results in comparisons for both 5-year and 50-year sediment estimations. Demirci et al. (2015) used an ANN approach for forecasting sediment concentration in Little Coal river, West Virginia, in the USA. It was found that the ANN model gives better estimates than other techniques. Kitsikoudis et al. (2015), derived sediment transport formulas for sand-bed rivers. They used ANN, ANFIS, and genetic programming based symbolic regression methods to derive these formulas. Partovian et al. (2016), made a study on the daily sediment and flow model of the Minnesota River. They applied the previously measured data to ANN and ANFIS models. They compared it with MLR and auto-regressive moving average (ARMA) models to evaluate the performance of their models. According to their results, ANN and ANFIS models performed better than MLR model.

Kisi and Zounemat (2016) conducted studies to forecast the amount of sediment in 2 stations on the Muddy river in the USA. The input parameters were the daily flow rate and the amount of sediment concentration data in the study. They used ANN, NF, SRC, and CNF models (Clustered Neuro-Fuzzy model, developed from classic NF). The CNF method has been shown to provide better sediment estimation results than others. CNF method can be presented as an alternative to ANN, NF, SRC methods in sediment prediction. Seyedian and Rouhani (2015) studied the capabilities of the ANFIS to estimate daily sediment loads for four stations in the USA. They compared the ANFIS model they created with the SRC model in terms of error amounts (RMSE, MBE), and determination coefficient (R^2) values. They stated that the ANFIS model performed better than the SRC model. Tasar et al. (2017) used M5tree (M5T), ANN approaches, and statistical approaches to estimate sediment load. Gunawan et al. (2017) estimated sediment load using the backpropagation network (BPNN) scheme, which is an ANN method. As a result, they stated that this model performs better than other known calculation methods with its correlation coefficient (R) and mean square error (MSE) stability. Buyukyildiz and Kumcu (2017) studied to predict sediment load which gauged at Ispir Bridge station, Çoruh River in Turkey. Choubin et al. (2018), estimated river sediment using the classification and regression tree (CART) model with machine learning techniques. Emamgholizadeh and Demneh (2019), compared artificial intelligence models for the estimation of daily suspended sediment of Telar and Kasilian rivers in Iran. Results showed that ANN and ANFIS models were better performance than the other models. Salih et al. (2020), predicted suspended sediment load in river-based on river discharge information by using newly developed data mining models. Among the applied data mining models, the M5P model gave the best prediction result. Meshram et al. (2020), were used radial basis function (RBF), support vector machine (SVM), artificial neural networks (ANNs), and multiple model (MM)-ANNs to predict sediment yield. Results showed that the MM-ANNs model results gave best performance in all models.

This study aims to investigate the performance of data-driven methods, including two types of perceptron support vector machines with radial basis function kernel (SVM-RBF), poly kernel learning algorithms (SVM-PK), adaptive neuro-fuzzy (NF), and statistical approaches such as sediment rating curves (SRC), multi linear regression (MLR) in sediment concentration predictions in rivers. Furthermore, the LibSVM model, which is one of the new modeling techniques, was used for sediment estimation in this study.

2. Methods

2.1. Sediment Rating Curve

Conventional sediment rating curve (SRC) shows the connection between the sediment amount and the streamflow measured in any control section of the rivers. If Q indicates the streamflow and S indicates the concentration of sediment, the connection between these two variables;

$$S = aQ^b \tag{1}$$

where a and b are rating curve constant coefficients. Williams (1978) who examined the S-Q relationship given in Equation 1, proved that there is no uniform relationship. In some rivers, the S-Q relationship is followed by two different values. That is, the amount of sediment at different times in the stream can be different due to the hydrological causes for the same discharge value. In many cases, accurate sediments cannot be forecasted and are inadequate using these curves.

2.2. Multi Linear Regression

Multi linear regression (MLR) is a type of analysis for predicting a dependent variable, depending on 2 or more independent variables associated with the dependent variable (Berk, 2004). In MLR analysis, the relationship between further than one independent variable (x_1, x_2 ... x_n) and a dependent variable (y) is examined. That is, if the dependent variable "y" is assumed to be impressed by "n" independent variables such as x_1, x_2, x_3 … x_n . If the relationship between them is assumed to be linear, MLR equation "y" dependent variable can be expressed as:

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n \tag{2}$$

When starting the regression analysis, the variables (two or more variables) to be searched first must be determined, an acceptance must be made then for the type of equation that shows the relationship between these variables.

2.3. Support Vector Machine

Support vector machine (SVM) is an approach of learning found by Cortes and Vapnik (1995) for solving the classification and regression problems. It is likely that classification of variables on a plane by drawing a boundary between them. The boundary which is drawn between variables must be as far as possible to each variable. The SVM defines how to draw this boundary between variables group. SVM studies according to statistical learning theory. A set of training data [$(x_1, y_1), (x_2, y_2), \ldots\ldots (x_n, y_n)$], where "$x_i$" value indicates that the input space of the sample and has a corresponding target "y_i" value. The SVM estimating function expressed as:

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

Where the Kernel function is K_{xi}, b is bias term of SVM network and W_{jk} is called the Lagrange multipliers that obtain to the significance of the training data sets for the output data. The network architecture of SVM is given in Fig. 1.

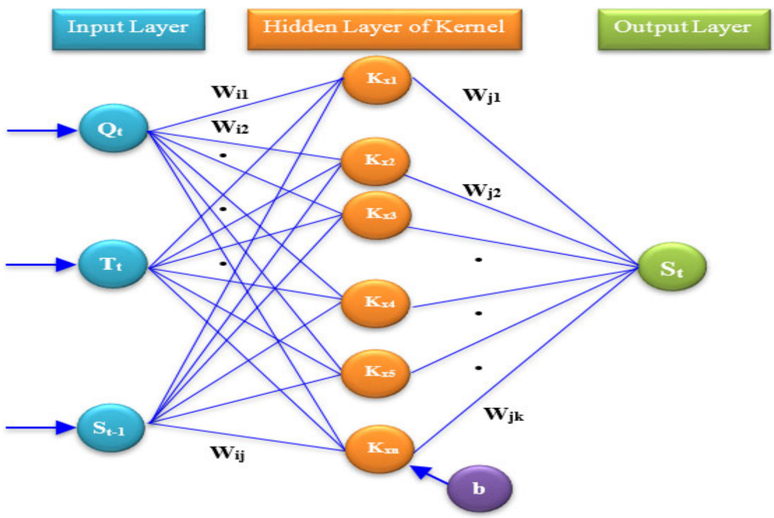

Fig. 1. Network architecture of SVM10 model

In the SVM model, firstly input data is determined and processing is started. Then Which Kernel type is decided (In this paper, radial basis function and polynomial Kernel is used). After decided Kernel type, parameters of kernel are obtained. Finally, Models trained and applied test results. All this SVM stage is given in Fig. 2.

The kernel function of non-linear radial basis (Hsu et al. 2003) is:

$$K_{xi} = e^{-\gamma \|x_i - y_i\|^2} \qquad \gamma > 0 \quad \text{and} \qquad i = 1,2,3,...n \tag{4}$$

where γ is a user-defined parameter. The kernel function of polynomial (Hsu et al. 2003) is:

$$K_{xi} = (x.y + c)^d \qquad i = 1,2,3,...n \tag{5}$$

where function degree define as d and c is constant parameter. If d = 1, function became linear condition.

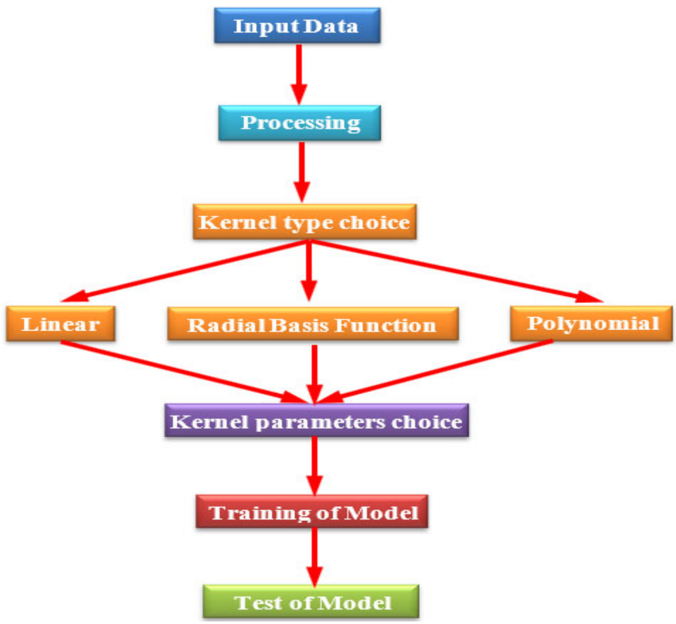

Fig. 2. SVM model stage

2.4. Library SVM

Library SVM (LibSVM) is one of the machine learning algorithms for support vector classification, regression (Chang & Lin 2002). C-SVC, nu-SVC, epsilon-SVR, and nu-SVR are the most commonly known LibSVM machine learning algorithms. It works with multi-class classification.

The difference between the LibSVM model from the classic SVM model is that SVM types can be selected. Once the SVM types have been determined, the processing steps are similar to the conventional SVM. In this model, nu-SVR SVM type and RBF Kernel type is selected. The RBF kernel portion was previously explained. Schölkopf et al. (1998) proposed a support vector classification algorithm called nu-SVR, which was self-adjusting the epsilon parameter. Nu-SVR provides a parameter that can set and control the number of support vectors based on the total number of samples in the data set.

2.5. Neuro-Fuzzy

Suitable methods for classical analysis (key curves, linear regression) cannot be applied successfully in the field of hydrology because hydrological events are dependent on many variables and have non-linear relationships. In order to analyze such hydrological events, simple, economical, and easy methods have been developed. Therefore, more accurate and efficient results can be obtained as the events are discussed in the perspective of fuzzy. Neuro-fuzzy (NF) is a very effective logical understanding that can be used for this approach (Üneş et al. 2015).

NF was initially represented by Jang (1993). The NF system works as a learning algorithm created with neural network functional rules. The parameters of neuro-fuzzy systems are obtained by neural network learning algorithms in fuzzy rule-based systems, and different analysis methods such as Sugeno can be applied. The NF with Sugeno type works according to "If-Then" rules and the NF structure uses the Sugeno-Fuzzy rules. It is possible to introduce fuzzy systems are logical models which is consisted of “If-Then” rules and membership functions. The Sugeno NF system, generated by two rules using three inputs, is shown in Fig. 3. Where, w_1 or w_2 is obtained by weighted mean of individual rule outputs. NF structure is shown in Fig. 4. NF is connected via directional links and contains several nodes. Every node has a node function that can be constant or adjustable parameters. During learning process, the fittest parameter values are obtained by adjusting training data. NF is a method with the basic learning rules that want to reduce the total of the squared errors between the network output data and the real output data.

Sugeno system in the first degree, two fuzzy If / Then rules with a typical set of rules can be specified as follows (Sayed et al. 2003).

1. Rule: If x is A_1, y is B_1 and z is C_1; then $f_1 = p_1x + q_1y + r_1z + s$
2. Rule: If x is A_2, y is B_2 and z is C_2; then $f_2 = p_2x + q_2y + r_2z + s$

where: A_1, B_1, C_1 and A_2, B_2, C_2 are linguistic labels (such as “low”, “medium” or “high”), f_1 and f_2 denote, respectively, output functions of 1. and 2. rule, $\{p_i, q_i, r_i, s\}$ specified as result parameters.

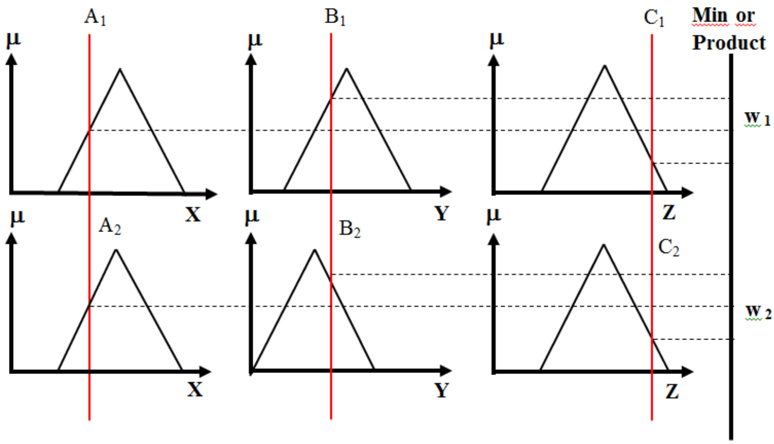

Fig. 3. Sugeno type fuzzy model generated by two rules using three inputs

With a Sugeno fuzzy inference system, NF network is generated. Researchers who want more information about NF can be found in Jang (1993).

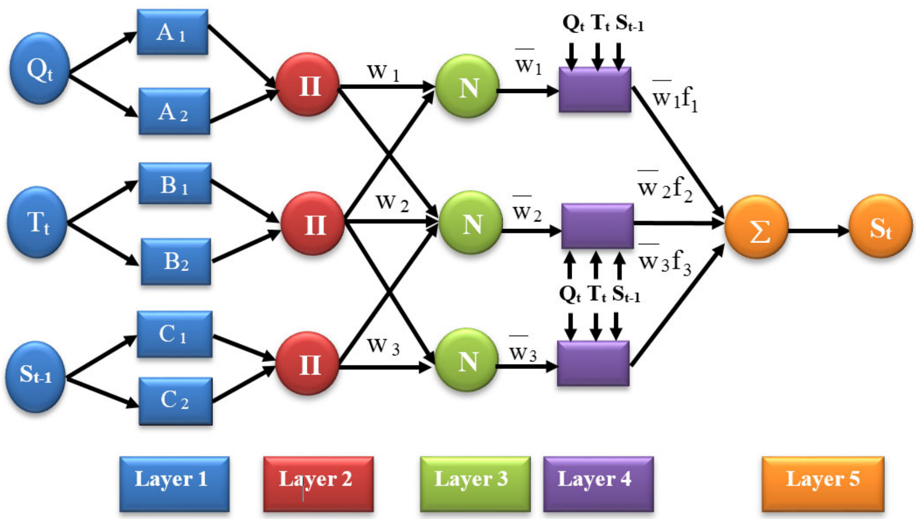

Fig. 4. Structure of the Neuro-Fuzzy (NF) system using three inputs

3. Study area

In this paper, Des Moines County area (Hydrologic Unit No: 18020109) in Iowa, United States was selected as the study area for estimating suspended sediment concentration amount. The Augusta station at the Skunk River has been studied (USGS Station No: 05474000]). Gage Datum at sea level is about 160 m (NGVD29, National Geodesy Vertical Datum). The location of the Augusta station at the Skunk River, which is the right tributary of Mississipi, is shown in Fig. 5.

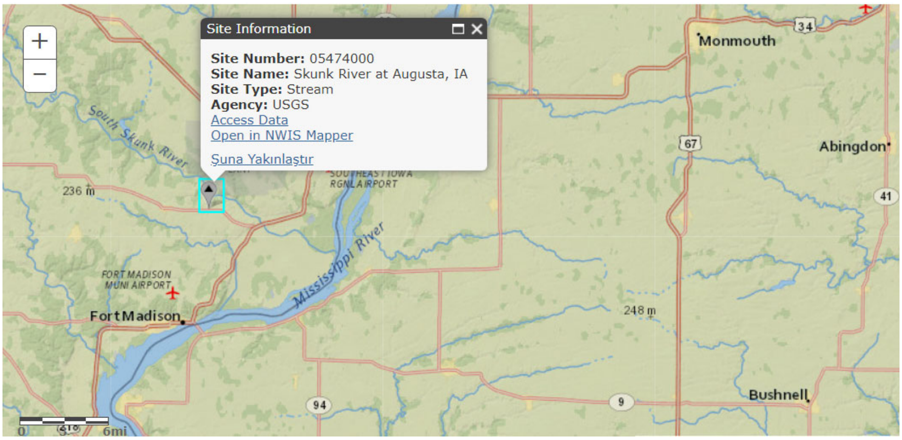

Fig. 5. Augusta station overviews on Skunk River (USGS)

4. Datasets

In this study, 3-year measurement data belonging to the Augusta station were used. Model performs was investigated by using daily average temperature of water, real-time streamflow, and sediment concentration data from Augusta Station at the Skunk River in the USA. A total of 1095 days of three years (2007-2009) was used for estimation. Daily mean temperature (T_{mean}), streamflow (Q) and sediment concentration (S) values are shown in Fig. 6, respectively.

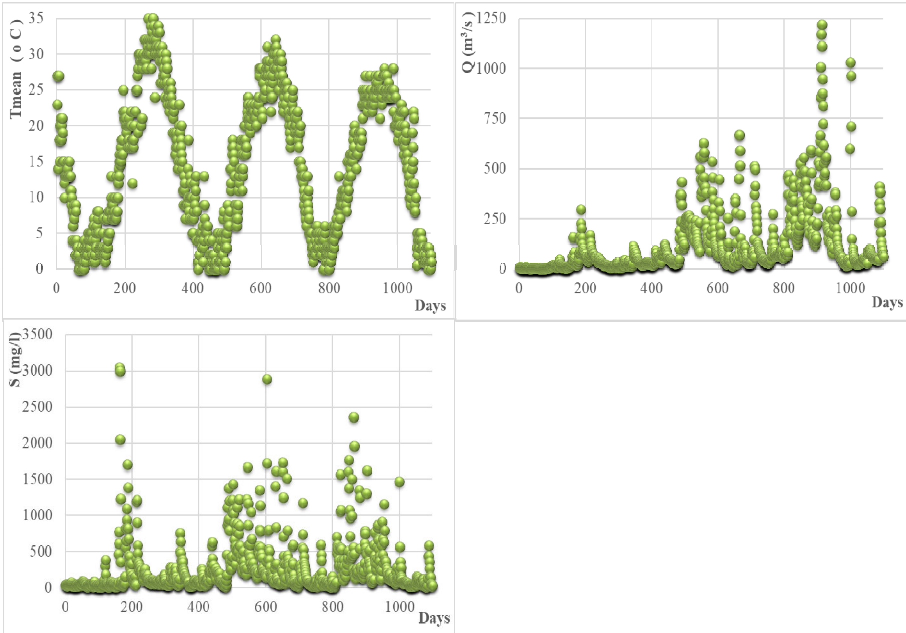

Fig. 6. Daily mean water temperature (T_{mean}), streamflow (Q), and sediment concentration (S) values for 3 years

5. Model results

5.1. Error analysis

The results of SRC, MLR, SVM-RBF, SVM-PK, LibSVM, and NF results for the models generated for 3 years data are as follows. For each model, the mean absolute errors (MAE), the root mean square errors (RMSE), and the correlation coefficients (R) between the models are also used to assess the performance of model estimations and observations. The MAE and RMSE were obtained as follows. The observed values were calculated.

$$MAE=\frac{1}{N}\sum_{i=1}^{N}\left|Y_{i_{observed}}-Y_{i_{estimate}}\right| \tag{6}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Y_{i_{observed}}-Y_{i_{estimate}}\right)^2} \tag{7}$$

Here, N represents data numbers and Y_i sediment concentration data. The model results of the statistical criteria calculated in the study are given in Table 1.

5.2. Model results

In the Sediment Rating Curve (SRC), 3-year data were assessed. The SRC was drawn by fitting a curve to the flow rate (Q_t) and sediment concentration (S_t) data from the station. From the equation expressing this curve, sediment concentration data for SRC were obtained by setting the measured flow values at the gauge station instead of the unknown x value. The SRC for the training data is shown in Fig. 7.

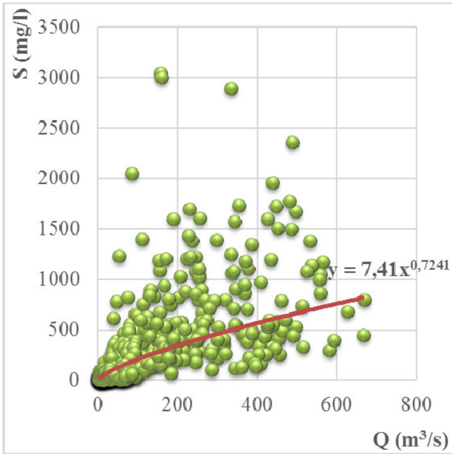

Fig. 7. Sediment Rating Curve (SRC) using train data

The established SRC model using training data is given in Equation 8 and this equation is also applied to the test data. Statistical results (RMSE, MAE and R) and inputs of SRC model are given in Table 1.

$$S_t = 7.4008 \text{ x } (Q_t)^{0.7241} \tag{8}$$

After determining the Sediment Rating Curve, a scatter plot is drawn for the test data. For the test phase, this scatter plot and distribution graph between measured values and the SRC results are shown in Fig. 8a.

The correlation coefficient, between measured values and SRC prediction results, R = 0.47 was obtained for the test. The sediment rating curve multiplies the observed value by the predicted value of real values. When the distribution graph for the test data is examined, it is seen that the SRC sediment estimated values differ than the actual values. According to the SRC results, SRC method has low R value for the test data and when the distribution graph is examined, it is seen that the desired estimates cannot be obtained and it is considerably lower than the station measurement data.

Table 1. Performances of SRC, MLR, SVM-RBF, SVM-PK, LibSVM, NF model for test data

Model No and Inputs		SRC			MLR			SVM-RBF		
Models	Model Inputs	RMSE (mg/l)	MAE (mg/l)	R	RMSE (mg/l)	MAE (mg/l)	R	RMSE (mg/l)	MAE (mg/l)	R
Model 1	Q_t	**278.77**	**163.19**	**0.47**	446.29	240.67	0.41	325.05	175.14	0.42
Model 2	Q_t, Q_{t-1}				437.52	223.22	0.55	333.82	181.58	0.39
Model 3	Q_t, Q_{t-1}, Q_{t-2}				441.91	224.55	0.54	324.08	178.40	0.40
Model 4	T_t				267.47	191.84	0.31	285.85	156.79	0.31
Model 5	T_t, T_{t-1}				266.79	191.65	0.32	284.52	156.40	0.31
Model 6	T_t, T_{t-1}, T_{t-2}				267.13	193.14	0.24	284.30	156.66	0.30
Model 7	S_{t-1}				172.65	87.13	0.75	**180.76**	**81.09**	**0.75**
Model 8	S_{t-1}, S_{t-2}				170.86	85.45	0.76	187.76	82.66	0.72
Model 9	S_{t-1}, S_{t-2}, S_{t-3}				168.09	83.86	0.77	187.66	82.71	0.72
Model 10	Q_t, T_t, S_{t-1}				222.63	122.34	0.69	213.61	112.07	0.66
Model 11	Q_t, Q_{t-1}, T_t, T_{t-1}, S_{t-1}, S_{t-2}				236.75	107.64	0.79	192.53	102.00	0.70
Model 12	Q_t, Q_{t-1}, T_t, T_{t-1}, T_{t-2} S_{t-1}, S_{t-2}				236.75	107.64	0.79	193.50	102.37	0.70
Model 13	Q_t, Q_{t-1}, T_t, T_{t-1},S_{t-1}, S_{t-2}, S_{t-3}				233.90	106.53	0.79	190.44	101.49	0.70
Model 14	Q_t, Q_{t-1}, T_t, T_{t-1}, T_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}				**230.92**	**103.42**	**0.80**	190.45	101.52	0.70
Model 15	Q_t, Q_{t-1}, Q_{t-2},T_t, T_{t-1}, T_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}				172.65	87.13	0.75	186.46	97.86	0.72

MAE: Mean absolute errors; **RMSE**: Root mean square errors **R**: Correlation coefficient

Table 1. cont.

Model No and Inputs		SVM-PK			LibSVM			NF		
Models	Model Inputs	RMSE (mg/l)	MAE (mg/l)	R	RMSE (mg/l)	MAE (mg/l)	R	RMSE (mg/l)	MAE (mg/l)	R
Model 1	Q_t	387.57	200.27	0.41	220.49	143.47	0.63	260.18	178.83	0.63
Model 2	Q_t, Q_{t-1}	385.81	190.90	0.52	155.90	108.75	0.84	182.53	121.09	0.85
Model 3	Q_t, Q_{t-1}, Q_{t-2}	394.10	193.27	0.51	165.37	107.42	0.83	219.16	153.48	0.81
Model 4	T_t	286.16	156.66	0.31	266.78	166.50	0.25	269.56	192.87	0.33
Model 5	T_t, T_{t-1}	284.95	156.46	0.32	266.36	166.52	0.25	281.84	209.67	0.19
Model 6	T_t, T_{t-1}, T_{t-2}	285.60	156.62	0.32	266.23	166.54	0.25	275.70	203.89	0.19
Model 7	S_{t-1}	174.57	73.66	0.75	168.69	72.62	0.77	166.42	85.06	0.78
Model 8	S_{t-1}, S_{t-2}	173.10	72.17	0.76	166.44	69.09	0.78	156.72	77.26	0.82
Model 9	S_{t-1}, S_{t-2}, S_{t-3}	169.48	69.78	0.77	166.32	68.83	0.78	164.80	79.48	0.81
Model 10	Q_t, T_t, S_{t-1}	177.54	84.29	0.74	215.12	77.02	0.81	169.07	102.82	0.80
Model 11	Q_t, Q_{t-1}, T_t, T_{t-1}, S_{t-1}, S_{t-2}	213.95	91.70	0.80	**121.38**	**64.39**	**0.90**	134.75	77.69	0.88
Model 12	Q_t, Q_{t-1}, T_t, T_{t-1}, T_{t-2} S_{t-1}, S_{t-2}	208.59	89.33	0.80	124.94	65.78	0.90	156.48	89.15	0.88
Model 13	Q_t, Q_{t-1}, T_t, T_{t-1},S_{t-1}, S_{t-2}, S_{t-3}	199.74	83.38	0.81	122.35	65.87	0.90	**131.25**	**73.71**	**0.89**
Model 14	Q_t, Q_{t-1}, T_t, T_{t-1}, T_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}	201.30	83.86	0.81	122.97	65.57	0.89	147.00	80.75	0.89
Model 15	Q_t, Q_{t-1}, Q_{t-2},T_t, T_{t-1}, T_{t-2}, S_{t-1}, S_{t-2}, S_{t-3}	**180.72**	**79.41**	**0.83**	125.74	64.53	0.89	151.62	82.64	0.88

MAE: Mean absolute errors; **RMSE**: Root mean square errors; **R**: Correlation coefficient

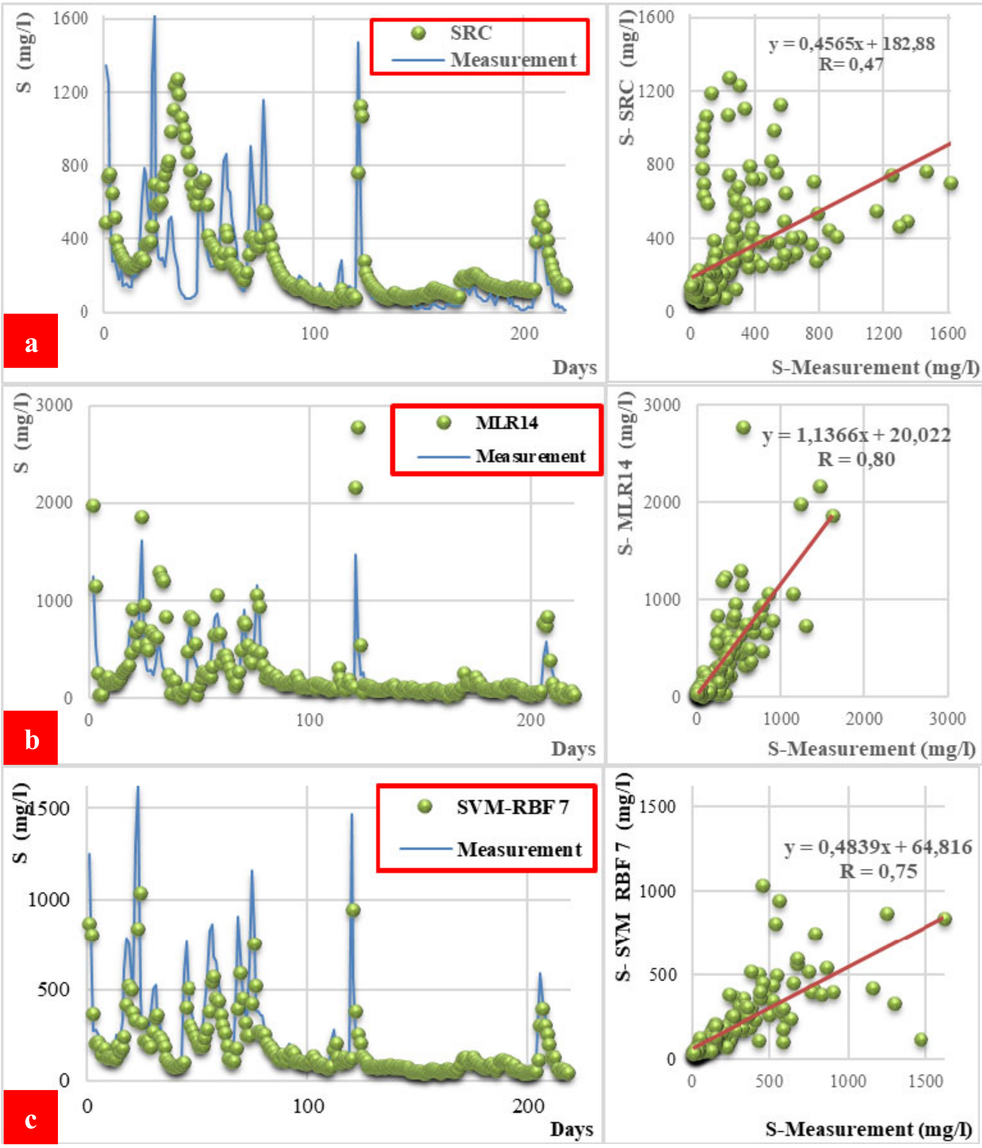

Fig. 8. Measurement and Model distribution-scatter graph for test data **a)** SRC **b)** MLR14 **c)** SVM-RBF 7

As for the Multi Linear Regression (MLR 1 to 15), 3-year data were assessed and the results were determined as follows. The mean temperature (T_t), the lagged time mean temperatures (T_{t-1}, T_{t-2}), the streamflow (Q_t), the lagged time stream-flows (Q_{t-1}, Q_{t-2}), and the lagged time sediment concentrations (S_{t-1},

S_{t-2}, S_{t-3}) at the time “t-1”; “t-2”; “t-3” were used as input values for the MLR model analysis.15 different models (MLR 1 to MLR 15) were used for MLR model studies. Statistical results (RMSE, MAE and R) and inputs of MLR model are given in Table 1.

According to Table 1, the best results in the MLR model belong to the "**MLR14**" model. The results of the MLR14 model are given below. The equation of MLR predictions is obtained by using training data in the MLR14 model and this equation is also applied to the test data. The equation used in MLR model estimation is given in Equation 15.

$$S_t = 232.21 + 3.43Q_t - 3.08Q_{t-1} - 3.7T + 4.14T_{t-1} + 0.27T_{t-2} + \\ + 0.7S_{t-1} + 0.05S_{t-2} + 0.06S_{t-3} \quad (9)$$

For the 3-year data generated, MLR 14 model was evaluated, and MLR 14 distribution and scatter graphs are shown, Fig. 8b.

In the scatter plot generated during the test phase, the correlation coefficient was obtained as R = 0.80. MLR1 test data estimations are lower than training data estimations. The MLR1 estimates in the test phase yield far-reaching estimates of actual values, although the daily sediment values yield better results than the SRC values. It has been observed that the MLR14 prediction values are smaller than the real sediment observation values in the scatter graphs.

For the SVM-RBF model analysis, as in MLR, the 3-year data are divided into training and test data. 15 different models (SVM-RBF 1 to SVM-RBF 15) were used for SVM-RBF model studies. Statistical results (RMSE, MAE, and R) and inputs of the SVM-RBF model are given in Table 1.

According to Table 1, the best results in the SVM-RBF model belong to the "SVM-RBF 7" model. The results of the SVM-RBF 7 model are given below. For the 3-year data generated, SVM-RBF 7 model was evaluated. Distribution and scatter graphs of the SVM-RBF 7 model are shown in Fig. 8c.

The correlation coefficient R = 0.75 was obtained for the graph generated for the test with the SVM-RBF 7 model results. The SVM-RBF 7 estimations at the test phase show better results than SRC model estimation values for the observed daily real-time sediment concentrations. But the MLR 14 model gave slightly better prediction results than the SVM-RBF 7.

For the SVM-PK model analysis, as in MLR and SVM-RBF, the 3-year data are divided into training and test data. 15 different models (SVM-PK 1 to SVM-PK 15) were used for SVM-PK model studies. Statistical results (RMSE, MAE, and R) and inputs of these models are given in Table 1.

According to Table 1, the best results in the SVM-PK model belong to the **"SVM-PK 15"** model. The results of the SVM-RBF 7 model are given below. Distribution and scatter graphs of the SVM-PK 15 model are shown in Fig. 9a. below, respectively.

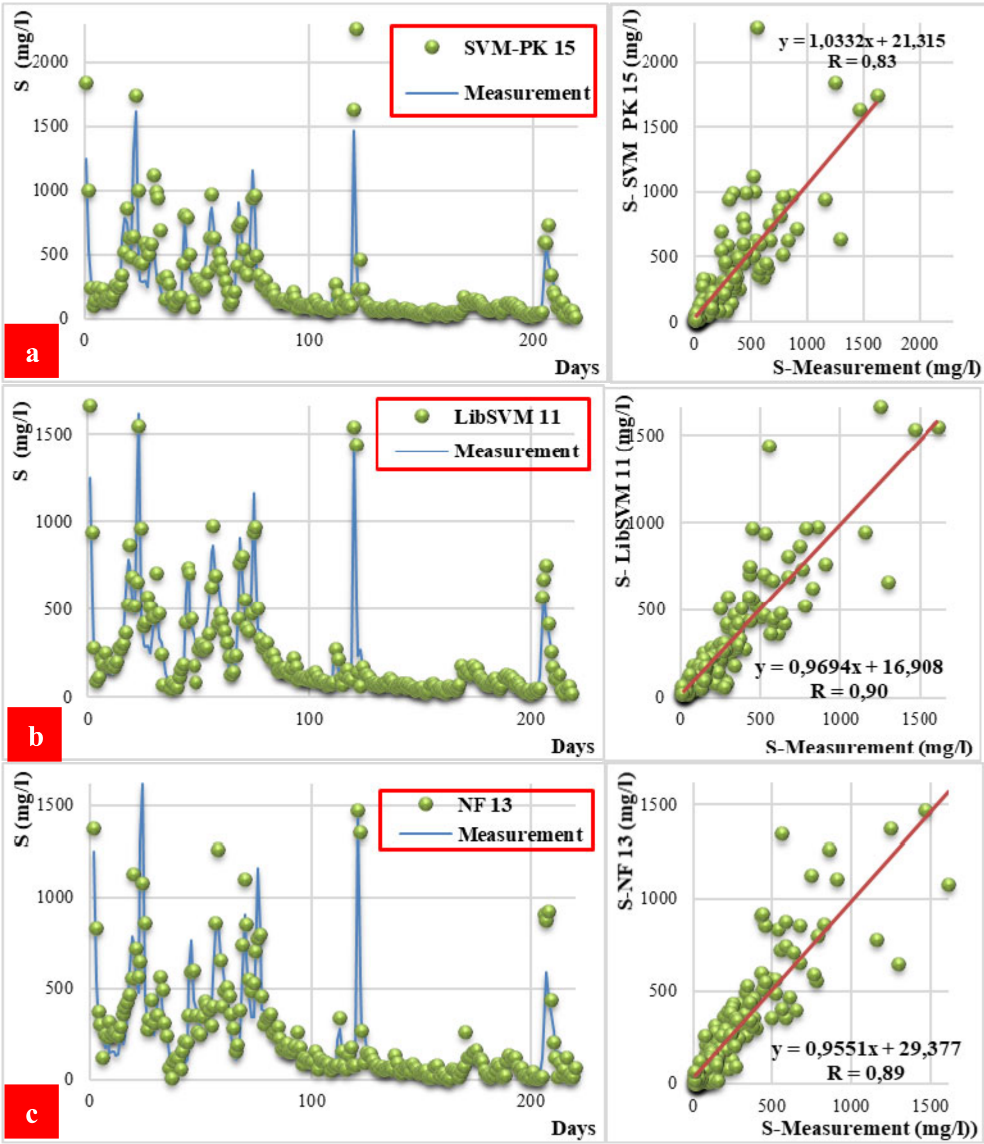

Fig. 9. Measurement and Model distribution-scatter graph for test data **a)** SVM- PK 15 **b)** LibSVM 11 **c)** NF 13

In the scatter plot generated during the test phase, the correlation coefficient was determined as R = 0.83. The SVM-PK 15 estimations at the test phase show better results than the other models (SRC, MLR 14, SVM-RBF 7) values for the observed daily real-time sediment concentrations. The SVM-PK 15 model predictions have close to the actual sediment measurements.

For the LibSVM model analysis, as in MLR; SVM-RBF, and SVM-PK, the 3-year data are divided into training and test data. 15 different models (LibSVM 1 to LibSVM 15) were used for LibSVM model studies. Statistical results (RMSE, MAE and R) and inputs of these models are given in Table 1.

According to Table 1, the best results in the LibSVM model belong to the **"LibSVM 11"** model. The results of the LibSVM 11 model are given below. Distribution and scatter graphs of the LibSVM model are shown in Fig. 9b. below, respectively.

The correlation coefficient R = 0.90 was obtained for the graph generated for the test with the LibSVM 11 model results. The LibSVM 11 estimations at the test phase show better results than the other models (SRC, MLR 14, SVM-RBF 7 and SVM-PK 15) values for the observed daily real-time sediment concentrations.

In NF analysis, Gaussian parabolic 3x3x4x4x3x4x3 Membership Functions (MFs) and Grid Partition section were analyzed with 300 iterations, assuming the output as linear. For the NF model analysis, as in MLR; SVM-RBF, SVM-PK and LibSVM the 3-year data are divided into training and test data. 15 different models (NF 1 to NF 15) were used for NF model studies. Statistical results (RMSE, MAE and R) and inputs of these models are given in Table 1.

According to Table 1, the best results in the NF model belong to the **"NF 13"** model. The results of the NF 13 model are given below. Distribution and scatter graphs of the NF 13 model are shown in Fig. 9c. below.

The correlation coefficient R = 0.89 was obtained for the graph generated for the test with the NF 13 model results. The NF 13 estimations at the test phase show better results than the other models (SRC, MLR 14, SVM-RBF 7, and SVM-PK 15) values for the observed daily real-time sediment concentrations. It is seen that NF and LibSVM models have low error rates and a high correlation when a general evaluation is carried out.

5.3. General evaluation

Sediment amounts from 875-day observations used in the training of the NF model were also trained for MLR, SVM models as input set. Then, the models created were applied to the inputs of the test data generated from 220-day observations. The model results were compared with the measured values. The correlation coefficient variation of prediction models are given in Fig. 10.

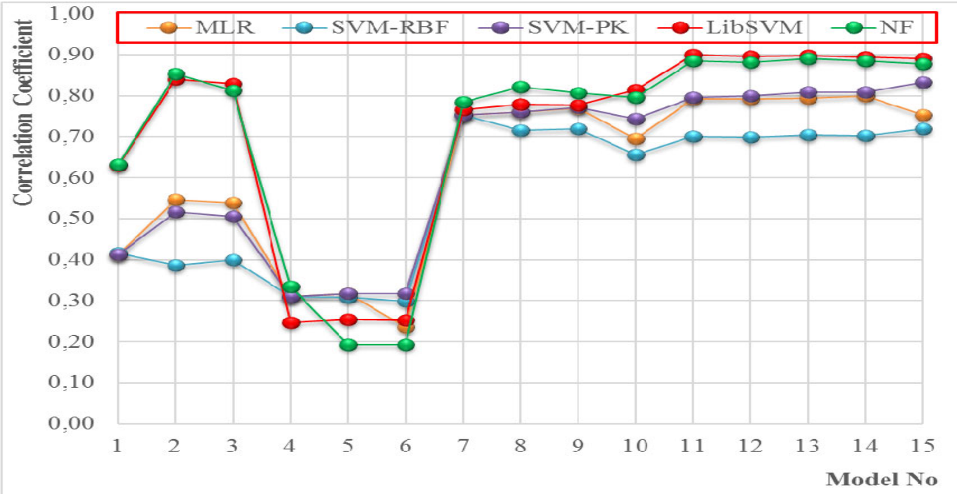

Fig. 10. Correlation coefficient variations according to SRC, MLR, SVM, and NF models

The model with the best result according to Table 1 is obtained when RMSE, MAE have the smallest, and R has the largest value. According to RMSE, MAE, and R, the SRC model (278,77-163,19-0,47) has the lowest success rate. The MLR 14 (230,92-103,42-0,80) model performed better than the SRC model. The SVM-PK 15 (180,72-79,41-0,83) model performed better than the SRC, MLR 14, and SVM-RBF 7 model. The NF 13 (131,25-73,71-0,89) and LibSVM 11 (121,38-64,39-0,90) models were found to perform better than the other classical methods at all estimation performance evaluations.

6. Conclusions

In the present study, the potential of sediment rating curve (SRC), multi linear regressions (MLR), neuro-fuzzy (NF) and support vector machines with radial basis function kernel (SVM-RBF), support vector machines with the poly kernel (SVM-PK), LibSVM for the predicting of daily suspended sediment concentration is questioned by comparing the results with the observed suspended sediment concentration. Daily mean temperature, real-time streamflow, sediment concentration, 3-year data from the Skunk River Augusta station in the US were used. As a result of this study, it is possible to draw the following conclusions:

For the 3-year data, the best results according to the correlation coefficient and error analysis criteria were obtained in the neuro-fuzzy and LibSVM models. But, the LibSVM approach slightly better than the NF model for forecasting daily sediment concentrations. The worst estimation results in all criteria were obtained in the sediment rating curve model.

The MLR model could not find the desired accuracy in the same question due to the nonlinearity of the suspended sediment behavior while explaining empirical relationships.

Support vector machines with poly kernel model has better performance than SRC, MLR, and SVM RBF models

The neuro-fuzzy and LibSVM models whose inputs are the air temperature, streamflow, lagged time stream-flows and suspended sediment concentrations performed the best among the input combinations tried in this paper. This indicates that all these variables are needed for better suspended sediment modeling.

As a result, it is demonstrated in this paper that the neuro-fuzzy and LibSVM models can be an applicable and alternative method for suspended sediment prediction in future studies.

The hydrological data in this study were obtained from the USGS.
The researchers would like to thank the USGS technical team,
the USGS staff involved in measuring and transmitting hydrological data.
Conflicts of interest: none

References

Berk, R. A. (2004). Regression analysis: A constructive critique (Vol. 11). Sage.

Buyukyildiz, M., & 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. DOI: https://doi.org/10.1007/s11269-017-1581-1.

Çatal, Y., & Saplioğlu, K. (2018). Comparison of adaptive neuro-fuzzy inference system, artificial neural networks and non-linear regression for bark volume estimation in brutian pine (Pinus brutia Ten.). *Applied ecology and environmental research, 16*(2), 2015-2027.

Chang, C.C., & Lin, C.J. (2002). Training v-support vector regression: theory and algorithms. *Neural computation, 14*(8), 1959-1977.

Choubin, B., Darabi, H., Rahmati, O., Sajedi-Hosseini, F., & Kløve, B. (2018). River suspended sediment modelling using the CART model: A comparative study of machine learning techniques. *Science of the Total Environment, 615*, 272-281.

Cortes, C., Vapnik, V. (1995) Support-vector networks. *Mach Learn, 20*, 273-297. DOI: https://doi.org/10.1007/BF00994018

Demirci, M., & Baltaci, A. (2013). Prediction of suspended sediment in river using fuzzy logic and multilinear regression approaches. *Neural Computing and Applications, 23*(1), 145-151. DOI: https://doi.org/10.1007/s00521-012-1280-z

Demirci, M., Üneş, F., & Saydemir, S. (2015). Suspended sediment estimation using an artificial intelligence approach. In Sediment matters, 83-95. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-14696-6_6

Demirci, M., Unes, F., Kaya, Y. Z., Mamak, M., Tasar, B., & Ispir, E. (2017). *Estimation of groundwater level using artificial neural networks: a case study of Hatay-Turkey*. In "10th International Conference Environmental Engineering".

Demirci, M., Unes, F., Kaya, Y. Z., Tasar, B., & Varcin, H. (2018). Modeling of dam reservoir volume using adaptive neuro fuzzy method. *Aerul si Apa. Componente ale Mediului*, 145-152.

Ebtehaj, I., & Bonakdari, H. (2017). Design of a fuzzy differential evolution algorithm to predict non-deposition sediment transport. *Applied Water Science, 7*(8), 4287-4299.

Ehteram, M., Ahmed, A. N., Latif, S. D., Huang, Y. F., Alizamir, M., Kisi, O., ... & El-Shafie, A. (2021). Design of a hybrid ANN multi-objective whale algorithm for suspended sediment load prediction. *Environmental Science and Pollution Research, 28*(2), 1596-1611.

Emamgholizadeh, S., & Demneh, R. K. (2019). A comparison of artificial intelligence models for the estimation of daily suspended sediment load: a case study on the Telar and Kasilian rivers in Iran. *Water Supply, 19*(1), 165-178.

Firat, M. & Güngör, M. 2010. Monthly total sediment forecasting using adaptive neuro fuzzy inference system. *Stochastic Environmental Research and Risk Assessment, 24*, 259-270. DOI: https://doi.org/10.1007/s00477-009-0315-1

Ghavidel, S. Z. Z., & Montaseri, M. (2014). Application of different data-driven methods for the prediction of total dissolved solids in the Zarinehroud basin. *Stochastic environmental research and risk assessment, 28*(8), 2101-2118.

Gunawan, T. A., Kusuma, M. S. B., Cahyono, M., & Nugroho, J. (2017). The application of backpropagation neural network method to estimate the sediment loads. In *MATEC Web of Conferences, 101*, 05016. EDP Sciences.

Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification.

Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics, 23*(3), 665-685. DOI: https://doi.org/10.1109/21.256541

Kişi, Ö. (2005). Daily river flow forecasting using artificial neural networks and autoregressive models. *Turkish Journal of Engineering and Environmental Sciences, 29*(1), 9-20.

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. DOI: https://doi.org/10.1007/s11269-016-1405-8

Kitsikoudis, V., Sidiropoulos, E., & Hrissanthou, V. (2015). Assessment of sediment transport approaches for sand-bed rivers by means of machine learning. *Hydrological sciences journal, 60*(9), 1566-1586.

Meshram, S. G., Singh, V. P., Kisi, O., Karimi, V., & Meshram, C. (2020). Application of artificial neural networks, support vector machine and multiple model-ANN to sediment yield prediction. *Water Resources Management, 34*(15), 4561-4575.

Mirbagheri, S. A., Nourani, V., Rajee, T. & Alikhani, A. (2010). Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. *Hydrological Sciences Journal 55,* 1175-1189. DOI: https://doi.org/10.1080/02626667.2010.508871

Partovian, A., Nourani, V., & Alami, M. T. (2016). Hybrid denoising-jittering data processing approach to enhance sediment load prediction of muddy rivers. *Journal of Mountain Science, 13*(12), 2135-2146.

Salih, S. Q., Sharafati, A., Khosravi, K., Faris, H., Kisi, O., Tao, H., ... & Yaseen, Z. M. (2020). River suspended sediment load prediction based on river discharge information: application of newly developed data mining models. *Hydrological Sciences Journal, 65*(4), 624-637.

Saplıoğlu, K., & Çimen, M. (2010). Yapay sinir ağlarını kullanarak günlük yağış miktarının tahmini. *Mühendislik Bilimleri ve Tasarım Dergisi, 1*(1), 14-21.

Sayed, T., Tavakolie, A. & Razavi, A. 2003. Comparison of adaptive network based fuzzy inference systems and B-spline neuro-fuzzy mode choice models. *Journal of Computing in Civil Engineering 17,* 123-130. DOI: https://doi.org/10.1061/(ASCE)0887-3801(2003)17:2(123)

Schölkopf, B., Bartlett, P., Smola, A. & Williamson, R. 1998. *Support vector regression with automatic accuracy control*. In ICANN 98, 111-116. Springer, London.

Seyedian, S. M., & Rouhani, H. (2015). Assessing ANFIS accuracy in estimation of suspended sediments. *Građevinar, 67*(12), 1165-1176.

Taşar, B., Kaya, Y. Z., Varçin, H., Üneş, F. & Demirci, M. (2017). Forecasting of suspended sediment in rivers using artificial neural networks approach. *International Journal of Advanced Engineering Research and Science, 4*(12).

Turhan, E., & Çağatay, H. Ö. (2016). Eksik akım verilerinin tahmin modelinin oluşturulmasında yapay sinir ağlarının kullanımı: Asi Nehri-Demirköprü akım gözlem istasyonu örneği. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 31*(1), 93-106 (in Turkish).

Turhan, E., Keleş, M. K., Tantekin, A., & Keleş, A. E. (2019). The investigation of the applicability of data-driven techniques in hydrological modeling: The case of seyhan basin. *Rocznik Ochrona Środowiska, 21*, 29-51.

Üneş, F., Joksimovic, D. & Kisi, O. 2015. Plunging flow depth estimation in a stratified dam reservoir using neuro-fuzzy technique. *Water Resources Management 29,* 3055-3077. DOI: https://doi.org/10.1007/s11269-015-0978-y

Üneş, F., Doğan, S., Taşar, B., Kaya, Y., & Demirci, M. (2018a). The Evaluation and Comparison of Daily Reference Evapotranspiration with ANN and Empirical Methods. *Natural and Engineering Sciences, 3*(3), 54-64.

Üneş, F., Bölük, O., Kaya, Y. Z., Taşar, B., & Varçin, H. (2018b). Estimation of Rainfall-Runoff Relationship Using Artificial Neural Network Models for Muskegon Basin. *International Journal of Advanced Engineering Research and Science, 5*(12), 198-205.

USGS.gov | Science for a changing world [WWW Document], n.d. URL https://www.usgs.gov/

Wieprecht, S., Tolossa, H. G., & Yang, C. T. (2013). A neuro-fuzzy-based modelling approach for sediment transport computation. *Hydrological sciences journal, 58*(3), 587-599.

Williams, J. R. (1978). A sediment graph model based on an instantaneous unit sediment graph. *Water Resources Research, 14*(4), 659-664. DOI: https://doi.org/10.1029/WR014i004p00659