Water and Environment Journal
Promoting Sustainable Solutions

CIWEM

Water and Environment Journal. Print ISSN 1747-6585

Online sequential extreme learning machine in river water quality (turbidity) prediction: a comparative study on different data mining approaches

Mohammad Zounemat-Kermani [1], Meysam Alizamir [2], Marzieh Fadaee[1], Adarsh Sankaran Namboothiri[3] & Jalal Shiri[4]

[1]Water Engineering Department, Shahid Bahonar University of Kerman, Kerman, Iran; [2]Department of Civil Engineering, Hamedan Branch, Islamic Azad University, Hamedan, Iran; [3]Department of Civil Engineering, TKM College of Engineering Kollam, Kollam, India; and [4]Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

Keywords
aquatic systems; data mining; environmental engineering; OSELM; soft computing.

Correspondence
Mohammad Zounemat-Kermani, Water Engineering Department, Shahid Bahonar University of Kerman, Kerman, Iran.
Email: zounemat@uk.ac.ir

doi:10.1111/wej.12630

Abstract

As a measure of water quality, water turbidity might be a source of water pollution in drinking water resources. Henceforth, having a reliable tool for predicting turbidity values based on common water quantity/quality measured parameters is of great importance. In the present paper, the performance of the online sequential extreme learning machine (OS-ELM) in predicting daily values of turbidity in Brandywine Creek, Pennsylvania, is evaluated. For this purpose, in addition to the developed OS-ELM, several data-driven models, that is, multilayer perceptron neural network (MLPANN), the classification and regression tree (CART), the group method of data handling (GMDH) and the response surface method (RSM) have been applied. The general findings of the study confirm the superiority of the OS-ELM model over the other applied models so that the OS-ELM improved the averaged RMSE of the predicted values 9.1, 11.7, 20.5 and 29.3% over the MLPANN, GMDH, RSM and CART models, respectively.

Introduction

Ensuring the quality of river water is one of the prime concerns of hydrologists and environmental modelers. Turbidity is the major visible indicator of the quality of river systems, existence (or non-existence) of aquatic habitat and aesthetic destructions in surface waters. Accurate modelling of turbidity is highly complex as it is an outcome of the joint influence of many inter-dependent parameters (Iglesias *et al*., 2014). It is a primary indicator of water quality and high turbidity values normally indicate the high value of other water quality parameters and pollutant indicators, like chemical oxygen demand (COD), ammonium, sulfate and nitrate (Vigil, 2003; Díaz Muñiz *et al*., 2012). Turbidity is a light-scattering property of water and the automatic monitoring of turbidity using a turbidimeter is expensive in initial investment, the necessity of frequent repair and maintenance, use of toxic compounds like formazine and poor accuracy in capturing low turbidity values. Lack of understanding about the physical properties of underlying mechanisms is one of the major constraints in performing physical modelling of turbidity levels in water bodies. Therefore, any effort for accurate modelling of turbidity using easily measurable river hydraulic (e.g. water discharge) and environmental parameters (e.g. water temperature) needs to be investigated.

Factors influencing the river water turbidity include hydrologic factors, meteorological factors and water quality parameters (Iglesias *et al*., 2014). There are several key parameters influencing the fate of the turbidity levels in rivers over time. The primary visible factors influencing the turbidity in rivers are the suspended sediment concentration (SSC) and the river discharge (Q). In addition to natural catchment erosion, construction and agricultural activities also lead to frequent loading of sediments to river and lake systems (USEPA, 2009). The increased sediment loading enhances the SSC levels, which eventually increases the turbidity levels in the surface water systems (Neupane *et al*., 2015). Numerous studies on the prediction of SSC using machine learning methods have already been reported (e.g. Rajaee *et al*., 2011; Kisi, 2012; Zounemat-Kermani *et al*., 2016, Zounemat-Kermani, 2017; Choubin *et al*., 2018; Adnan *et al*., 2019a).

River flow (discharge) is another crucial element in initiating and developing turbidity. However, in the absence of discharge measurements, the use of rainfall-runoff models could be helpful in investigating the rainfall-turbidity relationships (Beaudeau *et al*., 2001). Beaudeau *et al*. (2001) used ANN for prediction of turbidity levels a drinking water supply well in France, located at a karst aquifer for summer and winter seasons, considering precipitation measurements at the heart of the catchment

as input. They reported remarkable improvement in predictions up to 12 hours lead time when compared with the classical Box-Jenkin time series model. Iglesias *et al*. (2014) used ANN for turbidity predictions in the Grullos Nalón river basin (Northern Spain) considering electrical conductivity (EC), dissolved oxygen (DO), pH, ammonium content (NH_3) and water temperature (T) as predictor variables. They identified temperature as the most significant predictor variable for turbidity prediction and the introduction of a new synergetic variable defined as the product of two original variables could improve the turbidity predictions. Nieto *et al*. (2014) employed a hybrid support vector machine particle swarm optimisation (SVM-PSO) approach for turbidity prediction of the Nalón river basin (Northern Spain). Yang *et al*. (2014) presented an ANN model as well as a warning system for the prediction and management of turbidity levels at the Taipei Water Source Domain using rainfall as an input variable. The accumulative rainfall records of upstream locations are considered as inputs for downstream turbidity predictions. Subsequently, the observed and ANN model results were applied for developing the EWS. Mather and Johnson (2015) applied the conjunctive use of the cluster analysis and classification tree for the prediction of streamflow event model parameters at three moderately-sized streams in the U.S. considering many hydrological and meteorological variables. Subsequently, they predicated turbidity using streamflow, air temperature and precipitation as inputs. It was reported that the ratio of initial event discharge to the peak event discharge, as well as antecedent moisture conditions (AMC) of the catchment, are the dominant factors in the event turbidity predictions. Mather and Johnson (2016) performed 3-day ahead forecasts of turbidity during streamflow events for two streams in the United States using the same input variables as in Mather and Johnson (2015). They reported that the turbidity forecast error was independent of lead time and considerably less than the threshold error for all the lead times considered in the study. Delphla *et al*. (2018) conjunctively used a trend analysis and ANN for forecasting of mean and peak turbidity levels of the main drinking water source of Québec, Canada. ANN was used to forecast daily turbidity values considering AMCs as inputs. They reported that trend analyses helped to capture the time of occurrences of turbidity peaks in advance, which improved the accuracy in the prediction of mean turbidity levels. Stevenson and Cristian (2019) developed a decision support system capable of predicting turbidity events up to seven days in advance considering operational, meteorological and hydro-geological factors as data inputs. They applied the generalized linear model (GLM) and random forests (RF) model for predicting turbidity peaks. The RF model was found to be slightly better than the GLM in capturing the nonlinearities in the data sets. Savary *et al*. (2020) presented the scope for multi-resolution analysis in turbidity predictions at a port located in eastern Normandy, France. The method followed a hybrid framework involving a decomposition tool and ANN for modelling. They concluded that the forecast accuracy of the developed framework depends on precision in addressing the 'boundary effects'.

The advancements in novel computational tools and machine learning circumvent this limitation in modelling hydrological and environmental engineering (Adnan *et al*., 2020; Fadaee *et al*., 2020; Zounemat-Kermani *et al*., 2020). Such applications include a wide variety of problems including the prediction of hydrological variables such as precipitation, evaporation, streamflow and environmental hazards (Choubin *et al*., 2019a; 2019b). Recently, developed algorithms and practices such as Extreme Learning Machine (ELM), classification and regression tree (CART), and hybrid machine learning methods are gaining popularity for addressing different types of environmental challenges with an improved degree of accuracy (Adnan *et al*., 2019b; Choubin *et al*., 2019b). Nevertheless, the application of novel ML models in modelling turbidity level in water bodies is considered a research gap in the literature so that the artificial neural networks (ANNs), as the most commonly used data mining paradigm, have found their applications in turbidity predictions. Besides, from the review of literature, it is noted that even though hydrological, meteorological and water quality parameters are responsible for turbidity in water bodies, the role of an appropriate combination of such parameters is not yet investigated for the turbidity predictions. Also, even though the machine learning methods, especially ANNs, are applied for the prediction of turbidity in water bodies, the complete potential of many advanced mining methods have not been fully explored for turbidity predictions. Therefore, the specific objectives of the current study are formulated as (i) to investigate the role of appropriate combinations of water quality parameters and hydrologic data for turbidity predictions of Brandywine Creek, USA; and (ii) to compare the performance of novel machine learning methods, that is, the online sequential extreme learning machine (OS-ELM) versus the more commonly used data-driven methods such as the perceptron ANN, classification and regression tree (CART), group methods of data handling (GMDH) and response surface method (RSM) for turbidity predictions.

The next section presents the theoretical background of different data mining algorithms used in the study. Thereafter, the description of the study area and data statistics are presented. Subsequently, the modelling scenarios and model development details are provided. Results

and discussion are given thereafter and in the final section, important conclusions drawn from the study are presented.

Methods

In this research, methods include three types of common machine learning (ML) models including the most public type of artificial neural network (ANN), namely the multilayer perceptron (MLP), the commonly used decision tree model, namely classification and regression tree (CART) model, and inductive neural network, namely the group method of data handling (GMDH). In addition to the mentioned ML models, the novel method of extreme learning (ELM) machine, namely the online sequential ELM (OS-ELM) is considered for this study. In order to cover a wider range of types of predictive models for reaching a more precise and robust conclusion, the response surface method (RSM), which is a statistical data-driven method, is applied alongside the mentioned ML models. In the following, the predictive data mining methods applied in the current study are described.

Multilayer perceptron neural network

The multilayer perceptron neural network (MLPANN) is known as the one type of feedforward neural network (FFNN), which due to high-speed operating, its simple implement and requirements of the smaller training sets, is widely used. The standard MLPANN comprises three sequentially layers, namely the input layer, including the independent variables, the hidden layer with several numbers of processing neurons that are found by trial and error, and the third layer, output layer that includes the dependent variables. Figure 1 illustrates the basic structure of the MLPANN model.

In the study, an MLPANN model with a single hidden layer is used. Each neuron sums its input signals affecting on it after multiplying them by their mutual synaptic weights in the hidden layer. The general mathematical equation of the MLPANN is presented as bellows:

$$y_j = f\left(\sum w_{ji} x_i\right) \tag{1}$$

where j is a neuron in the hidden layer, x_i is the input variable, w_{ji} is the weight between the input i and neuron j, f is an activation function using the weighted sums of the inputs and y_j is the final output of neuron j. The sum of squared differences between the real and desired amounts of the output neurons E is described as bellows:

$$E = \frac{1}{2}\sum_j \left(y_{dj} - y_j\right)^2 \tag{2}$$

where y_{dj} and y_j are the desired and real values of output neuron j. Relying upon the adopted training algorithm, each w_{ji} is set to minimize the value of E. Several algorithms exist to train a multilayer perceptron. In this study, the backpropagation algorithm is applied, which due to its limitations, supported by the Levenberg–Marquardt algorithm (Zounemat-Kermani, 2012).

Classification and regression tree

The classification and regression tree (CART), which was suggested by Breiman *et al*. (1984), is one type of

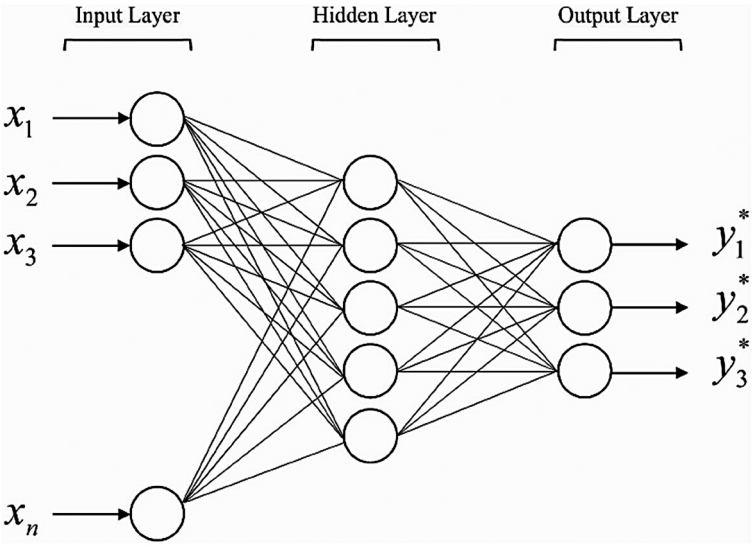

Fig. 1. The basic structure of the MLPANN model used in the study.

non-parametric methods of the decision tree methodology that is successfully applied in data mining and predicting engineering issues. Establishing a model in order to predict the dependent variables using the independent variables is the principal aim of decision trees. The decision trees is a rule-based approach that mimics the natural trees formed by branches, leaves, roots and nodes. There are three principal parts in this method, that is, node, production and condition; also the nodes are composed of decision, chance and end nodes (Fig. 2). Every single node in this method demonstrates a definite feature or an independent input variable; moreover, the nodes are able to connect to qualitative or quantitative variables to each other.

The CART algorithm is able to self-identify the critical variables and does not require initial assumptions about the variables. The major points of any decision tree algorithm, such as CART, can be interpreted as below:

(1) Starting the tree building using several rules of splitting data at each node based on the value of one variable,
(2) Existence of criteria to stop the tree-building process which is recognized as production node,
(3) Estimation of end nodes' predicted values.

The CART algorithm's process commences with the choice of a variable as the root of a tree (node #1). By asking the yes/no question about the range of this variable, the process is continued. After that, the process is separated by branches into sub-nodes according to the chosen answer. This method is continued until stopping rules are met. Ultimately, the end nodes determine the predicted values by the CART method.

Group method data handling

The group method data handling (GMDH) was introduced by Ivakhnenko (1968) in order to solve complicated network problems such as prediction. This method relies upon high-order polynomial input variables and solving the problem uses regression analysis. In this approach, the general input-output variables' relationship is represented by the Volterra-Kolmogorov-Gabor polynomial series form:

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{i=1}^{M}\sum_{j=1}^{M} a_{ij} x_i x_j + \sum_{i=1}^{M}\sum_{j=1}^{M}\sum_{k=1}^{M} a_{ijk} x_i x_j x_k \quad (3)$$

where y is the output, M is the number of independent variables, $X\left(x_1, x_2, \ldots, x_M\right)$ are the input variables vector and $A\left(a_1, a_2, \ldots, a_M\right)$ are the vector of polynomial coefficients.

The components of the input vector X can be introduced as independent variables, functional forms, or finite difference terms. Also, the other functions (e.g. harmonic, logistic, difference and probabilistic) can be used. In the basic combinational GMDH, these functions are types of first-order (linear) or second-order polynomial functions of two or three independent variables:

$$f_1\left(x_i, x_j\right) = a_0 + a_1 x_i + a_2 x_j \quad (4)$$

$$f_2\left(x_i, x_j\right) = f_1\left(x_i, x_j\right) + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \quad (5)$$

The second phase is to build the linear combination of the polynomial terms based on the variable coefficients

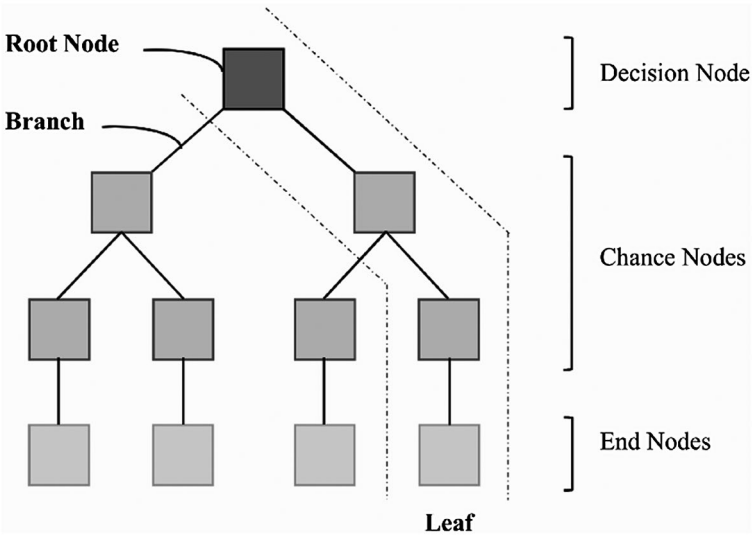

Fig. 2. The basic structure of a regression tree.

by minimising the squared sum of deviations amongst model predictions and sample outputs.

GMDH operates by creating successive layers with complex connections. Taking into account the input variables, each layer is formed by computing regressions of the values in the former layer. The best new input variables are retained and utilized for the next layer to create newer input variables of higher-order. This process will continue until a prespecified selection measure is met, then the new combined variables will be stored and the algorithm will stop.

Online Sequential Extreme Learning Machine (OS-ELM)

The online sequential extreme learning machine (OS-ELM) method is based on the classical feedforward neural network structure that was recommended by Liang *et al*. (2006). In this method, the data set, which consists of N arbitrary independent samples, is used for training with L number of hidden nodes. The target output is determined as below (Alizamir *et al*., 2018):

$$f_L\left(x_j\right)=\sum_{i=1}^{L}\beta_i G\left(a_i,b_i,x_j\right)=t_j, j=1,2,\ldots,N \tag{6}$$

where a_i and b_i are the learning parameters of the hidden nodes, β_i is the output weight vector and $G\left(a_i,b_i,x_j\right)$ is the activation function, which for the additive hidden node is defined as:

$$G\left(a_i,b_i,x_j\right)=g\left(a_i.x_j+b_i\right), b_i\in R \tag{7}$$

where a_i and b_i are the input weight vector and the bias of the i-th hidden node, respectively, and $a_i.x_j$ stands for the inner product of the two.

Training of the OS-ELM approach involves two phases, called the initialisation phase and online sequential learning phase. In the initialisation phase, a small training data set chunk (M_0) with hidden neurons (L) is used to training OS-ELM, where the data $M=\left\{\left(x_i,t_i\right)|x_i\in R^n, t_i\in R^m, i=1,\ldots,N\right\}$, $M_0=\left\{x_i,t_i\right\}_{i=1}^{N_0}\subset M$, and $N_0\geq L$.

This phase involves four steps:

(1) Assign randomly the input weights a_i and biases b_i, $i=1,\ldots,L$.

(2) Determine the initial hidden layer output matrix (H_0):

$$H_0=\begin{bmatrix} G\left(a_1,b_1,x_1\right) & \cdots & G\left(a_L,b_L,x_L\right)\\ \vdots & \cdots & \vdots\\ G\left(a_1,b_1,x_{N_0}\right) & \cdots & G\left(a_L,b_L,x_{N_0}\right)\end{bmatrix}_{N_0\times L} \tag{8}$$

(3) Determine the initial output weights ($\beta^{(0)}$)

$$\beta^{(0)}=P_0H_0^TT_0 \tag{9}$$

where $T_0=\left[t_1,\ldots,t_{N_0}\right]_{N_0\times m}^T$, $P_0=(H_0^TH_0)^{-1}$ and $K_0=P_0^{-1}=H_0^TH_0$

(4) Set $k=0$, where k denotes the number of chunks in the network

In the second phase, online sequential learning phase, a new set of observation data ((k + 1)th) has arrived.

$$M_{k+1}=\left\{\left(x_i+t_i\right)\right\}_{i=\left(\sum_{j=0}^{k}N_j\right)+1}^{\sum_{j=0}^{k+1}N_j} \tag{10}$$

where N_{k+1} indicates the number of observations in the $(k+1)$th chunk. This phase involves three steps:

(1) Determine the partially hidden layer output matrix (H_{k+1})

$$H_0=\begin{bmatrix} G\left(a_1,b_1,x_{\left(\sum_{j=0}^{k}N_j\right)+1}\right) & \cdots & G\left(a_L,b_L,x_{\left(\sum_{j=0}^{k}N_j\right)+1}\right)\\ \vdots & \cdots & \vdots\\ G\left(a_1,b_1,x_{\sum_{j=0}^{k+1}N_j}\right) & \cdots & G\left(a_L,b_L,x_{\sum_{j=0}^{k+1}N_j}\right)\end{bmatrix}_{N_{k+1}\times L} \tag{11}$$

(2) Determine the output weights ($\beta^{(k+1)}$)

$$\beta^{(k+1)}=\beta^{(k)}+K_{k+1}^{-1}H_{k+1}^T\left(T_{k+1}-H_{k+1}\beta^{(k)}\right) \tag{12}$$

where $T_{k+1}=\left[t_{\left(\sum_{j=0}^{k}N_j\right)+1},\ldots,t_{\sum_{j=0}^{k+1}N_j}\right]_{N_{k+1}\times m}^T$, $K_{k+1}=K_k+H_{k+1}^TH_{k+1}$.

In the recursive process, to avoid inverting matrices, the least-squares solution defined as below:

$$K_{k+1}^{-1}=K_k^{-1}-K_k^{-1}H_{k+1}^T\left(I+H_{k+1}K_k^{-1}H_{k+1}^T\right)^{-1}H_{k+1}K_k^{-1} \tag{13}$$

Since $P_{k+1}=K_{k+1}^{-1}$,

$$P_{k+1}=P_k-P_kH_{k+1}^T\left(I+H_{k+1}P_kH_{k+1}^T\right)^{-1}H_{k+1}P_k$$

$$\beta^{(k+1)}=\beta^{(k)}+P_{k+1}H_{k+1}^T\left(T_{k+1}-H_{k+1}\beta^{(k)}\right) \tag{14}$$

(1) Set $k=k+1$

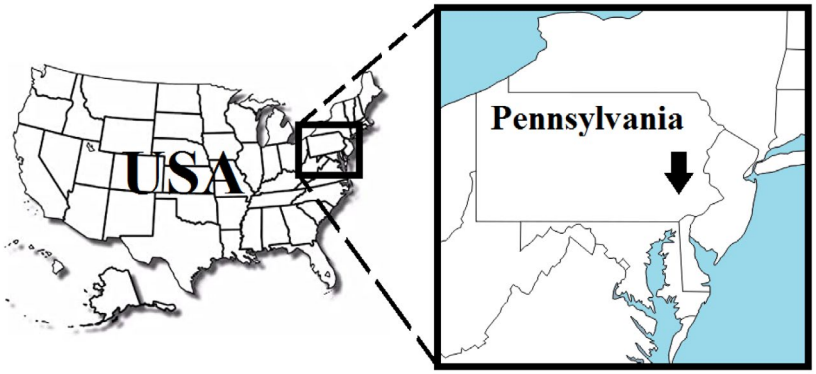

Fig. 3. The geographical position of the Brandywine Creek Station located in Chester County, Hydrologic Unit 02040205.

This process is repeated by going back to step 2 of the online sequential learning phase until the last chunk has arrived (Huang *et al*., 2005; Lima *et al*., 2016).

Response surface method

The response surface method (RSM) is a combination of statistical and mathematical approaches that was improved to analyse engineering issues and is widely used for process optimisation. For structural reliability analysis, Bucher and Bourgund (1990) proposed an efficient approximation to build a modified version of the RSM. In the principal form of the RSM, to approximate true limit state functions, polynomials are used. The response surface, approximated by a polynomial model including quadratic terms and without the cross-terms is formulated as below:

$$\widehat{G}(X) = a + \sum_{i=1}^{n} b_i x_i + \sum_{i=1}^{n} c_i x_i^2 \tag{15}$$

where $\widehat{G}$ stands for the quadratic polynomial and a, b_i and c_i are the polynomial constants. For each variable, the samples are taken along the coordinates axes at the $x_i = T_i(\pm h)$ distance, where T_i donates the probabilistic transformation of x_i, μ_i and σ_i are the mean value and standard deviation, and h is an arbitrary factor. Using Eq. (16), the new position of the sample point is determined by adaptive linear interpolation:

$$\mu_X^{i+1} = \mu_X^i + \left(X_i^* - \mu_X^i\right) \frac{G\left(\mu_X^i\right)}{\left[G\left(\mu_X^i\right) - G\left(X_i^*\right)\right]} \tag{16}$$

where μ_X^{i+1} donates the new centre point, μ_X^i shows the mean value of the random variables at the central point of samples and X_i^* indicates the approximated design point on the real space.

Study area and data source

In the present study, daily river data from Brandywine Creek with a drainage area equal to 233 km^2, which is a tributary of the Christina River in southeastern Pennsylvania in the USA, has been utilized for executing the models. From the geology point of view, the Brandywine Creek basin is covered mostly with crystalline rocks and carbonate rocks in the north-central part. Besides, land use areas in the Brandywine basin include about 20% of urban uses, agriculture (45%) and wetland/forest (35%) (http://delawarewatersheds.org). Figure 3 shows the location of the established station.

The water quantity and quality data of the study area were collected and presented by the Chester County Water Resources Authority in association with the U.S. Geological Survey and required daily data sets were gathered from the USGS webpage from 11 March 2009 to 12 June 2012. Obtained data include river discharge, precipitation, water temperature, specific conductance, dissolved oxygen, pH and turbidity. Table 1 represents the input and output water of parameters and their related statistical values.

In Fig. 4, the historical variations, as well as the boxplot diagram of the measured turbidity values, during the whole time period is shown. As can be seen in Fig. 4a, the turbidity follows an irregular fluctuation trend. Besides, the illustrated boxplot in Fig. 4b implies that the measured data does not follow a normal distribution, and there exist a lot of outliers in the data registered. This simply means that the turbidity time-series data have a chaotic or nonlinear nature, and therefore, proper modelling of the data needs sophisticated and resilient models. In this study based on a random split, 80% of the river data are selected for the training of the models and the remaining 20% data series are used for Validation of the models' performance (the validation set).

Table 1 The statistical summary of the input and output data used in this study

Data		Parameter	Symbol	Range	Mean	St.d.
Input parameters	Quantity river data	Discharge	Q (m^3/s)	0.7–107.6	4.96	7.15
		Precipitation	P (mm/d)	0.00–105.16	3.52	10.13
	Quality river data	Specific conductance	Sc (μs/cm)	118–475	316.90	57.59
		Pondus hydrogenii	pH	6.5–8.2	7.60	0.23
		Suspended sediment	SS (kg/s)	0.07–4100.00	30.48	187.12
		Dissolved oxygen	DO (mg/L)	6.9–13.6	9.7	1.36
		Water temperature	T (C°)	2.4–27.3	16.18	5.66
Output parameter		Turbidity	Tu (FNU)[a]	0.3–160.0	7.85	16.27

[a]FNU = Formazin Nephelometric Unit which is an index for turbidity estimation in water measured with an infrared light source.

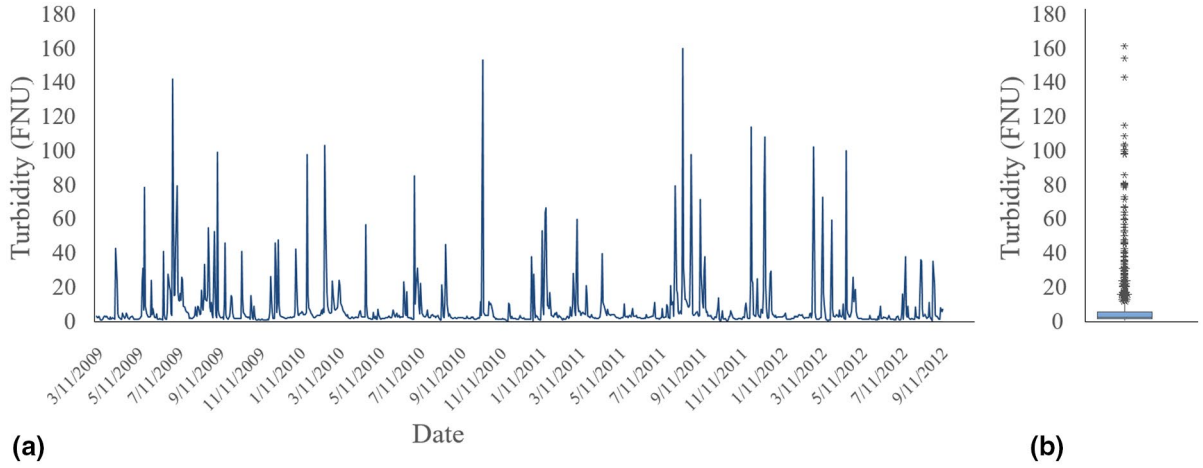

Fig. 4. (a) Historical variation curves of the measured turbidity at the Brandywine Creek Station; (b) boxplot diagram of the measured data.

Table 2 The Pearson's correlation coefficients (PCC) and corresponding p-values between the independent and target (turbidity) parameters

Parameter	Q	SC	pH	T	DO	P	SS
PCC	0.773	–0.521	–0.385	0.001	–0.085	0.555	0.762
P-value	0.000	0.000	0.000	0.968	0.007	0.000	0.000

Modelling scenarios and data preparation

Before submitting the input data into the predictive models, the input data series should be analysed in order to determine the effective independent parameters on the output variable (herein as to turbidity). Table 2 presents the corresponding Pearson's correlation coefficients and the correspondence p-value between the input and the output values.

The PCCs in Table 2 indicate that the turbidity is mostly affected by the discharge and suspended sediment and reversely affected by the specific conductivity (SC). Moreover, the PCC value corresponds to the water temperature (PCC = 0.001) implies that this parameter is not correlated to the turbidity at all, though the high value for the *P*-value (>0.05) reveals the fact that this conclusion cannot be completely supported based on the gathered range of water temperature. Hence, in this study, the coefficients of Mallows' Cp and R-squared have been considered for determining the best input vectors to the predictive models based on different input scenarios (Table 3). As can be seen in Table 3, three input scenarios including the univariate (Q as the input component), bivariate (Q and SS as the input components) and multivariate (Q, SS, SC, DO and P as the input components) are chosen for constructing the models. It is expected that the multivariate scenario – with the lowest value for the Mallow's Cp (=0.8) and the highest value of R-squared (= 70.1) – gives the best predictions in comparison to the univariate and bivariate scenarios.

Application and results

Tables 4–8 present the statistical indicators of the applied models during the training and validation periods. Focusing on the statistical indices, all the applied models provided the most accurate results with scenario-2, where flow

Table 3 The values of Mallows' Cp and R-squared coefficients for determining the best input vectors from all the available input parameters (see Table 1)

Number of Variables	Modelling scenario	R-Sq	Mallow's Cp[a]	Q	SC	pH	T	DO	P	SS
1	1 – Univariate	59.8	340.3	✓						
2	2 – Bivariate	65.3	159.1	✓						✓
7	3 – Multivariate	70.1	4.6	✓	✓			✓	✓	✓

[a]The written Mallow's Cp values are the smallest ones amongst all the possible univariate, bivariate and multivariate subsets. The smallest value of Mallow's Cp implies that the related scenario is comparatively more precise than the other ones.

Table 4 The calculated benchmark indicators for the evaluation of the OS-ELM models

			Training			Validation		
Scenario	Input combination	Computational time (s)	RMSE (FNU)	NSE	R^2	RMSE (FNU)	NSE	R^2
I	Q	0.031	9.612	0.636	0.636	11.031	0.604	0.609
II	Q, SS	0.046	3.858	0.941	0.941	4.431	0.936	0.936
III	Q, SC, DO, P, SS	0.062	3.696	0.946	0.946	5.062	0.916	0.921

Table 5 The calculated benchmark indicators for the evaluation of the RSM models

			Training			Validation		
Scenario	Input combination	Computational time (s)	RMSE (FNU)	NSE	R^2	RMSE (FNU)	NSE	R^2
I	Q	0.218	9.829	0.619	0.619	11.557	0.566	0.577
II	Q, SS	0.227	6.865	0.814	0.814	6.217	0.874	0.877
III	Q, SC, DO, P, SS	0.312	5.133	0.896	0.896	7.049	0.838	0.851

Table 6 The calculated benchmark indicators for the evaluation of the CART models

			Training			Validation		
Scenario	Input combination	Computational time (s)	RMSE (FNU)	NSE	R^2	RMSE (FNU)	NSE	R^2
I	Q	1.308	7.333	0.788	0.788	12.261	0.511	0.516
II	Q, SS	2.484	3.474	0.952	0.952	6.190	0.875	0.876
III	Q, SC, DO, P, SS	3.696	2.784	0.969	0.969	7.960	0.794	0.802

Table 7 The calculated benchmark indicators for the evaluation of the MLPANN models

			Training			Validation		
Scenario	Input combination	Computational time (seconds)	RMSE (FNU)	NSE	R^2	RMSE (FNU)	NSE	R^2
I	Q	11.046	9.560	0.640	0.640	10.665	0.630	0.636
II	Q, SS	12.890	3.156	0.960	0.960	5.379	0.906	0.906
III	Q, SC, DO, P, SS	18.171	5.109	0.897	0.897	6.309	0.870	0.871

discharge and suspended sediment load values were utilized as the input variables, except the GMDH model which provided the best simulations of turbidity by the third scenario (using five variables in the input vector). Comparing the applied models, the OS-ELM models gave the most accurate results with the lowest RMSE (4.431 FNU) and the highest NSE (0.936) and R^2 (0.936) values, followed by the MLPNANN, GMDH, CART and RSM models.

The correlation analysis amongst the applied inputs and models' errors provided some remarkable statements. First, except, OS-ELM, the model's errors did not show

Table 8 The calculated benchmark indicators for the evaluation of the GMDH models

Scenario	Input combination	Computational time (seconds)	Training			Validation		
			RMSE (FNU)	NSE	R^2	RMSE (FNU)	NSE	R^2
I	Q	0.084	9.697	0.629	0.63	11.123	0.598	0.602
II	Q, SS	0.140	6.973	0.808	0.812	5.982	0.883	0.884
III	Q, SC, DO, P, SS	0.274	6.403	0.838	0.838	5.829	0.889	0.891

considerable correlation with flow discharge records. Further, though specific conductance (SC) is not an input variable in the univariate and bivariate models, it showed high correlation values with the models' errors for all the models, except MLPNANN and OS-ELM, where the correlation was slightly lower. This might be linked to the importance of this variable in modelling water turbidity and its influence on modelling performance accuracy. Similar statements can be reported for precipitation and suspended sediment. In the latter case, the third scenario that comprises five independent variables showed exclusively lower correlations with suspended sediment. Hence, it might be concluded that, though the inclusion of a variable as an input parameter has increased the performance accuracy, the models include the shadow effects of this parameter in the target variable. Nevertheless, these parameters usually form the turbidity characteristic of water, so it might be expectable to see such correlation even with excluding some variables from the input matrix.

Figure 5 displays the scatter plots of the observed versus predicted values of the turbidity for the applied models during the validation period.

Similar to the previous statements, the simulated turbidity values by the OS-ELM model showed the closest agreement with the corresponded observed values. Nonetheless, all the models show an underestimation trend for higher turbidity values, while they overestimated its lower magnitudes. This might be anticipated as the standard deviation values of this parameter were high (with respect to its average values) in the data set (Table 1). It is observed from Fig. 5 that the peak estimated turbidity values (>50 FNU) are not as close to the agreement line as the lower turbidity values (<50 FNU). One potential reason for this shortcoming may be due to the presence of a small number of higher turbidity values in the training data sets (2.6% of the total training data). Nevertheless, it is worth noting that most of the data-driven hydrological models, that is, modelling river flow, sediment transportation and water quality parameters encounter the same drawback for simulating/predicting peak values. Several methods such as appropriate data transformation (Sudheer *et al*., 2013), using wavelet transform machine learning models (Adamowski and Sun, 2010) and low and high values separation (Corzo and Solomatine, 2007) have been proposed for improving the simulation/prediction peak values. Thus, it is recommended to consider these techniques to evaluate the models' performance in predicting high values of turbidity. Figure 6 illustrates the bar charts of the RMSE index related to the predicted values of all the modelling scenarios. It can be seen that the OS-ELM-2 provided the lowest RMSE values, which is an indication of its superiority.

Figures 7–9 present the residuals of the models relying on all three adopted scenarios. In the case of the univariate models, the highest scatter of the residuals belonged to the CART (with a mean residual value of 0.57 and a standard deviation of 12.278), while the GMDH gave the lowest values for the mean (0.262) and standard deviation (11.14) of the residuals. Attending to the bivariate models, the lowest and highest scatters of the residuals corresponded to the OS-ELM-2 (mean residual = 0.052; standard deviation = 4.442) and CART-2 (mean of residuals = 0.187; standard deviation = 6.202) models. Finally, for the third scenario (multivariate models), the OS-ELM-3 showed the highest residual scatters (mean = –3.703; standard deviation = 14.413), whereas the GMDH-3 (mean = –0.094; standard deviation = 5.843) had the lowest residual scaters. Comparing the three scenarios, the residual values showed the most scatters for the multivariate model, though it surpasses the univariate and bivariate models from the error indices viewpoint. This might show the variation of modelling performance amongst different events that have been experienced by the models when dealing with the available patterns in the data matrix. So, this would dictate that, besides giving optimum results, multivariate models have encountered serious problems in simulating some specific turbidity values that were probably the extreme values.

Discussion

The outcomes for finding the best input vector for constructing the models (see Table 3) shows that the water temperature (T) and pH do not have any effect on the variations of turbidity. Moreover, the flow rate (Q) and suspended sediment concentration (SS) have the most influence on the turbidity values (see Tables 2 and 3). The findings of this study also revealed that constructing the models based on

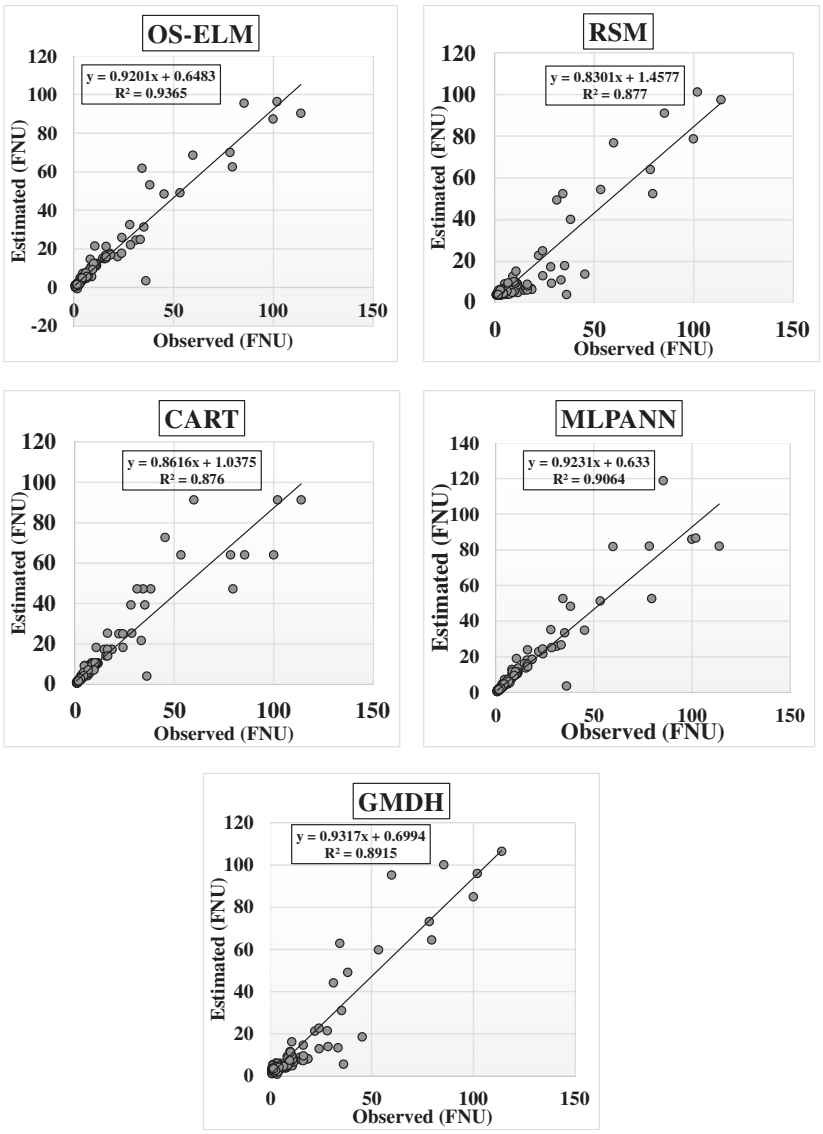

Fig. 5. Comparison of estimated values of proposed models versus observed data.

the input values of river discharge (Q) and suspended sediment concentration (SS) (herein known as scenario-2) surpassed the univariate scenario (Q as the input variable) and the multivariate scenario (Q, SC, DO and P as the input variables). This conclusion is in contrast to the preliminary judgement obtained from the results of choosing the best subset parameters (optimum input vector) using Mallow's Cp values (see Table 3). In other words, despite the fact that the multivariate scenario (3rd scenario) has the lowest value for the Mallow's Cp, all the predictive models – with the exception for the GMDH model – acted better using Q and SS as input parameters. The main potential reason for this paradox lies in the different utilized approached of linear (Mallow's Cp Test) and nonlinear (all the applied data mining codes) models in capturing the nature of data series and mapping the relationship between the independent (predictants) and dependent (predicted) variables. In contradiction to the results of Iglesias *et al*. (2014), this study showed that temperature has not a positive effect on the turbidity prediction. The reasons behind this paradox could be explained as the following items. First, in the present study, two types of quantitative (discharge and precipitation) and qualitative (Sc, pH, SS, DO, & T) input parameters were considered for the prediction of turbidity, whereas in the Iglesias *et al*. (2014) work, only the effect of qualitative parameters was considered. In other words, in the current study, the influence of temperature has been already reflected in the variations of river discharge, which is the main parameter on turbidity. Second, according to the results of Table 2, there was not any correlation between the temperature and

turbidity in this study area, which reveals the fact that in this specific basin, the water temperature would not play any direct effect on the variation of turbidity. Nevertheless, the same conclusion cannot be extended to other study areas. Gelda *et al*. (2013), moreover, argued that most of the turbidity sources in streams belong to surface runoff (especially the extreme runoff) events that carry vast amounts of fine particles (mostly as clay) and can be considered as sources of suspended sediment particles, too. From the hydraulic viewpoint, stream power that is mainly governed by flow discharge (Q) is a major attribute to make scouring/deposition of river bed/bank particles, which causes turbidity in the stream. Hence, in comparison to the multivariate model, the accurate performance of the bivariate model (comprising Q and SS) was not surprising, because increasing the input variable vectors in the training matrix might inversely affect the general performance due to the inclusion of redundant variables.

Another interesting point that was obtained from the results of this study was the fact that considering the precipitation did not help the predictive models to provide better results in the turbidity prediction. Though precipitation (mostly in form of rainfall events) is the main cause of soil erosion in watersheds, its direct effect on water turbidity might not be considered even if a feature selection technique might select it as an input variable. This is because the flow rate in perennial rivers (alike to the case we studied here) is considerably higher than the rate of

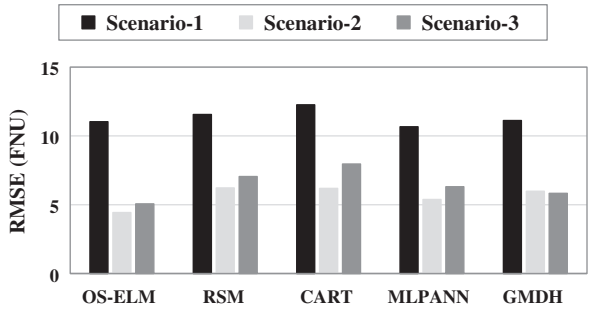

Fig. 6. Comparison of the predicted results in terms of RMSE (FNU) values of the proposed models in the validation set.

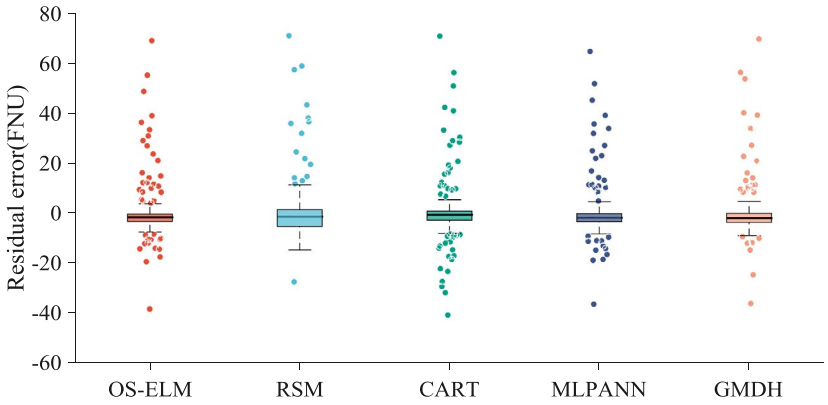

Fig. 7. Boxplots of residual error of applied models for the first scenario (univariate input structure).

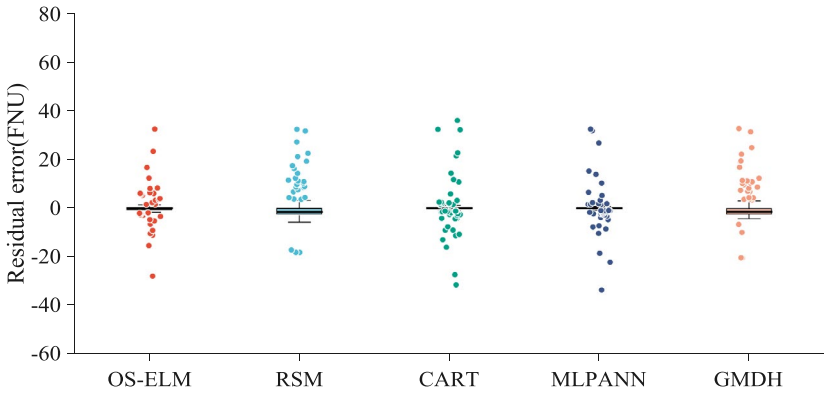

Fig. 8. Boxplots of residual error of applied models for the second scenario (bivariate input structure).

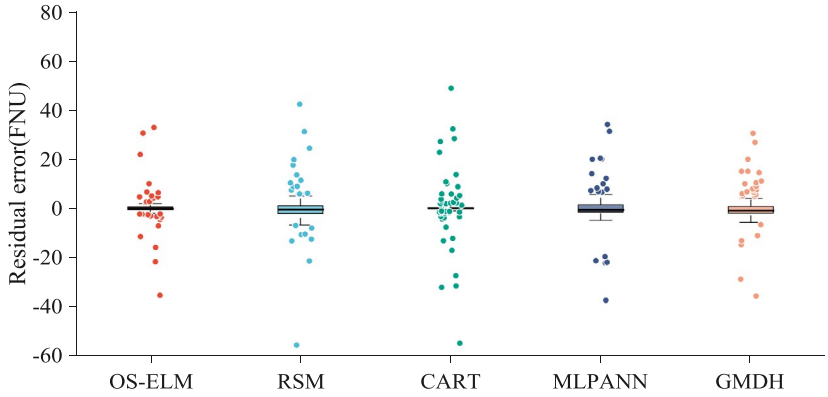

Fig. 9. Boxplots of residual error of applied models for the third scenario (multivariate input structure)

rainfall events, thus, its effect will be overlapped by streamflow velocity and discharge so that one can see the rainfall (in the present case) as a subsidiary vector of streamflow affecting the general flow regime in the river. According to Mulder *et al*. (1998), the main physical factors affecting the values of water turbidity are related to the geological and hydraulic characteristics of the flow (e.g. runout length of the river, bed shear strength and flow). Similar to this study, river flow was reported as the most important parameter on the river turbidity by Beaudeau *et al*. (2001). Iglesis *et al*. (2014) commented that several types of hydrologic, meteorological and water quality parameters have effects on water turbidity. In another study, Stevenson and Cristian (2019) stated that rainfall has a significant influence on the river turbidity in karstic systems.

Regarding the performance of the models, the MLPANN (scenario-2) was superior amongst all the predictive models in the training phase; however, the MLPANN could not hold its first position in the validation phase. Indeed, the OS-ELM model surpassed the MLPANN in predicting the turbidity and became the best predictive model. It should also be noted that all the neural-based models (MLPANN, GMDH and OS-ELM) acted better than the tree-based (CART) and statistical (RSM) models in this study. This conclusion implies that the interconnection architecture of the neural-based models could properly capture the nonlinearity and intercity of the dependent turbidity values.

Conclusions

(1) In this study, we employed an extended version of the extreme learning machine (ELM) method called the online sequential ELM (OS-ELM) for predicting river water quality parameter (turbidity) using different quantity and quality river flow data.

(2) Compared with several well-known and common machine learning and data-driven models, including the perceptron neural network (MLPANN), the classification and regression tree (CART), the group method of data handling (GMDH) and the response surface method (RSM), the proposed OS-ELM approach achieved better predictions.

(3) In contrast to the calculated Mallow's Cp and R-squared values for determining the best input subset for constructing the models, the findings of the study indicated that including five independent variables (i.e. discharge, precipitation, specific conductance, suspended sediment and dissolved oxygen) did not improve the prediction results of the OS-ELM model.

(4) In other words, constructing the bivariate predictive model based on the discharge and suspended sediment measured data (RMSE = 4.431 FNU), acted better than the multivariate OS-ELM model (RMSE = 5.06 FNU).

(5) In addition, all the applied machine learning models acted better in predicting turbidity in comparison to the employed statistical RSM approach.

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Data availability statement

The data that support the findings of this study are publicly available through the USGS website (https://www.usgs.gov).

To submit a comment on this article please go to http://mc.manuscriptcentral.com/wej. For further information please see the Author Guidelines at wileyonlinelibrary.com

References

Adamowski, J. and Sun, K. (2010) Development of a coupled wavelet transform and neural network method

for flow forecasting of non-perennial rivers in semi-arid watersheds. *Journal of Hydrology*, **390**(1–2), 85–91.

Adnan, R.M., Liang, Z., El-Shafie, A., Zounemat-Kermani, M. and Kisi, O. (2019) Prediction of suspended sediment load using data-driven models. *Water*, **11**(10), 2060.

Adnan, R.M., Liang, Z., Heddam, S., Zounemat-Kermani, M., Kisi, O. and Li, B. (2020) Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs. *Journal of Hydrology*, **586**, 124371.

Adnan, R.M., Liang, Z., Trajkovic, S., Zounemat-Kermani, M., Li, B. and Kisi, O. (2019) Daily streamflow prediction using optimally pruned extreme learning machine. *Journal of Hydrology*, **577**, 123981.

Alizamir, M., Kisi, O. and Zounemat-Kermani, M. (2018) Modeling long-term groundwater fluctuations by extreme learning machine using hydro-climatic data. *Hydrological Sciences Journal*, **63**(1), 63–73.

Beaudeau, P., Leboulanger, T., Lacroix, M., Hanneton, S. andWang, H. (2001) Forecasting of turbid floods in a coastal, chalk karstic drain using an artificial neural network. *Groundwater*, **39**(1), 109–118.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984) Classification and regression trees. Belmont, CA: Wadsworth. *International Group*, **432**, 151–166.

Bucher, C.G. and Bourgund, U. (1990) A fast and efficient response surface approach for structural reliability problems. *Structural Safety*, **7**(1), 57–66.

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

Choubin, B., Mosavi, A., Alamdarloo, E.H., Sajedi-Hosseini, F., Shamshirband, S., Dashtekian, K. and Ghamisi, P. (2019a) Earth fissure hazard prediction using machine learning models. *Environmental Research*, **179**, 108770.

Choubin, B., Rahmati, O., Soleimani, F., Alilou, H., Moradi, M. and Alamdari, N. (2019b). Regional groundwater potential analysis using classification and regression trees. Spatial modeling in GIS and R for earth and environmental sciences. In *book: Spatial Modeling in GIS and R for Earth and Environmental Sciences Chapter: 22*, Elsevier, pp. 485–498.

Corzo, G. and Solomatine, D. (2007) Baseflow separation techniques for modular artificial neural network modelling in flow forecasting. *Hydrological Sciences Journal*, **52**(3), 491–507.

Delphla, I., Florea, M. and Rodriguez, M.J. (2018) Drinking water source monitoring using early warning systems based on data mining techniques. *Water Resources Management*, **33**, 129–140.

Díaz Muñiz, C., García Nieto, P.J., Alonso Fernández, J.R., Martínez Torres, J. and Taboada, J. (2012) Detection of outliers in water quality monitoring samples using functional data analysis in San Esteban estuary (Northern Spain). *Science of the Total Environment*, **439**, 54–61.

Fadaee, M., Mahdavi-Meymand, A. and Zounemat-Kermani, M. (2020) Seasonal short-term prediction of dissolved oxygen in rivers via nature-inspired algorithms. *CLEAN–Soil, Air Water*, **48**(2), 1900300.

Gelda, R., Effler, S.W., Prestigiacomo, A.R., Peng, F., Effler, A.J.P., Wagner, B.A., *et al*. (2013) Characterizations and modeling of turbidity in a water supply reservoir following an extreme runoff event. *Inland Waters*, **3**(3), 377–390. Available at: http://delawarewatersheds.org/piedmont/brandywine-creek/

Huang, G.B., Liang, N.Y., Rong, H.J., Saratchandran, P. and Sundararajan, N. (2005) On-line sequential extreme learning machine. *Computational Intelligence*, **2005**, 232–237.

Iglesias, C., Torres, J.M., Nieto, P.J.G., Fernández, J.R.A., Díaz Muñiz, C., Piñeiro, J.I. and Taboada, J. (2014) Turbidity Prediction in a river basin by using artificial neural networks: a case study in Northern Spain. *Water Resources Management*, **28**, 319–331.

Ivakhnenko, A.G. (1968) The group method of data of handling; a rival of the method of stochastic approximation. *Soviet Automatic Control*, **13**, 43–55.

Kisi, O. (2012) Modeling discharge-suspended sediment relationship using least square support vector machine. *Journal of Hydrology*, **456–457**, 110–120.

Liang, N.Y., Huang, G.B., Saratchandran, P. and Sundararajan, N. (2006) A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks*, **17**(6), 1411–1423.

Lima, A.R., Cannon, A.J. and Hsieh, W.W. (2016) Forecasting daily streamflow using online sequential extreme learning machines. *Journal of Hydrology*, **537**, 431–443.

Mather, A. and Johnson, R. (2015) Event-based prediction of stream turbidity using a combined cluster analysis and classification tree approach. *Journal of Hydrology*, **530**, 751–761.

Mather, A. and Johnson, R. (2016) Forecasting turbidity during streamflow events for two Mid-Atlantic U.S. streams. *Water Resources Management*, **30**(13), 4899–4912.

Mulder, T., Syvitski, J.P. and Skene, K.I. (1998) Modeling of erosion and deposition by turbidity currents generated at river mouths. *Journal of Sedimentary Research*, **68**(1), 124–137.

Neupane, S., Vogel, J.R., Storm, D.E., Barfield, B.J. and Mittelstet, A.R. (2015) Development of a turbidity prediction methodology for runoff-erosion models. *Water, Air, and Soil Pollution*, **226**, 415.

Nieta, P.J.G., García-Gonzalo, E., Fernández, J.R.A. and Muñiz, C.D. (2014) Hybrid PSO–SVM-based method for long-term forecasting of turbidity in the Nalón river basin: a case study from Northern Spain. *Ecological Engineering*, **73**, 192–200.

Rajaee, T., Nourani, V., Zounemat-Kermani, M. and Kisi, O. (2011) River suspended sediment load prediction: application of ANN and wavelet conjunction model. *Journal of Hydrologic Engineering*, **16**, 613–627.

Savary, M., Johannet, A., Massei, N., Dupont, J.P., Hauchard, E. (2020) Limits in using multiresolution analysis to forecast turbidity by neural networks. Case study on the Yport Basin, Normandie-France. In: Bertrand, C., Denimal, S., Steinmann, M., and Renard, P. (Eds.) *Eurokarst 2018, Besançon. Advances in Karst Science*. Cham: Springer.

Stevenson, M. and Cristian, B. (2019) Advanced turbidity prediction for operational water supply planning. *Decision Support Systems*, **119**, 72–84.

Sudheer, K.P., Nayak, P.C. and Ramasastri, K.S. (2003) Improving peak flow estimates in artificial neural network river flow models. *Hydrological Processes*, **17**(3), 677–686.

USEPA. (2009). Environmental impact and benefits assessment for final effluent guidelines and standards for the construction and development category. In U. S. E. P. Agency (Ed.) Washington, DC: U.S. Environmental Protection Agency Office of Water (4303T).

Vigil, K.J. (2003) *Clean Water: An Introduction to Water Quality and Water Pollution Control*. Portland: Oregon State University Press.

Yang, T.M., Fan, S.K., Fan, C. and Hsu, N.-S. (2014) Establishment of turbidity forecasting model and early-warning system for source water turbidity management using back-propagation artificial neural network algorithm and probability analysis. *Environmental Monitoring and Assessment*, **186**, 4925.

Zounemat-Kermani, M. (2012) Hourly predictive Levenberg–Marquardt ANN and multilinear regression models for predicting of dew point temperature. *Meteorology and Atmospheric Physics*, **117**(3–4), 181–192.

Zounemat-Kermani, M. (2017) Assessment of several nonlinear methods in forecasting suspended sediment concentration in streams. *Hydrology Research*, **48**, 1240–1252. https://doi.org/10.2166/nh.2016.219.

Zounemat-Kermani, M., Kişi, Ö., Adamowski, J. and Ramezani-Charmahineh, A. (2016) Evaluation of data driven models for river suspended sediment concentration modeling. *Journal of Hydrology*, **535**, 457–472.

Zounemat-Kermani, M., Mahdavi-Meymand, A., Alizamir, M., Adarsh, S. and 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.