

Use of Optimally Pruned Extreme Learning Machine (OP-ELM) in Forecasting Dissolved Oxygen Concentration (DO) Several Hours in Advance: a Case Study from the Klamath River, Oregon, USA

Salim Heddam¹

Received: 18 February 2016 / Accepted: 1 June 2016 / Published online: 7 June 2016
© Springer International Publishing Switzerland 2016

Abstract This study presents a new method called optimally pruned extreme learning machine (OP-ELM) for forecasting dissolved oxygen concentration (DO) several hours in advance. The forecast time horizon ranges from 24-h ahead (one day) to 168-h ahead (seven days). The proposed OP-ELM model is compared to the standard multilayer perceptron neural network (MLPNN) with respect to their capabilities of forecasting DO in the Klamath River at Miller Island Boat Ramp, Oregon, USA. To demonstrate the forecasting capability of OP-ELM and MLPNN models, we used a long-term data set of hourly DO data for a ten-year period, from 1 January 2004 to 31 December 2013, collected by the United States Geological Survey (USGS Stations No: 420,853,121,505,500 [Top] and 420,853,121,505,501 [Bottom]). For developing the models, we split the data set into a training subset (from 2004 to 2010) that corresponded to 70 %, and a validation (from 2011 to 2013) that corresponded to 30 % of the total data set. We investigated the performance and accuracy of the proposed two models for three different horizons, i.e., short-term, medium-term and long-term forecasting; a total of six different models (FM1 to FM6), having the same data sets as inputs, were developed for short-term (24 h to 48 h), medium-term (72 h to 96 h) and long-term (120 h to 168 h) horizons. Input variables used in the six models were the six antecedent DO concentrations at $(t-5)$, $(t-4)$, $(t-3)$, $(t-2)$, $(t-1)$ and (t) . The performance of the OP-ELM and MLPNN models in training and validation sets were compared with the observed data. To get more accurate evaluation of the results of the two models, the following seven statistical performance indices were used: the coefficient of correlation (R), the Willmott index of agreement (d), the Nash-Sutcliffe efficiency (NSE), the root mean squared error (RMSE), the mean absolute error (MAE), the bias error (Bias), and the mean absolute percentage error (MAPE). The study reveals that OP-ELM and MLPNN provided good results and they were successful in forecasting DO at a high level of accuracy. The reliability of forecasting decreased

✉ Salim Heddam
heddamsalim@yahoo.fr

¹ Department of Agronomy, Faculty of Science, Division of Hydraulics University, 20 Août 1955, Route EL HADAÏK, BP 26, SKIKDA, Algeria

with increasing the step ahead. The measures of model performance fell within the acceptable ranges for the two stations. Regarding the fact that researches on medium and long-term forecasting are relatively limited, the present work aims to build and provide a good early warning system capable of preventing DO depletion and the associated problems of anoxia and hypoxia in river. Furthermore, the proposed forecasting models, when implemented appropriately, could be reliably used in detecting future change in DO concentration in rivers.

Keywords Forecasting · Dissolved oxygen · Optimally pruned extreme learning machine · Multilayer perceptron neural network · River

1 Introduction

Dissolved oxygen (DO) concentration is an important water quality parameter, which consists an important indicator of water pollution in rivers (Mohan and Pavan Kumar 2016). DO is used to assess the trophic state of rivers, canals and lakes (Gikas 2014; Chamoglou et al. 2014; Mellios et al. 2015). During the last decade, much effort has been devoted to the modelling of DO in river, lake and stream ecosystems using artificial intelligence (AI) techniques, and much work has been done on this subject. Of course, and not surprisingly, regarding the capabilities and usefulness of the AI techniques, the overwhelming attention given to their application is not just luck. Motivation for using AI techniques in DO modelling has received considerable attention in the last few years and an in-depth literature review suggests that the following four types of AI techniques have been broadly applied on the subject:

1. Artificial neural networks based approach (ANNs): generally, there are the following three types of ANNs that are widely used in modelling DO: (i) generalized regression neural network (GRNN) (Heddam 2014a; Antanasijević et al. 2013, 2014); (ii) radial basis function neural network (RBFNN) (Akkoyunlu et al. 2011; Ay and Kisi 2012; Emamgholizadeh et al. 2014); and (iii) multilayer perceptron neural network (MLPNN) (Ranković et al. 2010; Areerachakul et al. 2013).
2. Fuzzy logic and neurofuzzy system: generally, the famous adaptive neuro-fuzzy inference system (ANFIS) is the broadly reported model in modelling DO (Heddam 2014b; Nemati et al. 2015).
3. Evolutionary models: generally, gene expression programming (GEP) (Kisi et al. 2013), dynamic evolving neural-fuzzy inference system (DENFIS) (Heddam 2014c), Particle Swarm Optimization (PSO) (Liu et al. 2014), and genetic algorithm (GA) (Liu et al. 2013) are the widely used models.
4. Wavelet decomposition models (Evrendilek and Karakaya 2014, 2015).

Although the already developed models for estimating DO have shown good precision and accuracy, and they offer a good alternative in the absence of direct in situ measurements, these models suffer from a number of shortcomings and limitations. First, the models are linked to the water quality variables which already have been used in the calibration of the models; these will typically involve maintaining the input variables available regularly. Second, the models cannot be applied without timely and reliable water quality data. It would be interesting to investigate the capabilities of the AI models in forecasting DO at different level in advance using only the time series of DO and without the need for water quality variables as input to the models. Forecasting DO

in river ecosystems is a relatively difficult task, since the concentration of DO varies significantly over different time periods of the day, especially between morning and evening, within the seasons, and from year to year. Hence, according to what was stated above, we can define three categories of forecasting models: short-term, medium-term and long-term forecasting models. Long-term forecasting refers to more than 5 days in advance (120 h) and refers to one of two horizons, one is 5 days in advance (120 h), the other is 7 days in advance (168 h). In medium-term forecasting, the horizon is usually three to four days in advance (72 h to 96 h). Finally, short-term forecasting involves a period of about one to two days (24 h to 48 h). Obviously, for the long-term forecasting horizon, using a long historical data set, may improve the accuracy of the developed models.

Understanding the concentration and change of DO in freshwater ecosystems is necessary and of great importance. DO varies depending on many factors, including salinity, temperature and pressure (USGS 2008). These factors can be classified as follows (Cox 2003a, 2003b): (i) reaeration from the atmosphere; (ii) enhanced aeration at weirs and other structures; (iii) photosynthetic oxygen production; and (iv) the introduction of DO from other sources such as tributaries. The primary causes of DO depletion are: (i) the oxidation of organic material in the water column; (ii) degassing of oxygen in supersaturated water; (iii) respiration by aquatic plants; (iv) oxygen demand exerted by river bed sediments (Cox 2003a, 2003b). Low levels of DO in the water are accompanied by hypoxia (HY) and anoxia (AX) (Nürnberg 2004). Anoxia is defined as a condition of zero (a total loss of DO) or very low DO concentrations, while hypoxia is defined as a condition of low DO. Seasonal hypoxia, and eventual anoxia, can be caused by high anthropogenic nutrient loadings (Breitburg et al. 2003). Hypoxia profoundly affects ecosystem functioning, with strong implications for all ecosystem components, i.e., biodiversity, biological and cause physiological stress, and even death, to associated aquatic organisms, which indicate a stressed environment (Friedrich et al. 2014). Regarding the importance of the HY and AX processes, it will be very interesting to develop a model that will be used as an early warning system, and will have the ability to deliver more accurate and time sensitive information during stress events.

Few studies have attempted to develop forecasting models for DO time series without the need of using the water quality variable as input to the models. Alizadeh and Kavianpour (2015) compared two ANNs models, the standard MLPNN and wavelet-neural network (WNN), in forecasting DO at hourly and daily time steps, at the Hilo Bay on the east side of the Big Island of Hawaii, Pacific Ocean. DO at different time lags were considered as input variables for the MLPNN and WNN models. They have tested four forecasting models having different input combinations: $DO(t-1)$, $DO(t-2)$, $DO(t-3)$ and $DO(t-4)$. The models developed had only one output: the DO at $(t + 1)$. The t corresponds to the DO at the present time (day or hour), and the $(t-1)$ until $(t-4)$ correspond to the four previous values of DO. As a result of the study, for the models at daily time step, they have obtained a correlation coefficient (R) of 0.66 and 0.72 for the models having two inputs (i.e., $(t-1)$ and $(t-2)$) and the models having four inputs (i.e., $(t-1)$, $(t-2)$, $(t-3)$ and $(t-4)$), respectively. At hourly time step, the authors reported excellent results with an R of 0.94 in the validation phase. An et al. (2015) proposed a new forecasting model using the nonlinear grey Bernoulli model (NGBM (1, 1)) in forecasting DO in the Guanting reservoir (inlet and outlet), located at the upper reaches of the Yongding River in northwest Beijing, China. The authors used DO at weekly time step. Altunkaynak et al. (2005) examined the potential of fuzzy logic-based model (FL) to forecast monthly DO in the Golden Horn in Istanbul, Turkey. The authors have used two inputs (i.e., $DO(t-1)$ and $DO(t-2)$) and one output ($DO(t)$). As a result of the study, it was seen that one can forecast the next month's DO concentration from antecedent measurements within acceptable relative error

limits. Wang et al. (2013) compared the bootstrapped wavelet neural network (BWNN) as a new tool in forecasting monthly DO in Harbin, China. The proposed BWN model has been compared with the standard MLPNN and the wavelet neural network models (WNN). They have compared three types of models: using only $DO(t-1)$ as input; using only $DO(t-3)$ as input and using $DO(t-1)$ and $DO(t-3)$ as inputs. In all three developed models the output was the $DO(t)$. As a result of the study the authors reported a high Nash-Sutcliffe efficiency (NSE) value equal to 0.85 in the validation phase. Faruk (2010) applied the MLPNN in forecasting monthly DO time series using the data for the period between 1996 and 2004 from the Büyük Menderes basin, Turkey, and has obtained a NSE coefficient equal to 0.91.

As indicated above, a literature review demonstrates that modelling DO is broadly discussed and several kinds of models have been developed worldwide. However, few studies have been devoted to forecasting DO and many more are needed. In the previous forecasting studies, we can report the following drawbacks: (i) the time-step chosen is often constrained by the availability of DO data, and the majority of the models use data at monthly, weekly or infrequently daily time step, which is not really efficient as the amount of DO varies rapidly; (ii) in all reported models, the forecast horizon is generally the value at the next time ($t + 1$), and none of the studies has investigated a long term forecasting horizon. Therefore, in this paper we present the OP-ELM as a new and powerful tool in forecasting hourly dissolved oxygen concentration (DO) several hours in advance. The proposed model is compared to the standard MLPNN in order to demonstrate its superiority and usefulness. We test this new model with data from the Klamath River at Miller Island Boat Ramp, Oregon, USA, to illustrate how the proposed model can forecast DO very well. To the author's knowledge, DO forecasting with OP-ELM is the first study in the literature. Therefore, the present study investigates the use of OP-ELM in the development of a robust model for DO forecasting.

2 Materials and Methods

2.1 Study Area and Data Set

Historical hourly DO data from 1 January 2004 to 31 December 2013 were used in this study; they are available at the United States Geological Survey (USGS) website: http://or.water.usgs.gov/cgi-bin/grapher/table_setup.pl?site_id. Two stations were chosen: at Klamath, River Miller Island Boat Ramp, Oregon USA [Top] (USGS ID: 420,853,121,505,500; Latitude $42^{\circ}08'53''N$; Longitude $121^{\circ}50'55''E$; NAD83); and at Klamath, River Miller Island Boat Ramp, Oregon USA [Bottom] (USGS ID:420,853,121,505,501). Figure 1 shows the locations of the stations. For the two stations, the data set was divided into two sub-data sets: (i) training set (70 %); and (ii) validation set (30 %). Hence, the first seven years (2004 to 2010) were chosen for training and the last three years (2011–2013) for validation. Data were stored in a table, in which each row corresponded to one hour (one pattern) and contained a measured DO value. In some hours, the DO variable was not available, in which case we removed the incomplete patterns. A detailed description of the original and final patterns is reported in Table 1. From Table 1, for the top station a total of 1047 patterns were missing and for the bottom station a total of 3763 patterns were missing. Figures 2 and 3 show the plot of the DO time series for the training and validation, for the two stations. Table 2 summarizes the descriptive statistics of the data set, where X_{mean} , X_{max} , X_{min} , S_x and C_v denote the mean,

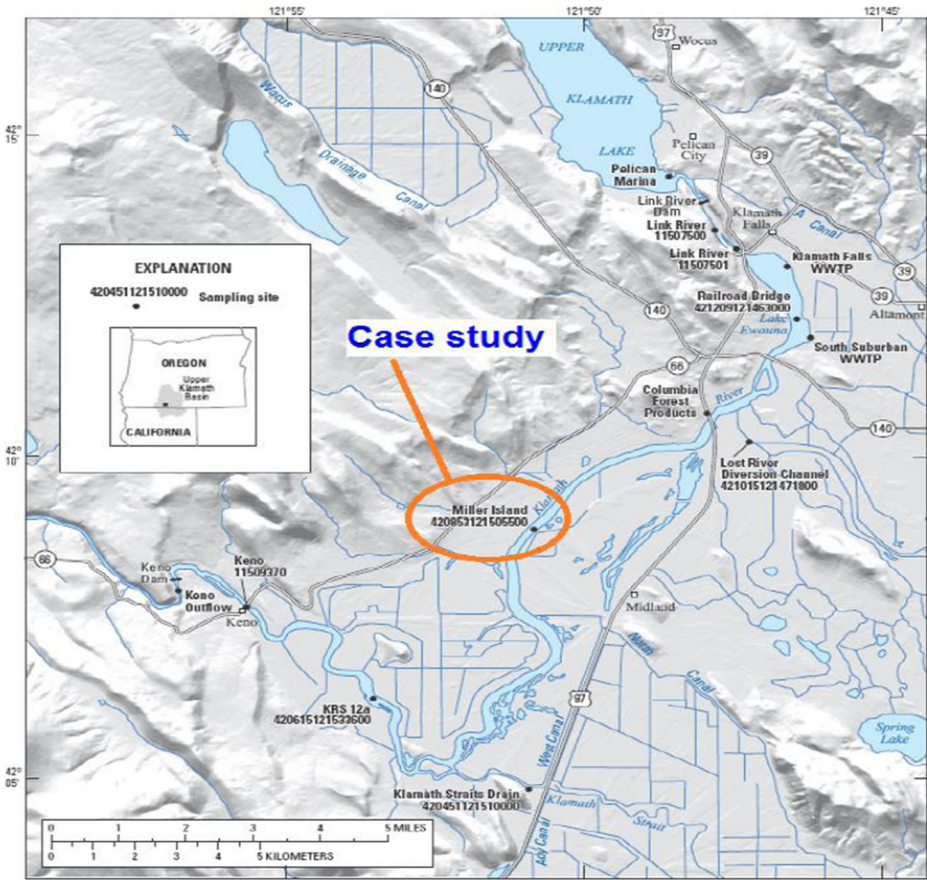


Fig. 1 Map showing the study area, Klamath River at Miller Island Boat Ramp [Top and Bottom] stations, Oregon, USA. (Adopted from Sullivan et al. 2012, 2013a, 2013b)

the maximum, the minimum, the standard deviation, and the coefficient of variation, respectively. All the data derived were normalized to have zero mean and unit variance using the Z-score method, calculated by the following formula:

$$x_{ni,k} = \frac{x_{i,k} - m_k}{Sd_k} \tag{1}$$

where $x_{ni,k}$ is the normalized value of the variable k (input or output) for each sample i . $x_{i,k}$ is the original value of the variable k (input or output). m_k and Sd_k are the mean value and standard deviation of the variable k (input or output). Normalization is an important process, which increases significantly the performance of the models (Kingston et al. 2005; Heddard et al. 2011, 2012, 2016; Heddard 2016a, 2016b).

2.2 Multilayer Perceptron Neural Network (MLPNN)

Artificial neural networks (ANNs) are an information processing system which belongs to the category of nonlinear models (Haykin 1999). ANNs are inspired from the function of the

Table 1 Dataset presentation includes years of available data, the description of the starting and ending dates covering the period of the study, the total patterns available, the incomplete patterns removed from the table, the final patterns used, and separating the data to the subset for training and validation

Station	Year	Begin Date	End Date	Total Patterns	Incomplete Patterns	Final Patterns	Subset
Klamath River at Miller Island Boat Ramp [Top] (420,853,121,505,500)	2004	01/01/2004	12/31/2004	8784*	0382	8402	Training
	2005	01/01/2005	12/31/2005	8760**	0119	8641	Training
	2006	01/01/2006	12/31/2006	8760	0238	8522	Training
	2007	01/01/2007	12/31/2007	8760	0022	8738	Training
	2008	01/01/2008	12/31/2008	8784	0097	8687	Training
	2009	01/01/2009	12/31/2009	8760	0023	8737	Training
	2010	01/01/2010	12/31/2010	8760	0002	8758	Training
	2011	01/01/2011	12/31/2011	8760	0004	8756	Validation
	2012	01/01/2012	12/31/2012	8784	0154	8630	Validation
	2013	01/01/2013	12/31/2013	8760	0006	8754	Validation
Klamath River at Miller Island Boat Ramp [Bottom] (420,853,121,505,501)	2004	01/01/2004	12/31/2004	8784	0403	8381	Training
	2005	01/01/2005	12/31/2005	8760	1381	7379	Training
	2006	01/01/2006	12/31/2006	8760	0191	8569	Training
	2007	01/01/2007	12/31/2007	8760	0603	8157	Training
	2008	01/01/2008	12/31/2008	8784	0413	8371	Training
	2009	01/01/2009	12/31/2009	8760	0180	8580	Training
	2010	01/01/2010	12/31/2010	8760	0047	8713	Training
	2011	01/01/2011	12/31/2011	8760	0151	8609	Validation
	2012	01/01/2012	12/31/2012	8784	0015	8769	Validation
	2013	01/01/2013	12/31/2013	8760	0379	8381	Validation

*8784 patterns correspond to: [7 months × 31 days × 24 h] = 5208 + [4 months × 30 days × 24 h] = 8088 + [1 month × 29 days × 24 h] = 8784. **only February with 28 days, while the others all had 30 or 31 days

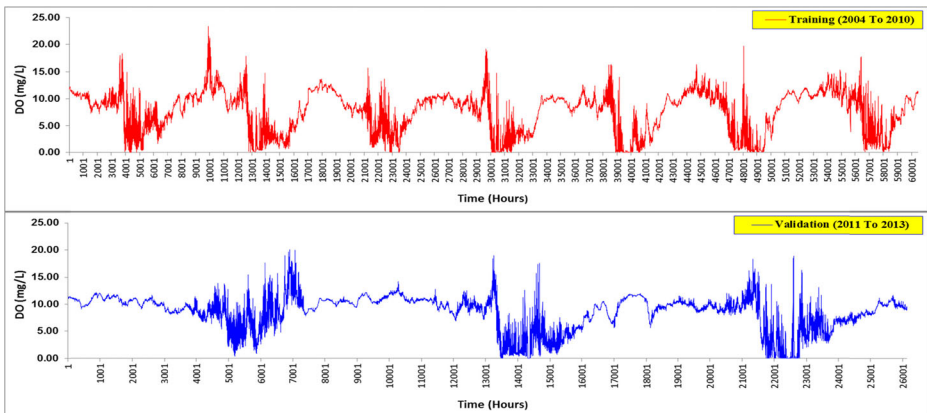


Fig. 2 Time series plot of dissolved oxygen concentration (DO) for the USGS [Top] (420,853,121,505,500) station, covering the period of the study. Here the beginning and ending dates are in the training phase from to 01/01/2004 at 01:00 until 12/31/2010 at 24:00; and in the validation phase from to 01/01/2011 at 01:00 until 12/31/2013 at 24:00

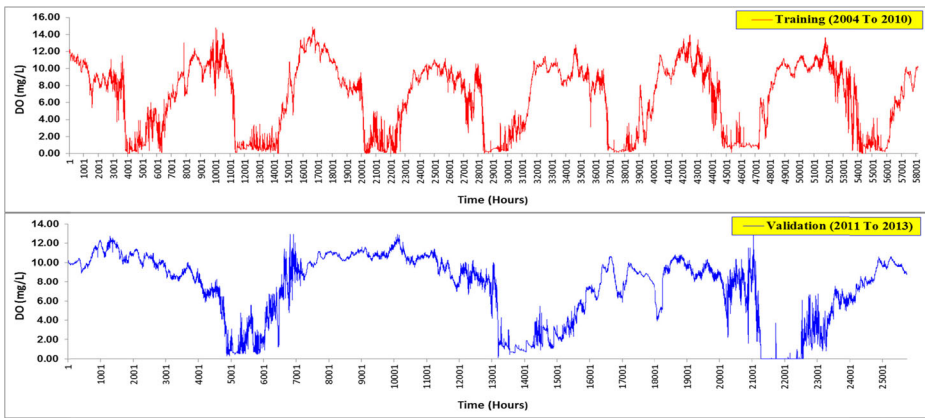


Fig. 3 Time series plot of dissolved oxygen concentration (DO) for the USGS 420853121505501 [Bottom] station, covering the period of the study. Here the beginning and ending dates are in the training phase from to 01/01/2004 at 01:00 until 12/31/2010 at 24:00; and in the validation phase from to 01/01/2011 at 01:00 until 12/31/2013 at 24:00

human brain. In ANNs, the neurons are tightly interconnected and organized into different layers. There are three types of layers, i.e., an input layer, one or more hidden layers, and an output layer. The layers are connected by weights (Haykin 1999). Multilayer perceptron neural network (MLPNN) (Rumelhart et al. 1986) is the most widely used type of ANNs reported in the literature, considered as a powerful tool in modeling nonlinear processes. The number of neurons contained in the input and output layers, is related to the problem that must be resolved. Each neuron in the input layer corresponds to a single variable, and the output neurons represent the solution of the problem being solved. Figure 4 presents the structure of the MLPNN models used in the present study. It consist of three layers: the input layer with six variables, denoted as x_1 to x_6 and corresponding to the DO at $(t-5)$, $(t-4)$, $(t-3)$, $(t-2)$, $(t-1)$ and (t) ; the hidden layer with a number of neurons defined by trial and error; and finally, the output layer with only one neuron: the DO at $(t + 24)$ to $(t + 168)$, according to the model developed. The MLPNN are capable of approximating any function with a finite number of discontinuities (Hornik et al. 1989) and are considered as a universal approximator (Hornik et al. 1989; Hornik 1991). Let us denote with k the number of input variables, and m the number of neurons in the hidden layer; the

Table 2 Statistical parameters of data set for forecasting models

Station	Period	Unit	X_{mean}	X_{max}	X_{min}	S_x	C_v	N
Klamath River at Miller Island Boat Ramp [Top] (420,853,121,505,500)	Training	mg/L	7.527	23.40	0.000	3.802	0.505	60,485
	Validation	mg/L	8.527	20.00	0.000	3.130	0.367	26,140
	Whole period	mg/L	7.829	23.40	0.000	3.641	0.465	86,625
Klamath River at Miller Island Boat Ramp [Bottom] (420,853,121,505,501)	Training	mg/L	6.581	14.90	0.000	4.097	0.623	58,150
	Validation	mg/L	7.332	13.30	0.000	3.514	0.479	25,759
	Whole period	mg/L	6.812	14.90	0.000	3.943	0.579	83,909

Abbreviations: X_{mean} , mean; X_{max} , maximum; X_{min} , minimum; S_x , standard deviation; C_v , coefficient of variation; N, number of patterns

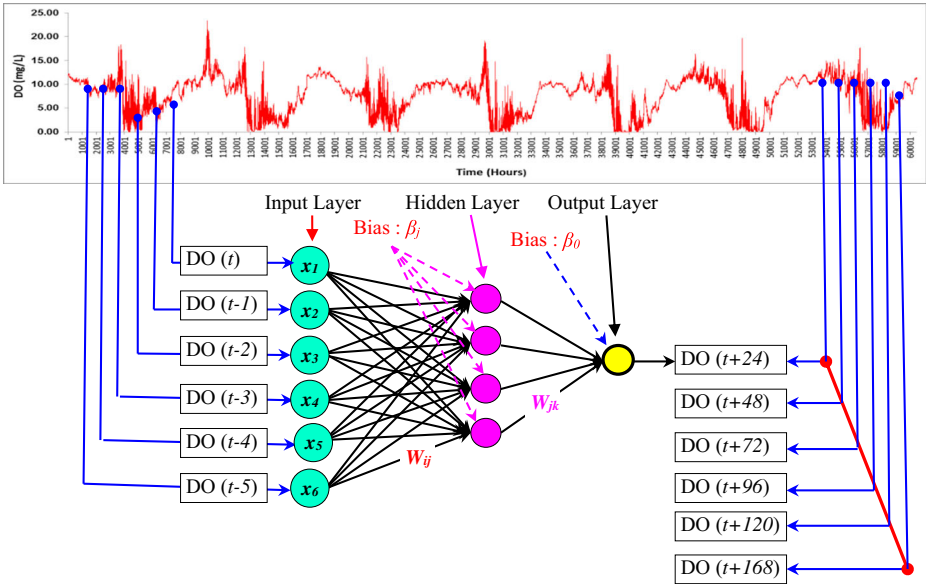


Fig. 4 Multilayer Perceptron Neural Network (MLPNN) structure used in forecasting dissolved oxygen concentration. The structure in this figure is valid for the two stations

mathematical structure of the MLPNN from the input to the output can be formulated as follows:

$$A_j = \beta_j + \sum_{i=1}^k w_{ij} \times x_i \tag{2}$$

where A_j is the weighted sum of the j hidden neuron, k is the total number of inputs, w_{ij} denotes the weight characterising the connection between the n^{th} input to the m^{th} hidden neuron, and β_j is the bias term of each hidden neuron. The output of the m^{th} hidden neuron is given by

$$Y_j = f(A_j) \tag{3}$$

The activation function f adopted in the present study was the sigmoid, given by Eq. (4):

$$f(A) = \frac{1}{1 + e^{-A}} \tag{4}$$

The neural network output is then given by:

$$O_k = \beta_0 + \sum_{j=1}^m w_{jk} \times Y_j \tag{5}$$

where w_{jk} denotes the weight characterising the connection between the m^{th} hidden neuron to the p^{th} output neuron, m the total number of hidden neurons, and β_0 is the bias term. A linear activation function is most commonly applied to the output layer.

During the last few years, AI techniques are largely used by scientists in developing different kind of models, and the ANNs are the most applied form of AI. Typical applications

include the following, among many others: predicting the dispersion coefficient (D) in a river ecosystem (Antonopoulos et al. 2015); modelling the permeability losses in permeable reactive barriers (Santisukkasaem et al. 2015); estimating the reference evapotranspiration (ET₀) in India (Adamala et al. 2015); calculating the dynamic coefficient in porous media (Das et al. 2015); predicting Indian monsoon rainfall (Azad et al. 2015); modeling of arsenic (III) removal (Mandal et al. 2015); predicting effluent biochemical oxygen demand (BOD) in a wastewater treatment plant (Heddham et al. 2016); modeling Secchi disk depth (SD) in river (Heddham 2016a); and predicting phycocyanin (PC) pigment concentration in river (Heddham 2016b).

Unsurprisingly, regarding the high capabilities of ANNs in developing environmental models, they have rapidly gained much popularity. But even with this, ANNs models are not without criticism. The performance of the ANNs models depends mainly on (Maier and Dandy 2000; Maier et al. 2010): (i) the selection of the appropriate input; (ii) data division (the best selection of the splitting ratio between training and validation); (iii) the model structure (the number of hidden neurons, number of hidden layers and the type of transfer function); (iv) model calibration; and finally (v) model evaluation using different kind of statistical metrics. Some important details about the appropriate application of the ANN models can be found in: Bowden et al. (2005a, 2005b); Maier and Dandy (2000); and Maier et al. (2010).

2.3 Optimally Pruned Extreme Learning Machine (OP-ELM)

Extreme learning machine (ELM) has been proposed by Huang et al. (2006a, 2006b) for training single hidden layer feedforward neural networks (SLFN) (Huang et al. 2015). ELM is a universal function approximator (Huang and Chen 2007, 2008). In SLFN, we have two weight matrices, from the input layer to the hidden layer (input layer weights) and from the hidden layer to the output layer (output layer weights). These two matrices must be optimized during the training process, generally using the backpropagation training algorithm. According to Huang et al. (2006a, 2006b), when using the ELM algorithm, the input layer weights do not need to be tuned iteratively and to be generated randomly; however, the output layer weights are determined analytically using least-squares method (Huang et al. 2015). The ELM algorithm is presented briefly below.

Let be N arbitrary training samples represented by (x_i, y_i), where x_i = [x_{i1}, x_{i2}, ..., x_{iD}]^T ∈ R^D and [y_{i1}, y_{i2}, ..., y_{iD}]^T ∈ R^D. Single hidden layer feedforward neural networks with L hidden nodes and activation function f(x) are mathematically presented as (Huang et al. 2006a, 2006b):

$$\sum_{i=1}^L \beta_i f_i(x_j) = \sum_{i=1}^L \beta_i f_i(w_i \cdot x_j + b_i) = o_j \tag{6}$$

where j = 1, 2, ..., N. Here w_i = [w_{i1}, w_{i2}, ..., w_{iD}]^T is the weight vector connecting the ith hidden node and the input nodes, β_i = [β_{i1}, β_{i2}, ..., β_{ik}]^T is the weight vector connecting the ith hidden node and the output nodes, and b_i is the threshold of the ith hidden node (Zeng et al. 2013). The standard SLFN with L hidden nodes with activation function f(x) can be written as:

$$H\beta = T \tag{7}$$

where:

$$H = \begin{bmatrix} f_1(w_1 \cdot x_1 + b_1) & \cdots & f_L(w_L \cdot x_1 + b_L) \\ \vdots & & \vdots \\ f_1(w_1 \cdot x_N + b_1) & \cdots & f_L(w_L \cdot x_N + b_L) \end{bmatrix} \tag{8}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix} \text{ and } T = \begin{bmatrix} t_1^T \\ \vdots \\ t_L^T \end{bmatrix} \tag{9}$$

Different from the conventional gradient-based solution of SLFN, ELM simply solves the function by:

$$\beta = H^+ T \tag{10}$$

where H^+ is the Moore-Penrose generalized inverse of matrix H (Zeng et al. 2013). More details about ELM can be found in: Huang et al. (2006a, 2006b); Huang et al. (2011, 2015); and Huang (2015).

Since it was proposed, ELM has received much popularity and several modifications were proposed to the original algorithm. Among them, we could name: the optimally pruned extreme learning machine (OP-ELM) (Miche et al. 2008, 2010); online sequential extreme learning machine (OS-ELM) (Liang et al. 2006); Bidirectional extreme learning machine B-ELM (Yang et al. 2012); Evolutionary extreme learning machine SaDE-ELM (Cao et al. 2012); Fully Complex Extreme Learning Machine C-ELM (Li et al. 2005); and others. In the present study we focused on the famous OP-ELM.

The optimally pruned extreme learning machine was developed by Miche et al. (2008, 2010). It is simply called OP-ELM (Miche et al. 2008, 2010; Sorjamaa et al. 2008), and is a modified version of the original extreme learning machines (ELM). According to Pouzols and Lendasse (2010a, 2010b), OP-ELM is designed for use for up to three different types of kernel functions, i.e., Gaussian, sigmoid and linear kernels, and the creation of an OP-ELM model occurs in three stages. Considering the fact that the original ELM suffer from an important drawback, which is the presence of irrelevant or correlated variables in the training data set, Miche et al. (2008, 2010) introduced the OP-ELM model as a good solution for pruning away irrelevant variables, thereby, further improving the reliability and efficiency of the ELM approach.

Given a training set (x_i, y_i) , $x_i \in \mathbb{R}^{d_1}$ and $y_i \in \mathbb{R}^{d_2}$; OP-ELM implies three important steps (Grigorievskiy et al. 2014): (i) build a regular ELM model with initially large number of neurons; (ii) rank neurons using multiresponse sparse regression (MRSR) (Similä and Tikka 2005) or least angle regression (LARS) (Efron et al. 2004) if the output is one-dimensional; and (iii) use leave-one-out (LOO) validation to decide how many neurons to prune (Fig. 5). In the first step, a construction of a SLFN using the original ELM algorithm with a large number of neurons is made (by default one hundred neurons). It is important to note that while the ELM use only sigmoid kernels, the OP-ELM offers three possibilities of kernel types: sigmoid,

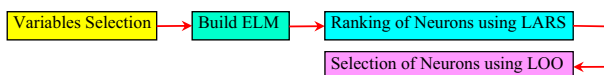


Fig. 5 The important steps in developing an OP-ELM (From Sorjamaa et al. 2008)

Gaussian and linear (Pouzols and Lendasse 2010a, 2010b). Ranking the hidden neurons according to their accuracy is done in the second step using the MRSR algorithm. Hereafter, we present the MRSR algorithm briefly, according to Similä and Tikka (2005). Suppose that the targets are denoted by an $n \times p$ matrix $T = [t_1 \dots t_p]$ and the regressors are denoted by an $n \times m$ matrix $X = [x_1 \dots x_m]$. The MRSR algorithm adds sequentially active regressors to the model:

$$\gamma^k = XW^k \tag{11}$$

such that the $n \times p$ matrix $Y^k = [y_1^k \dots y_p^k]$ models the targets T appropriately. The $m \times p$ weight matrix W^k includes k nonzero rows at the beginning of the k th step. Each step introduces a new nonzero row and, thus, a new regressor to the model. It is reported that MRSR is a variable ranking rather than selection algorithm (Pouzols and Lendasse 2010a). Finally, the third and last step in OP-ELM is the application of the leave-one-out (LOO) validation method where the decision over the actual best number of neurons for the model is taken. According to Grigorievskiy et al. (2014), using the so-called application of the PRESS (PREdiction Sum of Squares) statistic, provides an exact LOO.

Since it was proposed, OP-ELM has gained much popularity in engineering applications during the past few years. Moreno et al. (2014) compared the ELM and the OP-ELM for the classification of remote sensing hyperspectral data; Grigorievskiy et al. (2014) compared the OP-ELM with the linear model and the Least-Squares Support Vector Machines (LS-SVM) for long-term time series prediction; Akusok et al. (2015) applied the OP-ELM for the identification of the likeliest samples in a dataset to be mislabelled; and Sovilj et al. (2010) compared the OP-ELM and optimally pruned k-nearest neighbors OP-KNN for long-term time series prediction.

2.4 Model Performance Indices

In the present study, we used seven performance indices in order to evaluate performance and compare the developed models. These seven indices are: the coefficient of correlation (R) (Eq. 12) (Legates and McCabe 1999), the Willmott index of agreement (d) (Eq. 13) (Willmott 1982; Willmott et al. 1985), the Nash-Sutcliffe efficiency (NSE) (Eq. 14) (Nash and Sutcliffe 1970; Dawson and Wilby 2001), the root mean squared error (RMSE) (Eq. 15) (Chai and Draxler 2014; Dawson et al. 2002), the mean absolute error (MAE) (Eq. 16) (Dawson et al. 2002; Gebremariam et al. 2014), the bias error (Bias) (Eq. 17) (Dawson et al. 2007), and the mean absolute percentage error (MAPE) (Dawson et al. 2007) (Eq. 18):

$$NSE = 1 - \frac{\sum_{i=1}^N [O_i - P_i]^2}{\sum_{i=1}^N [O_i - O_m]^2} \tag{12}$$

$$d = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (|P_i - O_m| + |O_i - O_m|)^2} \tag{13}$$

$$R = \left[\frac{\frac{1}{N} \sum (O_i - O_m)(P_i - P_m)}{\sqrt{\frac{1}{N} \sum_{i=1}^n (O_i - O_m)^2} \sqrt{\frac{1}{N} \sum_{i=1}^n (P_i - P_m)^2}} \right] \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (O_i - P_i)^2} \quad (15)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |O_i - P_i| \quad (16)$$

$$\text{Bias} = \frac{1}{N} \sum_{i=1}^N (O_i - P_i) \quad (17)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{O_i - P_i}{O_i} \right| \quad (18)$$

where N is the number of data points, O_i is the measured value and P_i is the corresponding model prediction, and O_m and P_m are the average values of O_i and P_i . The Willmott index of agreement (d) is an indicator of model performance. It carries a value from 0 to 1, where a value of 1 indicates perfect agreement (Willmott 1982). The bias error (Bias), also reported as the mean error (ME), for a perfect match has the value zero. A bias index expresses the tendency of the model to underestimate (positive value) or overestimate (negative value). According to Dawson et al. (2007), a low value for the Bias does not necessarily indicate a good model in terms of accurate forecasts, since positive and negative errors will tend to cancel out each other; for this reason, MAE is often preferred to Bias. The Nash-Sutcliffe efficiency (NSE) is a measure of statistical association, which indicates the percentage of the observed variance that is explained by the predicted data (Nash and Sutcliffe 1970). NSE ranges between $-\infty$ and 1, with $\text{NSE} = 1$ being the optimal value (Moriassi et al. 2007). The R statistic describes the degree of colinearity between the observed and predicted values. A perfect model has an R equal to 1; typically, values greater than 0.70 are considered acceptable (Moriassi et al. 2007). Both RMSE and MAE describe the difference between model simulations and observations in the units of the variable. For both, values close to zero indicate perfect fit. Some advantages of MAE over RMSE in the evaluation and comparison of model performance can be found in Willmott and Matsuura (2005). The MAPE, also called MARE (Dawson et al. 2007) or average absolute percentage error (AAPE) (Maier and Dandy 1996; Bowden et al. 2002), is one of the most widely used measures of forecast accuracy for developing models (Kim and Kim 2016; Deo et al. 2016). The MAPE has no upper bound and for a perfect match the result would be zero (Dawson et al. 2007). Although some researchers worldwide already recommend the use of the MAPE as a measure of forecast accuracy, the MAPE has some advantages and disadvantages. One of the most important disadvantages of the MAPE is that it

produces infinite or undefined values for zero or close-to-zero measured (observed) values (Kim and Kim 2016), which is the case in many environmental case studies. According to Dawson et al. (2007), the MAPE is subject to “*potential fouling by small numbers in the observed record*”. The MAPE is susceptible to outliers and it tends to overstate forecast error (Tayman and Swanson 1999; Rayer 2007). If the observed values are very small, MAPE yields extremely large percentage errors (outliers), while zero actual values result in infinite MAPE (Kim and Kim 2016). Hence, when the observed values are close to zero, the MAPE value can explode to a huge number, and if averaged with the other values, it can give a distorted image of the magnitude of the error (Chase 2013).

In addition to all the reported seven performance indices, we also calculated the so-called RPD (Eq. 19) (ratio of standard deviation to RMSE), the scatter index (SI) (Eq. 20), also reported as the normalized objective function (NOF) (Gikas 2014), and the ratio of the mean absolute error to the standard deviation (RMAE) (Eq. 21). The RPD is an important index of model accuracy; in general, RPD values less than one are unacceptable while values greater than 3 are considered excellent (Chang et al. 2001; Ackerson et al. 2015). A value of the RMAE less than 0.5 may be considered low, acceptable and appropriate for model evaluation (Singh et al. 2004; Moriasi et al. 2007). For the scatter index (SI), a perfect model has a SI equal to 0.0 (Gikas 2014), and for values between 0.0 and 1.0, the model predictions are acceptable (Boskidis et al. 2010, 2011; Gikas et al. 2006; Pisinaras et al. 2010).

$$RPD = \frac{S_x}{RMSE} \tag{19}$$

$$SI = \frac{RMSE}{O_m} \tag{20}$$

$$RMAE = \frac{MAE}{S_x} \tag{21}$$

Finally, we report a comparison between measured and calculated mean and standard deviation of DO, in the training and validation phases.

3 Results and Discussion

An attempt has been made in this study to develop a robust model for forecasting DO several hours in advance, in the Klamath River, USA, using OP-ELM model. To draw robust conclusions, it is interesting to compare the results using OP-ELM with the results obtained using the standard MLPNN. We have developed six forecasting models called FM1 to FM6 using the same input data and having different outputs. The structures are shown in Table 3. The six models correspond to three different horizons: short-term (FM1 and FM2), medium-term (FM3 and FM4) and long-term forecasting horizons (FM5 and FM6). Forecasting DO time series refers to estimating the future values at different hour intervals by using a set of previous observed values. This will be interesting, as we use only the time series of DO without any other input variables. For the OP-ELM, we used the implementation available from www.cis.hut.fi/projects/tsp/index.php. In the present study, we applied the OP-ELM with all possible kernels, i.e., linear, sigmoid, and Gaussian, using a maximum number of 100

Table 3 Combinations of input variables considered in developing forecasting models

Forecasting horizons	Model	Input structure	Output
Short-Term	FM1	DO (t-5), DO (t-4), DO (t-3), DO (t-2) , DO (t-1) , DO (t)	DO (t + 24): +24 h ahead
	FM2	DO (t-5), DO (t-4), DO (t-3), DO (t-2) , DO (t-1) , DO (t)	DO (t + 48): +48 h ahead
Medium-Term	FM3	DO (t-5), DO (t-4), DO (t-3), DO (t-2) , DO (t-1) , DO (t)	DO (t + 72): +72 h ahead
	FM4	DO (t-5), DO (t-4), DO (t-3), DO (t-2) , DO (t-1) , DO (t)	DO (t + 96): +96 h ahead
Long-Term	FM5	DO (t-5), DO (t-4), DO (t-3), DO (t-2) , DO (t-1) , DO (t)	DO (t + 120):+120 h ahead
	FM6	DO (t-5), DO (t-4), DO (t-3), DO (t-2) , DO (t-1) , DO (t)	DO (t + 168):+168 h ahead

neurons (Fig. 6). For all the six models developed, the relationship among the inputs and the output can be defined by Eq. (22):

$$DO(t + n) = f[DO(t-5), DO(t-4), DO(t-3), DO(t-2), DO(t-1), DO(t)] \quad (22)$$

For example, for model FM1, the output corresponds to DO at $(t + 24)$, and the forecasting process is the following: with the N observations of the time series of DO at $(t = 1, t = 2, t = 3, t = 4, \dots, t = N)$, using a model with n input nodes (6 inputs in our study), we have a training data composed of $N-n$ training patterns, in which the first patterns will be composed of DO at $(t = 1, 2, 3, \dots, n)$ as inputs and DO $(n + 24)$ as the output; the second pattern will contain DO at $(t = 2, 3, 4, \dots, n + 1)$ as inputs and DO $(n + 25)$ as the output, and the last pattern will be DO $(N-(n-6 + 24), N-(n-5 + 24), N-(n-4 + 24), \dots, N-(n-1 + 24))$ as inputs and DO (N) as the output, where N is the total

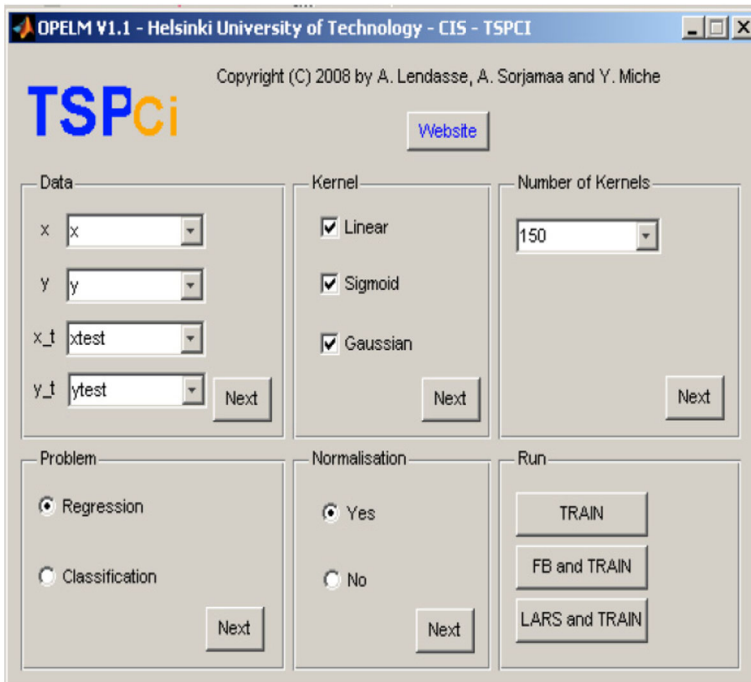


Fig. 6 Block diagram of the OP-ELM Toolbox (From Miche et al. 2008)

number of patterns. It is clear from the proposed structure that for the FM6 model that has DO at $(t + 168)$ as output, we have a training data composed of $(N-168)$ training patterns, in which the first patterns will be composed of DO at $(t = 1, 2, 3, 4, 5, 6)$ as inputs and DO $(t = 172)$ as the output. This structure is valuable for the all six developed models. Therefore, all the six developed models are basically approximators of the general equation (19). In the following sections, the forecasting capabilities of the two proposed models (OP-ELM and MLPNN) are presented and compared.

3.1 Forecasting DO Concentration at the USGS 420853121505500 [Top] Station

According to the seven indices for assessing the performance of the models, the optimal model should have the lowest RMSE, MAE, MAPE and Bias, and the values of NSE, R and d should be close to 1. Table 4 displays the RMSE, MAE, MAPE, Bias, NSE, R and d model performance statistics for both the training and validation sets. The two models (OP-ELM

Table 4 Performance indices for the OP-ELM and MLPNN forecasting models in the training and validation phases several hours in advance for USGS 420853121505500 [Top] station

Station: Klamath River at Miller Island Boat Ramp [Top] (420,853,121,505,500)

	Model	Forecasting Interval	RMSE (mg/L)	MAE (mg/L)	Bias (mg/L)	MAPE (%)	NSE /	R /	d /
Training Phase									
Short-Term	OP-ELM	+24 h	1.081	0.648	0.0000	0.337	0.919	0.959	0.979
		+48 h	1.292	0.839	0.0000	0.442	0.884	0.940	0.969
	MLPNN	+24 h	1.054	0.635	-0.0001	0.315	0.923	0.961	0.980
		+48 h	1.277	0.832	0.0005	0.429	0.887	0.942	0.969
Medium-Term	OP-ELM	+72 h	1.427	0.962	0.0000	0.527	0.859	0.927	0.961
		+96 h	1.538	1.052	0.0000	0.596	0.836	0.914	0.954
	MLPNN	+72 h	1.406	0.950	0.0007	0.515	0.863	0.929	0.962
		+96 h	1.512	1.037	0.0059	0.599	0.842	0.918	0.955
Long-Term	OP-ELM	+120 h	1.641	1.136	0.0000	0.675	0.814	0.902	0.946
		+168 h	1.785	1.268	0.0000	0.784	0.779	0.883	0.935
	MLPNN	+120 h	1.613	1.122	-0.0021	0.644	0.820	0.906	0.949
		+168 h	1.741	1.241	0.0033	0.750	0.790	0.889	0.938
Validation Phase									
Short-Term	OP-ELM	+24 h	1.225	0.665	0.0430	0.176	0.848	0.922	0.959
		+48 h	1.446	0.854	0.0344	0.237	0.788	0.888	0.940
	MLPNN	+24 h	1.270	0.697	0.1055	0.210	0.836	0.919	0.958
		+48 h	1.527	0.888	0.1336	0.237	0.763	0.882	0.938
Medium-Term	OP-ELM	+72 h	1.588	0.969	0.0661	0.248	0.744	0.865	0.928
		+96 h	1.751	1.082	0.0920	0.279	0.688	0.837	0.913
	MLPNN	+72 h	1.594	0.980	0.0172	0.267	0.742	0.862	0.925
		+96 h	1.720	1.088	0.0176	0.315	0.699	0.837	0.909
Long-Term	OP-ELM	+120 h	1.784	1.151	0.0279	0.325	0.676	0.824	0.902
		+168 h	1.930	1.258	0.0816	0.333	0.621	0.794	0.887
	MLPNN	+120 h	1.794	1.139	0.0913	0.291	0.672	0.825	0.905
		+168 h	1.933	1.264	0.0479	0.342	0.619	0.791	0.883

and MLPNN) displayed the highest level of accuracy for both training and validation. For the OP-ELM model, in the training phase, the RMSE, MAE and Bias ranged from 1.081 to 1.785, 0.648 to 1.268, and 0, respectively. The NSE, R and d ranged from 0.779 to 0.919, 0.883 to 0.959, and 0.935 to 0.979, respectively. Likewise, for MLPNN, the RMSE, MAE and Bias ranged from 1.054 to 1.741, 0.635 to 1.241, and -0.0001 to 0.0033, respectively. The NSE, R and d ranged from 0.790 to 0.923, 0.889 to 0.961, and 0.938 to 0.980, respectively. From Table 4, it can be observed that model FM1, whose output is the DO at $(t + 24)$, performed better than the other models in training; while the worst results were obtained for FM6 model, whose output is the DO at $(t + 168)$. In the validation phase, both MLPNN and OP-ELM models performed quite well in forecasting hourly DO. According to Table 4, for the OP-ELM model, the RMSE, MAE and Bias ranged from 1.225 to 1.930, 0.665 to 1.258, and 0.0430 to 0.0920, respectively. The NSE, R and d ranged from 0.621 to 0.848, 0.794 to 0.922, and 0.887 to 0.959, respectively. Likewise, for MLPNN, the RMSE, MAE and Bias ranged from 1.270 to 1.933, 0.697 to 1.264, and 0.0479 to 0.1055. The NSE, R and d ranged from 0.619 to 0.836, 0.791 to 0.919, and 0.883 to 0.958, respectively. From Table 4, we conclude that the accuracies of the OP-ELM model predictions were usually slightly better than those of the MLPNN model. This demonstrates that the OP-ELM model is more suitable for forecasting DO.

For reasons explained earlier in the section on model performance indices, an analysis of the MAPE is required to be presented. In the training phase, as shown in Table 4, the values of the MAPE ranged from 0.337 % to 0.784 % for OP-ELM and from 0.315 % to 0.750 % for MLPNN. For all the models, the MAPE of the OP-ELM shows slightly larger value than that of MLPNN models. In the validation phase, as shown in Table 4, the MAPE values ranged from 0.176 % to 0.333 % for OP-ELM and from 0.201 % to 0.342 % for MLPNN. For all the models, the MAPE values of the MLPNN were slightly larger than those of OP-ELM models. However, generally OP-ELM models performed better than MLPNN in the validation phase. To sum up, among the six forecasting models (FM1 to FM6), OP-ELM models showed the best performances in the validation phase. Examination of the MAPE values is another way to evaluate the developed models, and an important question must be answered: although the performances of the models are good according to the others six performance criteria (i.e., RMSE, MAE, Bias, R, NSE and d), why the values of the MAPE are very high, especially in the training phase? First, using the MAPE, the performance of the model can be easily judged. The MAPE is only 0.5 % if the measured value is 0.10 mg/L and the predicted value is 0.105 mg/L. But MAPE is 84 % if the measured value is 0.10 mg/L and the predicted value 0.184 mg/L. Second, as mentioned, the disadvantages of the MAPE include that it can assume a very high value (the case of the present study) if the actual measured value is very small (Kim and Kim 2016; Dawson et al. 2007; Tayman and Swanson 1999; Rayer 2007). For example, for a measured DO of 0.10 mg/L, which is frequently observed in the present study, and a calculated DO value of 8.567 mg/L, the MAPE is 8467 %, which is very high and unacceptable.

In order to demonstrate the capabilities and usefulness of the proposed approaches, we present in Table 5 a comparison between measured and calculated mean and standard deviation, the RPD index, the SI and the RMAE. Values of the RMSE and MAE less than half of the standard deviations (SD) of the observations may be considered low (Singh et al. 2004; Moriasi et al. 2007). As stated above, all RPD values less than one are unacceptable while values greater than 3 are considered excellent. According to Table 5, for both training and validation, there were small differences between model estimates and measured DO values, according to the mean and standard deviation. All RMAE values were less than half of the standard deviation and all RPD values were higher than 1.60. It can be seen from

Table 5 Comparison between Mean and Standard deviation of the measured (Mes) and forecasted (Cal) DO values, ratio of standard deviation (S_x) to RMSE (RPD), scatter index (SI) and ratio of the mean absolute error to standard deviation (RMAE), for USGS 420853121505500 [Top] station

Station: Klamath River at Miller Island Boat Ramp [Top] (420,853,121,505,500)

Horizon	Models	Forecasting Interval	X_{mean}		S_x		RPD	RMAE	SI
			Cal	Mes	Cal	Mes			
Training Phase									
Short-Term	OP-ELM	+24 h	7.516	7.516	3.645	3.802	3.610	0.167	0.144
		+48 h	7.516	7.516	3.575	3.801	2.976	0.219	0.172
	MLPNN	+24 h	7.516	7.516	3.652	3.802	3.521	0.170	0.140
		+48 h	7.515	7.516	3.580	3.801	2.941	0.221	0.170
Medium-Term	OP-ELM	+72 h	7.516	7.516	3.523	3.801	2.703	0.250	0.190
		+96 h	7.515	7.515	3.476	3.801	2.513	0.273	0.205
	MLPNN	+72 h	7.515	7.516	3.531	3.801	2.667	0.253	0.187
		+96 h	7.509	7.515	3.483	3.801	2.469	0.277	0.201
Long-Term	OP-ELM	+120 h	7.515	7.515	3.428	3.800	2.358	0.295	0.218
		+168 h	7.515	7.515	3.355	3.800	2.183	0.327	0.238
	MLPNN	+120 h	7.517	7.515	3.442	3.800	2.315	0.299	0.215
		+168 h	7.512	7.515	3.376	3.800	2.128	0.334	0.232
Validation Phase									
Short-Term	OP-ELM	+24 h	8.477	8.520	3.023	3.139	2.564	0.212	0.144
		+48 h	8.484	8.518	2.882	3.138	2.169	0.272	0.170
	MLPNN	+24 h	8.414	8.520	3.142	3.139	2.469	0.222	0.187
		+48 h	8.385	8.518	3.136	3.138	2.053	0.283	0.202
Medium-Term	OP-ELM	+72 h	8.451	8.517	2.920	3.137	1.976	0.309	0.149
		+96 h	8.423	8.515	2.956	3.136	1.792	0.345	0.179
	MLPNN	+72 h	8.499	8.517	2.816	3.137	1.969	0.312	0.210
		+96 h	8.497	8.515	2.734	3.136	1.821	0.347	0.227
Long-Term	OP-ELM	+120 h	8.485	8.513	2.728	3.135	1.757	0.367	0.186
		+168 h	8.428	8.509	2.791	3.133	1.623	0.402	0.206
	MLPNN	+120 h	8.422	8.513	2.859	3.135	1.748	0.363	0.211
		+168 h	8.461	8.509	2.703	3.133	1.621	0.403	0.227

Table 5, in the validation phase, for the MLPNN and OP-ELM models, that all the differences between the mean of the measurements and the corresponding calculated DO values are less than 1.0 %. However, it is worth noting that there are only 5 cases in which the difference between the measured and the corresponding calculated standard deviations are higher than 10 % with a maximum of 13.72 % and a minimum of 0.6 %. This finding implies that the proposed models are very good with a high level of accuracy and robustness.

For a final verification of the accuracy of the developed forecasting models, we also present in Table 5 the values of the SI index in the training and validation phases, for both MLPNN and OP-ELM models. In the training phase, the SI of the six developed models (FM1 to FM6) ranged between 0.144 and 0.238 for the OP-ELM models, and between 0.140 and 0.232 for the MLPNN models. Furthermore, for all the developed models, the SI was less than 0.240

which indicates excellent modeling results. Similarly, in the validation phase, the SI of the six models ranged between 0.144 and 0.206 for the OP-ELM models, and between 0.187 and 0.227 for the MLPNN models. Furthermore, for all developed models, the SI was less than 0.230, which indicates excellent modeling results. Once again, according to the SI, the OP-ELM is again better. In conclusion, according to the results obtained in the validation phase, the OP-ELM models must be ranked as best. Figures 7 and 8 show scatter plots of the calculated versus observed DO for the OP-ELM and MLPNN models, respectively, in the validation phase. Figures 9 and 10 show the comparison between the calculated and measured DO for OP-ELM and MLPNN models, respectively, in the validation phase.

3.2 Forecasting DO Concentration at the USGS 420853121505501 [Bottom] Station

The MLPNN and OP-ELM models with the same input and different outputs were compared based on their performance in training and validation sets. The results are summarized in Tables 6 and 7. It is observed that all models generally gave low values of the RMSE, MAE and bias, and high values of R, d and NSE. Generally, the performances of both the MLPNN and OP-ELM models in DO forecasting were very satisfactory. The first important conclusion is that the two models (MLPNN and OP-ELM) showed higher accuracy in prediction and performed better compared to the results obtained in Top station. Overall, they deliver a higher degree of reliability. It can be seen from the results that the values of all seven performance indices change with respect to the step ahead. Therefore, the two models appear to be performing at the same level of accuracy, but the OP-ELM offers more accuracy in the validation phase. In the training phase, for the OP-ELM model, the RMSE, MAE, and bias ranged from 0.662 to 1.534, 0.425 to 1.084, and 0, respectively. The NSE, R, and d ranged from 0.860 to 0.974, 0.927 to 0.987, and 0.961 to 0.993, respectively. Likewise, for MLPNN, the RMSE, MAE, and bias ranged from 0.656 to

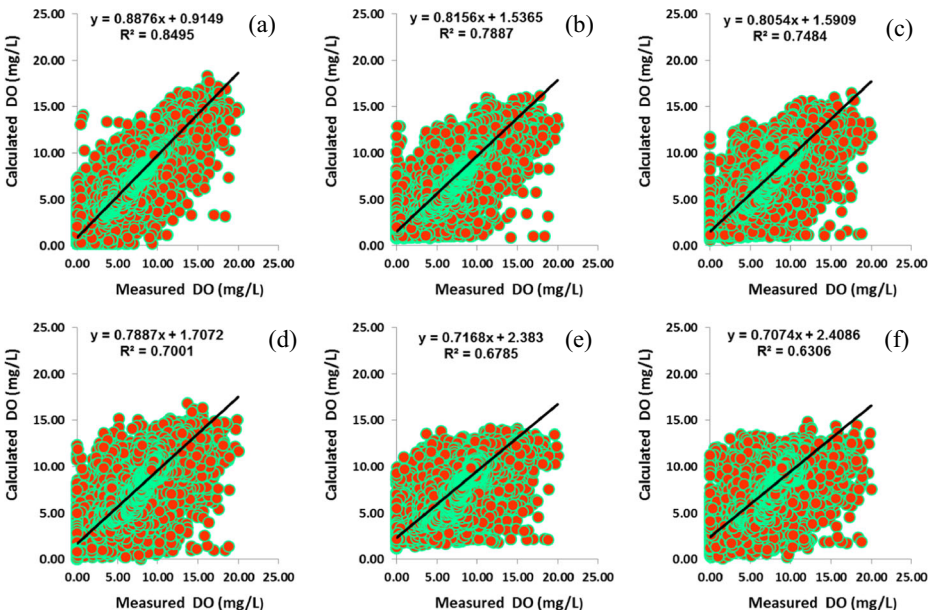


Fig. 7 Scatterplots of calculated versus measured values of dissolved oxygen (DO) in the validation phase, using OP-ELM model for the USGS 420853121505500 [Top] Station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

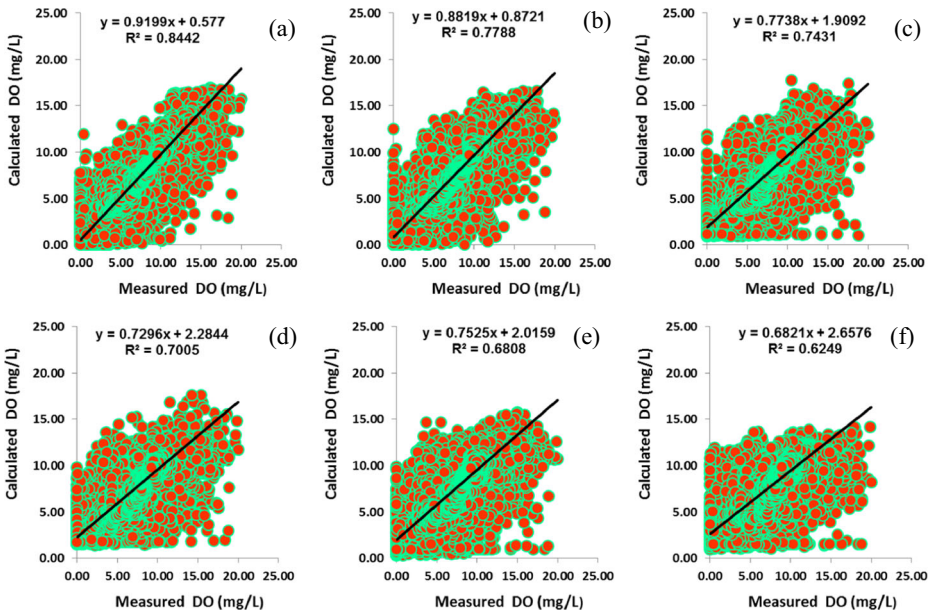


Fig. 8 Scatterplots of calculated versus measured values of dissolved oxygen (DO) in the validation phase, using MLPNN model for the USGS 420853121505500 [Top] Station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

1.502, 0.421 to 1.065, and -0.0042 to 0.0198 , respectively, and the NSE, R, and d ranged from 0.865 to 0.974, 0.930 to 0.987, and 0.963 to 0.993, respectively. In the validation phase, both MLPNN and OP-ELM models performed quite well in forecasting hourly DO. According to

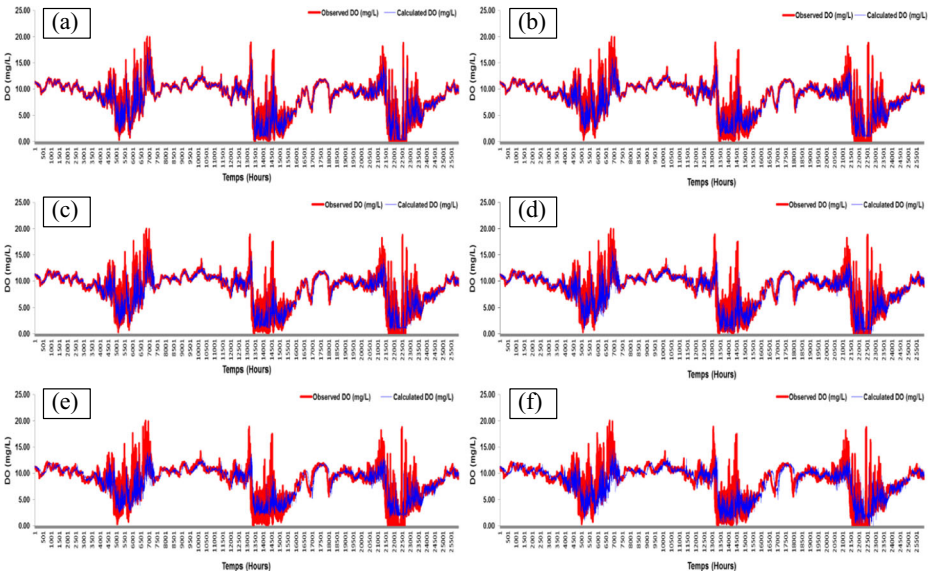


Fig. 9 Comparison between the calculated and measured data of dissolved oxygen (DO) in the validation phase, using OP-ELM model for the USGS 420853121505500 [Top] Station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

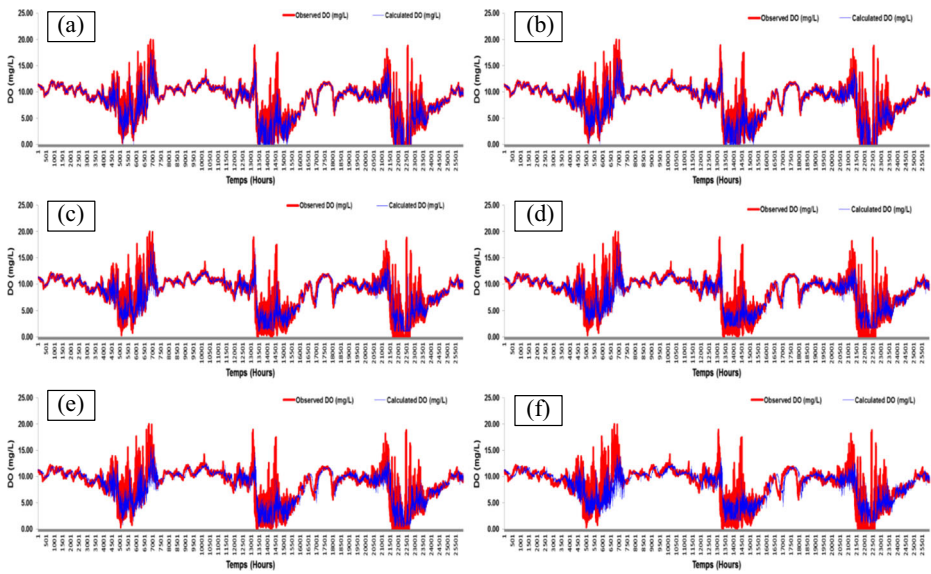


Fig. 10 Comparison between the calculated and measured data of dissolved oxygen (DO) in the validation phase, using MLPNN model for the USGS 420853121505500 [Top] Station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

Table 6, for the OP-ELM model, the RMSE, MAE, and bias ranged from 0.708 to 1.454, 0.425 to 0.995, and 0.00 to -0.0736 , respectively, and the NSE, R and d ranged from 0.829 to 0.960, 0.911 to 0.980, and 0.952 to 0.990, respectively. Likewise, for MLPNN, the RMSE, MAE, and bias ranged from 0.730 to 1.489, 0.455 to 1.057, and -0.0582 to -0.1224 , respectively, and the NSE, R and d ranged from 0.821 to 0.957, 0.906 to 0.979, and 0.948 to 0.989, respectively. According to Table 6, for all six models, the performance in the training phase was better than the performance in the validation phase. Nevertheless, the FM1 model DO output at (+24 h) must be considered as the best model developed. The FM6 model (+168 h) performed poorer than all the other models in terms of the seven performance indices. For the OP-ELM models, in the training phase, the lowest values of the RMSE, MAE, and Bias of forecasting models were 0.662 (mg/L), 0.425 (mg/L), and 0.00 (mg/L) (for OP-ELM FM1), and the highest values of the NSE, R and d were 0.974, 0.987 and 0.993 (for OP-ELM FM1). Table 6 indicates that the OP-ELM (FM1) had the lowest values of the RMSE and MAE (0.708 mg/L, 0.425 mg/L), and the highest NSE, R and d (0.960, 0.980, and 0.990) in the validation phase. According to Table 6, for the MLPNN models, the lowest values of the RMSE, MAE, and Bias of forecasting models were 0.656 (mg/L), 0.421 (mg/L), and -0.0042 (mg/L) (in MLPNN FM1), and the highest values of the NSE, R, and d were 0.974, 0.987 and 0.993 (in MLPNN FM1). Table 6 indicates that the MLPNN (FM1) had the lowest values of the RMSE and MAE (0.730 mg/L, 0.455 mg/L), and the highest NSE, R and d (0.957, 0.979 and 0.989) in the validation phase. From the results of training and validation, all six models developed in this study were evaluated all together, and results are conspicuous. In the validation phase, the OP-ELM models outperformed MLPNN models in terms of the various performance criteria. Finally, we report again the MAPE values in the training and validation phases. In the training phase, as shown in Table 6, MAPE values ranged from 0.243 % to 0.825 % for OP-ELM and from 0.244 % to 0.783 % for MLPNN. MAPE values for all OP-ELM models were slightly larger than those for MLPNN models. In the validation phase, as shown in Table 6, MAPE values ranged from 0.118 % to 0.278 % for OP-

Table 6 Performance of the OP-ELM and MLPNN forecasting models in the training and validation phases several hours in advance for USGS 420853121505501 [Bottom] station

Station: Klamath River at Miller Island Boat Ramp [Bottom] (420,853,121,505,501)

	Model	Forecasting Interval	RMSE (mg/L)	MAE (mg/L)	Bias (mg/L)	MAPE (%)	NSE /	R /	d /
Training Phase									
Short-Term	OP-ELM	+24 h	0.662	0.425	0.0000	0.243	0.974	0.987	0.993
		+48 h	0.883	0.597	0.0000	0.365	0.954	0.977	0.988
	MLPNN	+24 h	0.656	0.421	-0.0042	0.244	0.974	0.987	0.993
		+48 h	0.874	0.591	0.0024	0.357	0.954	0.977	0.988
Medium-Term	OP-ELM	+72 h	1.057	0.724	0.0000	0.455	0.933	0.966	0.983
		+96 h	1.201	0.825	0.0000	0.524	0.914	0.956	0.977
	MLPNN	+72 h	1.042	0.716	0.0028	0.442	0.935	0.967	0.983
		+96 h	1.173	0.809	0.0027	0.510	0.918	0.958	0.978
Long-Term	OP-ELM	+120 h	1.313	0.912	0.0000	0.611	0.897	0.947	0.972
		+168 h	1.534	1.084	0.0000	0.825	0.860	0.927	0.961
	MLPNN	+120 h	1.300	0.906	-0.0037	0.602	0.899	0.948	0.973
		+168 h	1.502	1.065	0.0198	0.783	0.865	0.930	0.963
Validation Phase									
Short-Term	OP-ELM	+24 h	0.708	0.425	-0.0304	0.118	0.960	0.980	0.990
		+48 h	0.916	0.594	-0.0383	0.167	0.932	0.966	0.982
	MLPNN	+24 h	0.730	0.455	-0.0692	0.134	0.957	0.979	0.989
		+48 h	0.954	0.635	-0.0870	0.181	0.927	0.963	0.980
Medium-Term	OP-ELM	+72 h	1.057	0.700	-0.0336	0.192	0.910	0.954	0.976
		+96 h	1.171	0.784	0.0000	0.215	0.889	0.943	0.970
	MLPNN	+72 h	1.087	0.738	-0.0591	0.202	0.905	0.951	0.974
		+96 h	1.181	0.820	-0.0476	0.225	0.887	0.942	0.969
Long-Term	OP-ELM	+120 h	1.308	0.912	-0.0736	0.259	0.862	0.929	0.961
		+168 h	1.454	0.995	-0.0036	0.278	0.829	0.911	0.952
	MLPNN	+120 h	1.379	0.969	-0.1224	0.271	0.847	0.922	0.955
		+168 h	1.489	1.057	-0.0582	0.295	0.821	0.906	0.948

ELM and from 0.134 % to 0.295 % for MLPNN. MAPE values for all MLPNN models were slightly larger compared to those for OP-ELM models. However, generally OP-ELM performed better than MLPNN in the validation phase. To sum up, among the six forecasting models (FM1 to FM6), OP-ELM models showed better performances in the validation phase. In conclusion, the OP-ELM models are better for DO forecasting.

We present in Table 7 a comparison between measured and calculated mean and standard deviation, the RPD index, the SI and the ratio of the RMAE. According to Table 7, for both training and validation there were small differences between model estimates and measured DO values, based on the mean and standard deviation. In the training phase, all the ratios of RMAE were less than 0.5 and all the RPD were higher than 3, except for the FM6 model for both OP-ELM and MLPNN. It can be seen from Table 7, in the validation phase, for both MLPNN and OP-ELM models, that all the difference between the mean of the measurements and the corresponding calculated DO values were less than 1.0 %. However, it is worth noting

Table 7 Comparison between Mean and Standard deviation of the measured (Mes) and forecasted (Cal) values of DO, ratio of standard deviation (S_x) to RMSE (RPD), scatter index (SI), and ratio of the mean absolute error to standard deviation (RMAE), for USGS 420853121505501 [Bottom] station

Station: Klamath River at Miller Island Boat Ramp [Bottom] (420,853,121,505,501)

Horizon	Models	Forecasting Interval	X_{mean}		S_x		RPD	RMAE	SI
			Cal	Mes	Cal	Mes			
Training Phase									
Short-Term	OP-ELM	+24 h	6.570	6.570	4.044	4.097	6.173	0.104	0.101
		+48 h	6.569	6.569	4.000	4.097	4.630	0.146	0.134
Medium-Term	MLPNN	+24 h	6.574	6.570	4.060	4.097	6.250	0.103	0.159
		+48 h	6.566	6.569	4.006	4.097	4.695	0.144	0.179
	OP-ELM	+72 h	6.568	6.568	3.957	4.096	3.876	0.177	0.100
		+96 h	6.568	6.568	3.915	4.095	3.413	0.201	0.133
Long-Term	MLPNN	+72 h	6.565	6.568	3.957	4.096	3.937	0.175	0.200
		+96 h	6.565	6.568	3.922	4.095	3.497	0.198	0.234
	OP-ELM	+120 h	6.567	6.567	3.879	4.095	3.115	0.223	0.161
		+168 h	6.566	6.566	3.795	4.094	2.667	0.265	0.183
	MLPNN	+120 h	6.571	6.567	3.875	4.095	3.155	0.221	0.198
		+168 h	6.546	6.566	3.811	4.094	2.725	0.260	0.229
Validation Phase									
Short-Term	OP-ELM	+24 h	7.350	7.320	3.434	3.522	4.975	0.121	0.097
		+48 h	7.357	7.319	3.368	3.522	3.846	0.169	0.125
	MLPNN	+24 h	7.389	7.320	3.374	3.522	4.831	0.129	0.148
		+48 h	7.406	7.319	3.284	3.522	3.690	0.180	0.161
Medium-Term	OP-ELM	+72 h	7.352	7.319	3.352	3.522	3.333	0.199	0.100
		+96 h	7.318	7.318	3.363	3.521	3.003	0.223	0.130
	MLPNN	+72 h	7.378	7.319	3.288	3.522	3.236	0.210	0.179
		+96 h	7.366	7.318	3.279	3.521	2.985	0.233	0.199
Long-Term	OP-ELM	+120 h	7.391	7.317	3.196	3.521	2.688	0.259	0.144
		+168 h	7.319	7.315	3.254	3.520	2.421	0.283	0.160
	MLPNN	+120 h	7.440	7.317	3.105	3.521	2.551	0.275	0.188
		+168 h	7.374	7.315	3.126	3.520	2.364	0.300	0.204

that there were only 2 cases where the difference between the measured and its corresponding calculated standard deviation were higher than 10 % (FM5 and FM6 for MLPNN), with a maximum of 11.81 % and a minimum of 2 %. This finding implies that the proposed models are very good with a high level of accuracy and robustness. We also present in Table 7 the SI index values in the training and validations phases, for both MLPNN and OP-ELM models. In the training phase, the SI of the six developed models (FM1 to FM6) ranged between 0.101 and 0.183 for OP-ELM models, and between 0.159 and 0.229 for MLPNN models. Furthermore, for all developed models, the SI was less than 0.240, which corresponds to excellent modeling. Similarly, in the validation phase, the SI of the six developed models ranged between 0.097 and 0.160 for OP-ELM models, and between 0.148 and 0.204 for MLPNN models. Furthermore, for all developed models, the SI was less than 0.210, which corresponds

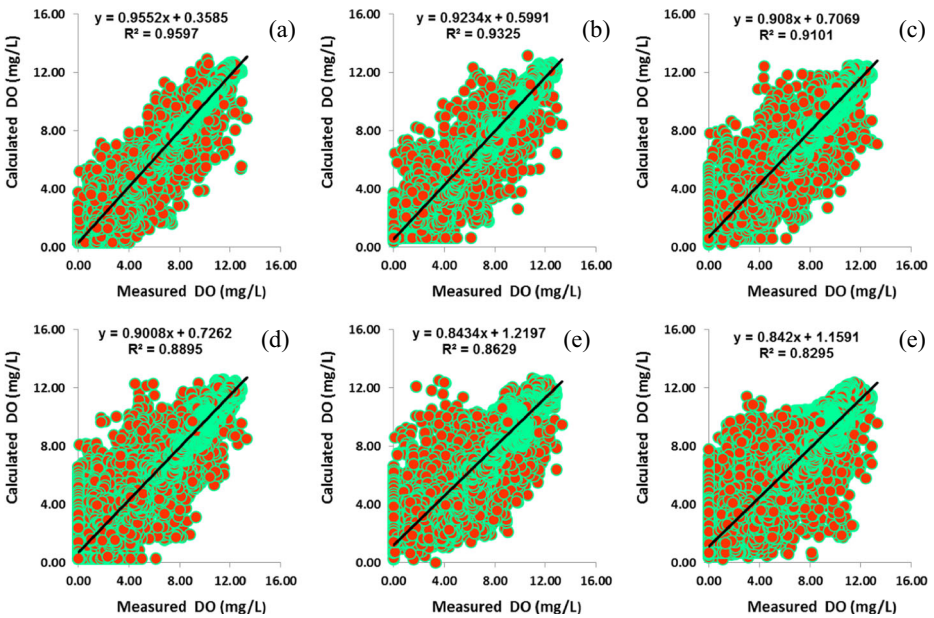


Fig. 11 Scatterplots of calculated versus measured values of dissolved oxygen (DO) in the validation phase, using OP-ELM model for the USGS 420853121505501 [Bottom] Station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

to excellent modeling. Once again, according to the SI values, the OP-ELM is better. In conclusion, according to the results obtained in the validation phase, the OP-ELM model must be ranked as better model. Figures 11 and 12 show scatter plots of the calculated versus

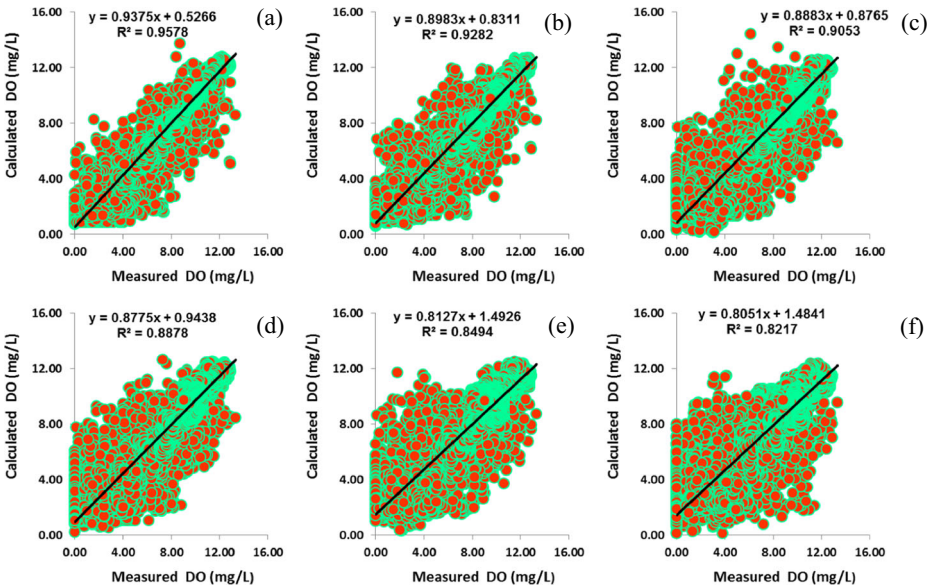


Fig. 12 Scatterplots of calculated versus measured values of dissolved oxygen (DO) in the validation phase, using MLPNN model for the USGS 420853121505501 [Bottom] Station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

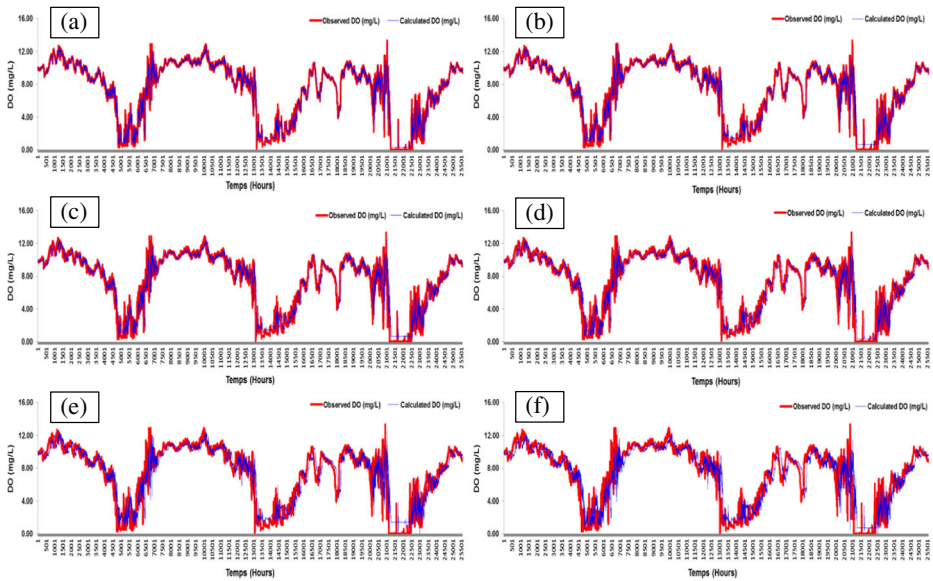


Fig. 13 Comparison between the calculated and measured data of dissolved oxygen (DO) in the validation phase, using OP-ELM model for the USGS 420853121505501 [Bottom] station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

observed DO for OP-ELM and MLPNN models, respectively, in the validation phase. Figures 13 and 14 show the comparison between the calculated and measured DO for OP-ELM and MLPNN models, respectively, in the validation phase.

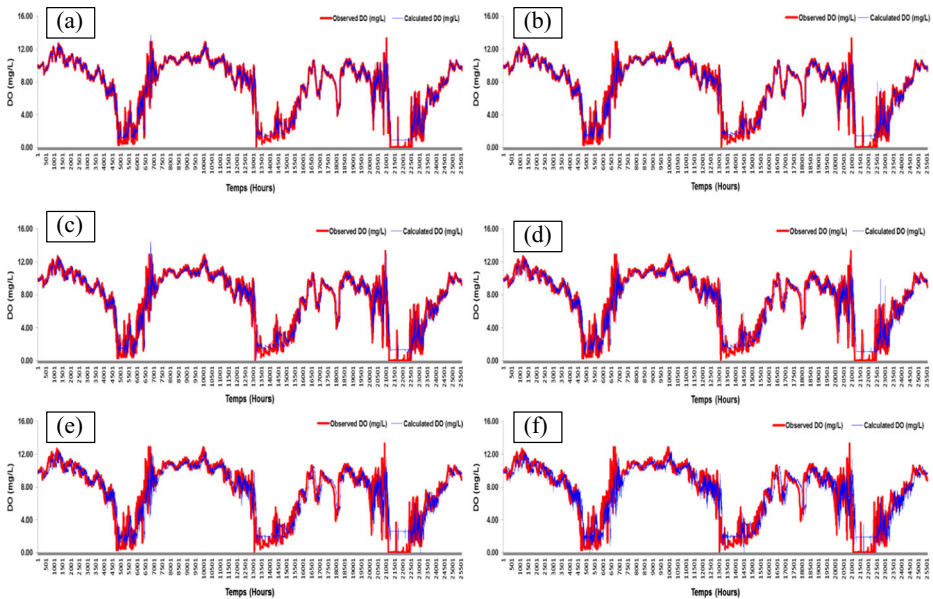


Fig. 14 Comparison between the calculated and measured data of dissolved oxygen (DO) in the validation phase, using MLPNN model for the USGS 420853121505501 [Bottom] station: **a** +24 h, **b** +48 h, **c** +72 h, **d** +96 h, **e** +120 h, and **f** +168 h

4 Conclusions

In this paper, we presented and compared two models to forecast DO several hours in advance, MLPNN and OP-ELM, which are based on artificial intelligence paradigm. To show the applicability of the proposed methods, actual values of measured DO in Klamath River at Miller Island Boat Ramp, Oregon, USA were used. The data were collected from 2004 to 2013, at an hourly time step. The study showed the development of a powerful tool, forecasting DO values at good accuracy (R between 0.91 and 0.98) in the validation phase. The results are very encouraging because the development of these tools was undertaken to support the actual methods and models for estimating DO in rivers without using other water quality parameters. This is the first publication that we are aware of, which demonstrates the use of OP-ELM and MLPNN, in forecasting DO several hours in advance up to (7 days). In the validation phase, the results show that for both stations, Top and Bottom, the OP-ELM model ensured more accurate forecasting than the MLPNN model.

Acknowledgments The authors would like to thank the staff of the United States Geological Survey (USGS) for providing the data that made this research possible. We would like to thank the staff of the Research Group Applications of Machine Learning (AML), Department of Computer Science at the Aalto University School of Science, Espoo, Finland, for providing the OP-ELM toolbox. Once again, we would like to thank the anonymous reviewers and the editor of Environmental Processes for their invaluable comments and suggestions on the contents of the manuscript which significantly improved the quality of the paper. Note that it has not gone unnoticed nor unappreciated. This makes me want to submit more papers to ENPR journal in the future.

References

- Ackerson JP, Dematté JAM, Morgan CLS (2015) Predicting clay content on field-moist intact tropical soils using a dried, ground VisNIR library with external parameter orthogonalization. *Geoderma* 259–260:196–204. doi:[10.1016/j.geoderma.2015.06.002](https://doi.org/10.1016/j.geoderma.2015.06.002)
- Adamala S, Raghuvanshi NS, Mishra A (2015) Generalized quadratic synaptic neural networks for ET₀ modeling. *Environ Process* 2:309–329. doi:[10.1007/s40710-015-0066-6](https://doi.org/10.1007/s40710-015-0066-6)
- Akkoyunlu A, Altun H, Cigizoglu H (2011) Depth-integrated estimation of dissolved oxygen in a lake. *ASCE J Environ Eng* 137(10):961–967. doi:[10.1061/\(ASCE\)EE.1943-7870.0000376](https://doi.org/10.1061/(ASCE)EE.1943-7870.0000376)
- Akusok A, Veganzones D, Miche Y, Björk K-M, du Jardin P, Severin E, Lendasse A (2015) MD-ELM: originally mislabeled samples detection using OP-ELM model. *Neurocomputing* 159:242–250. doi:[10.1016/j.neucom.2015.01.055](https://doi.org/10.1016/j.neucom.2015.01.055)
- Alizadeh MJ, Kavianpour MR (2015) Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Mar Pollut Bull* 98:171–178. doi:[10.1016/j.marpolbul.2015.06.052](https://doi.org/10.1016/j.marpolbul.2015.06.052)
- Altunkaynak A, Ozger M, Cakmakci M (2005) Fuzzy logic modeling of the dissolved oxygen fluctuations in Golden Horn. *Ecol Model* 189:436–446. doi:[10.1016/j.ecolmodel.2005.03.007](https://doi.org/10.1016/j.ecolmodel.2005.03.007)
- An Y, Zou Z, Zhao Y (2015) Forecasting of dissolved oxygen in the Guanting reservoir using an optimized NGBM (1,1) model. *Journal of Environmental Sciences*. (29):158–164. doi:[10.1016/j.jes.2014.10.005](https://doi.org/10.1016/j.jes.2014.10.005)
- Antanasijević D, Pocažt V, Povrenović D, Perić-Grujić A, Ristić M (2013) Modelling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study. *Environ Sci Pollut Res* 20: 9006–9013. doi:[10.1007/s11356-013-1876-6](https://doi.org/10.1007/s11356-013-1876-6)
- Antanasijević D, Pocažt V, Povrenović D, Perić-Grujić A, Ristić M (2014) Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo simulation uncertainty analysis. *J Hydrol* 519:1895–1907. doi:[10.1016/j.jhydrol.2014.10.009](https://doi.org/10.1016/j.jhydrol.2014.10.009)
- Antonopoulos VZ, Georgiou PE, Antonopoulos ZV (2015) Dispersion coefficient prediction using empirical models and ANNs. *Environ Process* 2:379–394. doi:[10.1007/s40710-015-0074-6](https://doi.org/10.1007/s40710-015-0074-6)
- Areerachakul S, Sophatsathit P, Lursinsap C (2013) Integration of unsupervised and supervised neural networks to predict dissolved oxygen concentration in canals. *Ecol Model* 261(262):1–7. doi:[10.1016/j.ecolmodel.2013.04.002](https://doi.org/10.1016/j.ecolmodel.2013.04.002)

- Ay M, Kisi O (2012) Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. *ASCE J Environ Eng* 138(6):654–662. doi:10.1061/(ASCE)EE.1943-7870.0000511
- Azad S, Debnath S, Rajeevan M (2015) Analysing predictability in Indian monsoon rainfall: a data analytic approach. *Environ Process* 2(1):717–727. doi:10.1007/s40710-015-0108-0
- Boskidis I, Gikas GD, Pisinaras V, Tsihrintzis VA (2010) Spatial and temporal changes of water quality, and SWAT modeling of Vosvozis river basin, North Greece. *J Environ Sci Health-Part A* 45(11):1421–1440. doi:10.1080/10934529.2010.500936
- Boskidis I, Gikas GD, Sylaios G, Tsihrintzis VA (2011) Water quantity and quality assessment of lower Nestos river, Greece. *J Environ Sci Health-Part A* 46:1050–1067. doi:10.1080/10934529.2011.590381
- Bowden GJ, Maier HR, Dandy GC (2002) Optimal division of data for neural network models in water resources applications. *Water Resour Res* 38(2):1010. doi:10.1029/2001WR000266
- Bowden GJ, Dandy GC, Maier HR (2005a) Input determination for neural network models in water resources applications. Part 1-background and methodology. *J Hydrol* 301(1–4):75–92. doi:10.1016/j.jhydrol.2004.06.021
- Bowden GJ, Dandy GC, Maier HR (2005b) Input determination for neural network models in water resources applications. Part 2. Case study: forecasting salinity in a river. *J Hydrol* 301(1–4):93–104. doi:10.1016/j.jhydrol.2004.06.021
- Breitburg DL, Adamack A, Rose KA, Kolesar SE, Decker MB, Purcell JE, Keister JE, Cowan JH (2003) The pattern and influence of low dissolved oxygen in the Patuxent River, a seasonally hypoxic estuary. *Estuaries* 26(2):280–297. doi:10.1007/BF02695967
- Cao J, Lin Z, Huang GB (2012) Self-adaptive evolutionary extreme learning machine. *Neural Process Lett* 36:285–305. doi:10.1007/s11063-012-9236-y
- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature *Geosci Model Dev* 7:1247–1250. doi:10.5194/gmd-7-1247-2014
- Chamoglou M, Papadimitriou T, Kagalou I (2014) Key-descriptors for the functioning of a Mediterranean reservoir: the case of the new Lake Karla-Greece. *Environ Process* 1:127–135. doi:10.1007/s40710-014-0011-0
- Chang CW, Laird DA, Mausbach MJ, Hurlburgh CR (2001) Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci Soc Am J* 65:480–490. doi:10.2136/sssaj2001.652480x
- Chase C (2013) *Demand-Driven Forecasting: A Structured Approach to Forecasting*, 2nd Edition. Hoboken, NJ, USA: Wiley. ISBN: 978-1-118-66939-6, pp 384.
- Cox BA (2003a) A review of dissolved oxygen modelling techniques for lowland rivers. *Sci Total Environ* 314(316):303–334. doi:10.1016/S0048-9697(03)00062-7
- Cox BA (2003b) A review of currently available in-stream water quality models and their applicability for simulating dissolved oxygen in lowland rivers. *Sci Total Environ* 314-316:335–377. doi:10.1016/S0048-9697(03)00063-9
- Das DB, Thirakulchaya T, Deka L, Hanspal NS (2015) Artificial neural network to determine dynamic effect in capillary pressure relationship for two-phase flow in porous media with micro-heterogeneities. *Environ Process* 2:1–18. doi:10.1007/s40710-014-0045-3
- Dawson CW, Wilby RL (2001) Hydrological modelling using artificial neural networks. *Prog Phys Geogr* 25(1):80–108. doi:10.1177/030913330102500104
- Dawson CW, Harpham C, Wilby RL, Chen Y (2002) Evaluation of artificial neural network techniques for flow forecasting in the river Yangtze. *China Hydrol Earth Syst Sci* 6:619–626. doi:10.5194/hess-6-619-2002
- Dawson CW, Abraham RJ, See LM (2007) HydroTest: a web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environ Model Softw* 22:1034–1052. doi:10.1016/j.envsoft.2006.06.008
- Deo RC, Wen X, Qi F (2016) A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl Energy* 168:568–593. doi:10.1016/j.apenergy.2016.01.130
- Efron B, Hastie T, Johnstone I, Tibshirani R (2004) Least angle regression. *Ann Stat* 32:407–499. doi:10.1214/009053604000000067
- Emamgholizadeh S, Kashi H, Marofpoor I, Zalaghi E (2014) Prediction of water quality parameters of Karoon River (Iran) by artificial intelligence-based models. *Int J Environ Sci Technol* 11:645–656. doi:10.1007/s13762-013-0378-x
- Evrendilek F, Karakaya N (2014) Regression model-based predictions of diel, diurnal and nocturnal dissolved oxygen dynamics after wavelet denoising of noisy time series. *Physica A* 404:8–15. doi:10.1016/j.physa.2014.02.062
- Evrendilek F, Karakaya N (2015) Spatiotemporal modeling of saturated dissolved oxygen through regressions after wavelet denoising of remotely and proximally sensed data. *Earth Sci Inf* 8:247–254. doi:10.1007/s12145-014-0148-4

- Faruk DÖ (2010) A hybrid neural network and ARIMA model for water quality time series prediction. *Eng Appl Artif Intell* 23:586–594. doi:10.1016/j.engappai.2009.09.015
- Friedrich et al (2014) Investigating hypoxia in aquatic environments: diverse approaches to addressing a complex phenomenon. *Biogeosciences*, 11:1215–1259. doi:10.5194/bg-11-1215-2014.
- Gebremariam SY, Martin JF, DeMarchi C, Bosch NS, Confesor R, Ludsin SA (2014) A comprehensive approach to evaluating watershed models for predicting river flow regimes critical to downstream ecosystem services. *Environ Model Softw* 61:121–134. doi:10.1016/j.envsoft.2014.07.004
- Gikas GD (2014) Water quality of drainage canals and assessment of nutrient loads using QUAL2Kw. *Environ Process* 1:369–385. doi:10.1007/s40710-014-0027-5
- Gikas GD, Yiannakopoulou T, Tsihrintzis VA (2006) Modeling of non-point source pollution in a Mediterranean drainage basin. *Environ Model Assess* 11:219–233. doi:10.1007/s10666-005-9017-3
- Grigorievskiy A, Miche Y, Ventelä AM, Séverin E, Lendasse A (2014) Long-term time series prediction using OP-ELM. *Neural Netw* 51:50–56. doi:10.1016/j.neunet.2013.12.002
- Haykin S (1999) Neural networks: a comprehensive foundation. Prentice Hall, Upper Saddle River
- Heddam S (2014a) Generalized regression neural network (GRNN) based approach for modelling hourly dissolved oxygen concentration in the upper Klamath River, Oregon, USA. *Environ Technol* 35(13): 1650–1657. doi:10.1080/09593330.2013.878396
- Heddam S (2014b) Modelling hourly dissolved oxygen concentration (DO) using two different adaptive neuro-fuzzy inference systems (ANFIS): a comparative study. *Environ Monit Assess* 186:597–619. doi:10.1007/s10661-013-3402-1
- Heddam S (2014c) Modelling hourly dissolved oxygen concentration (DO) using dynamic evolving neural-fuzzy inference system (DENFIS) based approach: case study of Klamath River at Miller Island Boat Ramp, Oregon, USA. *Environ Sci Pollut Res* 21:9212–9227. doi:10.1007/s11356-014-2842-7
- Heddam S (2016a) Secchi disk depth estimation from water quality parameters: artificial neural network versus multiple linear regression models? *Environ Process* 3(1):525–536. doi:10.1007/s40710-016-0144-4
- Heddam S (2016b) Multilayer perceptron neural network based approach for modelling phycoyanin pigment concentrations: case study from lower Charles River buoy, USA. *Environmental Science and Pollution Research*. doi:10.1007/s11356-016-6905-9
- Heddam S, Bermad A, Dechemi N (2011) Applications of radial basis function and generalized regression neural networks for modelling of coagulant dosage in a drinking water treatment: a comparative study. *ASCE J Environ Eng* 137(12):1209–1214. doi:10.1061/(ASCE)JEE.1943-7870.0000435
- Heddam S, Bermad A, Dechemi N (2012) ANFIS-based modelling for coagulant dosage in drinking water treatment plant: a case study. *Environ Monit Assess* 184:1953–1971. doi:10.1007/s10661-011-2091-x
- Heddam S, Lamda H, Filali S (2016) Predicting effluent biochemical oxygen demand in a wastewater treatment plant using generalized regression neural network based approach: a comparative study. *Environ Process* 3(1):153–165. doi:10.1007/s40710-016-0129-3
- Hornik K (1991) Approximation capabilities of multilayer feedforward networks. *Neural Netw*, 4(2):251–257, 1991. doi:10.1016/0893-6080(91)90009-T.
- Hornik K, Stinchcombe M, White H (1989) Multilayer feedforward networks are universal approximators. *Neural Netw* 2:359–366. doi:10.1016/0893-6080(89)90020-8
- Huang GB (2015) What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. *Cogn Comput* 7:263–278. doi:10.1007/s12559-015-9333-0
- Huang GB, Chen L (2007) Convex incremental extreme learning machine. *Neurocomputing* 70(16–18):3056–3062. doi:10.1016/j.neucom.2007.02.009
- Huang GB, Chen L (2008) Enhanced random search based incremental extreme learning machine. *Neurocomputing* 71(16–18):3460–3468. doi:10.1016/j.neucom.2007.10.008
- Huang GB, Chen L, Siew CK (2006a) Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans Neural Netw* 17(4):879–892. doi:10.1109/TNN.2006.875977
- Huang GB, Zhu QY, Siew CK (2006b) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501. doi:10.1016/j.neucom.2005.12.126
- Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. *Int J Mach Learn Cybern* 2:107–122. doi:10.1007/s13042-011-0019-y
- Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. *Neural Netw* 61:32–48. doi:10.1016/j.neunet.2014.10.001
- Kim S, Kim H (2016) A new metric of absolute percentage error for intermittent demand forecasts. *Int J Forecast* 32:669–679. doi:10.1016/j.ijforecast.2015.12.003
- Kingston GB, Maier HR, Lambert MF (2005) Calibration and validation of neural networks to ensure physically plausible hydrological modeling. *J Hydrol* 314:158–176. doi:10.1016/j.jhydrol.2005.03.013

- Kisi O, Akbari N, Sanatipour M, Hashemi A, Teimourzadeh K, Shiri J (2013) Modeling of dissolved oxygen in river water using artificial intelligence techniques. *Journal of Environmental Informatics* 22(2):92–101. doi:[10.3808/jei.201300248](https://doi.org/10.3808/jei.201300248)
- Legates DR, McCabe GJ (1999) Evaluating the use of “goodness of fit” measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35:233–241. doi:[10.1029/1998WR900018](https://doi.org/10.1029/1998WR900018)
- Li MB, Huang GB, Saratchandran P (2005) Sundarajan N (2005) fully complex extreme learning machine. *Neurocomputing* 68:306–314. doi:[10.1016/j.neucom.2005.03.002](https://doi.org/10.1016/j.neucom.2005.03.002)
- Liang NY, Huang GB, Saratchandran P, Sundarajan N (2006) A fast and accurate online sequential learning algorithm for feedforward networks, *IEEE trans. Neural Netw* 17(6):1411–1423. doi:[10.1109/TNN.2006.880583](https://doi.org/10.1109/TNN.2006.880583)
- Liu S, Tai H, Ding Q, Li D, Xu L, Wei Y (2013) A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. *Math Comput Model* 58:458–465. doi:[10.1016/j.mcm.2011.11.021](https://doi.org/10.1016/j.mcm.2011.11.021)
- Liu S, Xu L, Jiang Y, Li D, Chen Y, Li Z (2014) A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture. *Eng Appl Artif Intell* 29:114–124. doi:[10.1016/j.engappai.2013.09.019](https://doi.org/10.1016/j.engappai.2013.09.019)
- Maier HR, Dandy GC (1996) The use of artificial neural networks for the prediction of water quality parameters. *Water Resour Res* 32(4):1013–1022. doi:[10.1029/96WR03529](https://doi.org/10.1029/96WR03529)
- Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environ Model Softw* 15(1):101–124. doi:[10.1016/S1364-8152\(99\)00007-9](https://doi.org/10.1016/S1364-8152(99)00007-9)
- Maier HR, Jain A, Dandy GC, Sudheer KP (2010) Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. *Environ Model Softw* 25(8):891–909. doi:[10.1016/j.envsoft.2014.11.028](https://doi.org/10.1016/j.envsoft.2014.11.028)
- Mandal S, Mahapatra SS, Adhikari S, Patel RK (2015) Modeling of arsenic (III) removal by evolutionary genetic programming and least square support vector machine models. *Environ Process* 2:145–172. doi:[10.1007/s40710-014-0050-6](https://doi.org/10.1007/s40710-014-0050-6)
- Mellios N, Kofinas D, Laspidou C, Papadimitriou T (2015) Mathematical modeling of trophic state and nutrient flows of Lake Karla using the PCLake model. *Environ Process* 2(Suppl 1):S85–S100. doi:[10.1007/s40710-015-0098-y](https://doi.org/10.1007/s40710-015-0098-y)
- Miche Y, Sorjamaa A, Lendasse A (2008) OP-ELM: theory, experiments and a toolbox. In: In: proceedings of the international conference on artificial neural networks. Lecture notes in computer science, Vol. 5163, Prague, Czech Republic, pp. 145–154. doi:[10.1007/978-3-540-87536-9_16](https://doi.org/10.1007/978-3-540-87536-9_16)
- Miche Y, Sorjamaa A, Bas P, Simula O, Jutten C, Lendasse A (2010) OP-ELM: optimally pruned extreme learning machine. *IEEE Trans Neural Netw* 21(1):158–162. doi:[10.1109/TNN.2009.2036259](https://doi.org/10.1109/TNN.2009.2036259)
- Mohan S, Pavan Kumar K (2016) Waste load allocation using machine scheduling: model application. *Environ Process* 3(1):139–151. doi:[10.1007/s40710-016-0122-x](https://doi.org/10.1007/s40710-016-0122-x)
- Moreno R, Corona F, Lendasse A, Graña M, Galvão LS (2014) Extreme learning machines for soybean classification in remote sensing hyperspectral images. *Neurocomputing* 128:207–216. doi:[10.1016/j.neucom.2013.03.057](https://doi.org/10.1016/j.neucom.2013.03.057)
- Moriasi DN, Arnold JG, Van Liew MW, Bingner RL, Harmel RD, Veith TL (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans ASABE* 50(3):885–900
- Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part 1- a discussion of principles. *J Hydrol* 10:282–290. doi:[10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)
- Nemati S, Fazelifard MH, Terzi O, Ghorbani MA (2015) Estimation of dissolved oxygen using data-driven techniques in the Tai Po River, Hong Kong. *Environ Earth Sci* 74:4065–4073. doi:[10.1007/s12665-015-4450-3](https://doi.org/10.1007/s12665-015-4450-3)
- Nürnberg GK (2004) Quantified hypoxia and anoxia in lakes and reservoirs. *Sci World J* 4:42–54. doi:[10.1100/tsw.2004.5](https://doi.org/10.1100/tsw.2004.5)
- Pisinaras V, Petalas C, Gikas GD, Gemtzi A, Tsihrintzis VA (2010) Hydrological and water quality modeling in a medium-sized basin using the Soil and Water Assessment Tool (SWAT). *Desalination* 250:274–286. doi:[10.1016/j.desal.2009.09.044](https://doi.org/10.1016/j.desal.2009.09.044)
- Pouzols FM, Lendasse A (2010a) Evolving fuzzy optimally pruned extreme learning machine: a comparative analysis. *IEEE International Conference on Fuzzy Systems (FUZZ)*, pp.1–8. doi:[10.1109/FUZZY.2010.5584327](https://doi.org/10.1109/FUZZY.2010.5584327).
- Pouzols FM, Lendasse A (2010b) Evolving fuzzy optimally pruned extreme learning machine for regression problems. *Evol Syst* 1:43–58. doi:[10.1007/s12530-010-9005-y](https://doi.org/10.1007/s12530-010-9005-y)
- Ranković V, Radulović J, Radojević I, Ostojić A, Ćomić L (2010) Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. *Ecol Model* 221:1239–1244. doi:[10.1016/j.ecolmodel.2009.12.023](https://doi.org/10.1016/j.ecolmodel.2009.12.023)
- Rayer S (2007) Population forecast accuracy: does the choice of summary measure of error matter? *Popul Res Policy Rev* 26:163–184. doi:[10.1007/s11113-007-9030-0](https://doi.org/10.1007/s11113-007-9030-0)

- Rumelhart DE, Hinton GE, Williams RJ (1986) Learning internal representations by error propagation. In: Rumelhart DE, McClelland PDP, Research Group, editors. *Parallel distributed processing: explorations in the microstructure of cognition. Foundations, Vol. I.* Cambridge, MA: MIT Press; pp. 318–362.
- Santisukkasaem U, Olawuyi F, Oye P, Das DB (2015) Artificial neural network (ANN) for evaluating permeability decline in permeable reactive barrier (PRB). *Environ Process* 2:291–307. doi:10.1007/s40710-015-0076-4
- Similä T, Tikka J (2005) Multiresponse sparse regression with application to multidimensional scaling. In: *Artificial neural networks: formal models and their applications-ICANN 2005, Vol. 3697/2005*, pp. 97–102. doi:10.1007/11550907_16.
- Singh J, Knapp HV, Demissie M (2004) Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. ISWS CR 2004–08. www.isws.illinois.edu.
- Sorjamaa A, Miche Y, Weiss R, Lendasse A (2008) Long-term prediction of time series using NNE-based projection and OP-ELM. In: *Proceedings of the IEEE international joint conference on neural networks (IJCNN)*. Hong Kong, China, pp. 2674–2680. doi:10.1109/IJCNN.2008.4634173
- Sovilj D, Sorjamaa A, Yu Q, Miche Y, Séverin E (2010) OPELM and OPKNN in long-term prediction of time series using projected input data. *Neurocomputing* 73:1976–1986. doi:10.1016/j.neucom.2009.11.033
- Sullivan AB, Rounds SA, Deas ML, Sogutlugil IE (2012) Dissolved oxygen analysis, TMDL model comparison, and particulate matter shunting-preliminary results from three model scenarios for the Klamath River upstream of keno dam, Oregon: U.S. Geological Survey Open-File Report 2012–1101, 30 p. <http://pubs.usgs.gov/of/2012/1101/>.
- Sullivan AB, Rounds SA, Asbill-Case JR, Deas ML (2013a) Macrophyte and pH buffering updates to the Klamath River water-quality model upstream of Keno dam, Oregon: U.S. Geological Survey Scientific Investigations Report 2013–5016, 52 p. <http://pubs.usgs.gov/sir/2013/5016/>
- Sullivan AB, Sogutlugil IE, Rounds SA, Deas ML (2013b) Modeling the water-quality effects of changes to the Klamath River upstream of Keno dam, Oregon: U.S. Geological Survey Scientific Investigations Report 2013–5135, 60 p. <http://pubs.usgs.gov/sir/2013/5135>.
- Tayman J, Swanson DA (1999) On the validity of MAPE as a measure of population forecast accuracy. *Population Research and Policy Review* 18(4):299–322. doi:10.1023/A:1006166418051.
- U.S. Geological Survey (2008) National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, Book 9, Chaps. A1–A9 variously dated. Chapter A6, 6–2 dissolved oxygen, p 48. http://water.usgs.gov/owq/FieldManual/Chapter6/6.2_contents.html.
- Wang Y, Zheng T, Zhao Y, Jiang J, Wan YG, Guo L, Wang P (2013) Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. *Environ Sci Pollut Res* 20:8909–8923. doi:10.1007/s11356-013-1874-8
- Willmott CJ (1982) Some comments on the evaluation of model performance. *Bull of Am Meteorol Soc* 63:1309–1313. doi:10.1175/1520-0477(1982)063
- Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim Res* 30(1):79–82. doi:10.3354/cr030079
- Willmott CJ, Ackleson SG, Davis RE, Feddema JJ, Klink KM, Legates DR, O'Donnell J, Rowe CM (1985) Statistics for the evaluation and comparison of models. *J Geophys Res* 90:8995–9005. doi:10.1029/JC090iC05p08995
- Yang Y, Wang Y, Yuan X (2012) Bidirectional extreme learning machine for regression problem and its learning effectiveness. *IEEE Transactions on Neural Networks and Learning Systems* 23:1498–1505. doi:10.1109/TNNLS.2012.2202289
- Zeng Z, Jiang YL, Liu Y, Liu W (2013) Efficient Data Representation Combining with ELM and GNMF, pp.13–23. In Sun F, Toh KA, Romay MG, Mao K (eds.), *Extreme Learning Machines 2013: Algorithms and Applications, Adaptation, Learning, and Optimization* 16. doi:10.1007/978-3-319-04741-6_2.