CrossMark

SHORT COMMUNICATION

New modelling strategy based on radial basis function neural network (RBFNN) for predicting dissolved oxygen concentration using the components of the Gregorian calendar as inputs: case study of Clackamas River, Oregon, USA

Salim Heddam[1] (ORCID)

Abstract The purpose of the present study is to investigate the possibilities of applying radial basis function neural network (RBFNN) as a new modelling approach for predicting dissolved oxygen concentration (DO) without water quality variables a input, and based on the components of the Gregorian calendar: (1) the year in the usual Gregorian calendar (YY), (2) the month of the year (MM), (3) the day of the month (DD), and (4) the number of complete hours that have passed since midnight (HH). Results obtained are compared with those of multiple linear regression (MLR). Results have shown good agreement between the predicted and measured values of DO, with a correlation coefficient of 0.97 in the testing phase using the RBFNN model, and R equal to 0.654 using the MLR model.

Keywords Modelling · Dissolved oxygen concentration · Artificial neural network · Multiple linear regression · Gregorian calendar

Introduction

Monitoring and control of water quality is extremely important because of its implications for the environment. In-situ measurements of water quality parameters have received a great importance worldwide and various programs have been introduced emphasizing the measure and identification of the important water quality parameters. Hence, direct in situ measurements are a useful tool for monitoring water quality in rivers, lakes and streams ecosystems. Nowadays, new methods are to be very welcome, and help to estimate water quality parameters in the absence of any measurement. One of the most important water qualities is certainly dissolved oxygen concentration (DO). DO is used as water quality indices, has been integrated as a part integral of some water quality models, and it is an important indicator of water pollution. Regarding the importance of DO, an extensive effort is being made to develop robust models that can help to estimates DO. Methods such as artificial neural network, fuzzy logic and neurofuzzy, evolutionary models, wavelet decomposition models, and Extreme learning machines, have been employed and specifically developed for this subject (Heddam 2014a, b, c; 2016a, b, c). In the all developed models, DO is linked to water quality variables used as inputs. Furthermore, it would be very interesting to investigate if the use of new modelling strategy leads to more accurate models or not. In the present investigation we present a new method for predicting DO without water quality variables. We used only the components of the Gregorian calendar which are: (1) the year in the usual Gregorian calendar (YY), (2) the month of the year (MM), (3) the day of the month (DD), and (4) the number of complete hours that have passed since midnight (HH).

Materials and methods

Study area and data set

In the present study we selected DO data from two stations: (USGS ID 14210000) at Clackamas River at Estacada, Oregon, USA, (Latitude 45°18′00″, Longitude 122°21′10″ NAD27), and (USGS ID 14211010) at Clackamas River

✉ Salim Heddam
heddamsalim@yahoo.fr

[1] Hydraulics Division, Agronomy Department, Faculty of Science, University 20 Août 1955, Route EL HADAIK, BP 26, Skikda, Algeria

near Oregon City, USA, (Latitude 45°22′46″, Longitude 122°34′34″ NAD27). Figure 1 shows the locations of the two stations in study area. For the two stations the data set is divided into three sub-data sets: (1) a training set (60 %), (2) a validation set (20 %) and (3) a test set (20 %). The historical DO data are available at the United States Geological Survey (USGS) website, http://or.water.usgs.gov/cgi-bin/grapher/table_setup.pl?site_id. DO were collected at every thirty (30) min (1/2 h) intervals of time. The dataset had a total of 5848 records for the USGS ID 14210000 station, and 5882 records for the USGS ID 14211010 station. These data cover all of the four seasons, and 1 month for each season was selected: January for the winter, April for the spring, July for the summer, and October for the autumn. At the present DO corresponds to the output and the four following component of the Gregorian calendar are identified as the inputs of the developed models: (1) the year in the usual Gregorian calendar (YY) and equal to 2015, (2) the month of the year between 01 (January) and 12 (December) (MM), (3) the day of the month between 01 and 31(DD), and (4) the number of

Table 1 Statistical parameters of data set for the two stations

Station	Unit	X_{mean}	X_{max}	X_{min}	S_x	C_v	N
14211010	mg/L	10.76	13.40	6.90	1.53	0.14	5882
14210000	mg/L	10.69	13.80	7.90	1.60	0.15	5848

complete hours that have passed since midnight, between 00 and 23 (HH). A summary of descriptive statistics of the data set for the selected stations is shown in Table 1, where X_{mean}, X_{max}, X_{min}, S_x and C_v denote the mean, the maximum, the minimum, the standard deviation, and the coefficient of variation, respectively.

Radial basis function neural network (RBFNN)

Radial basis function neural network has a feed forward architecture with only three layers: an input layer, a hidden layer, and an output layer as shown in Fig. 2. In the hidden layer, each neuron implements a radial basis function. The RBFNN uses a linear transfer function for the output

Fig. 1 Map showing the location of the two stations in Clackamas River, Oregon, USA, with USGS station identification number. (Adopted from [Lee 2011])
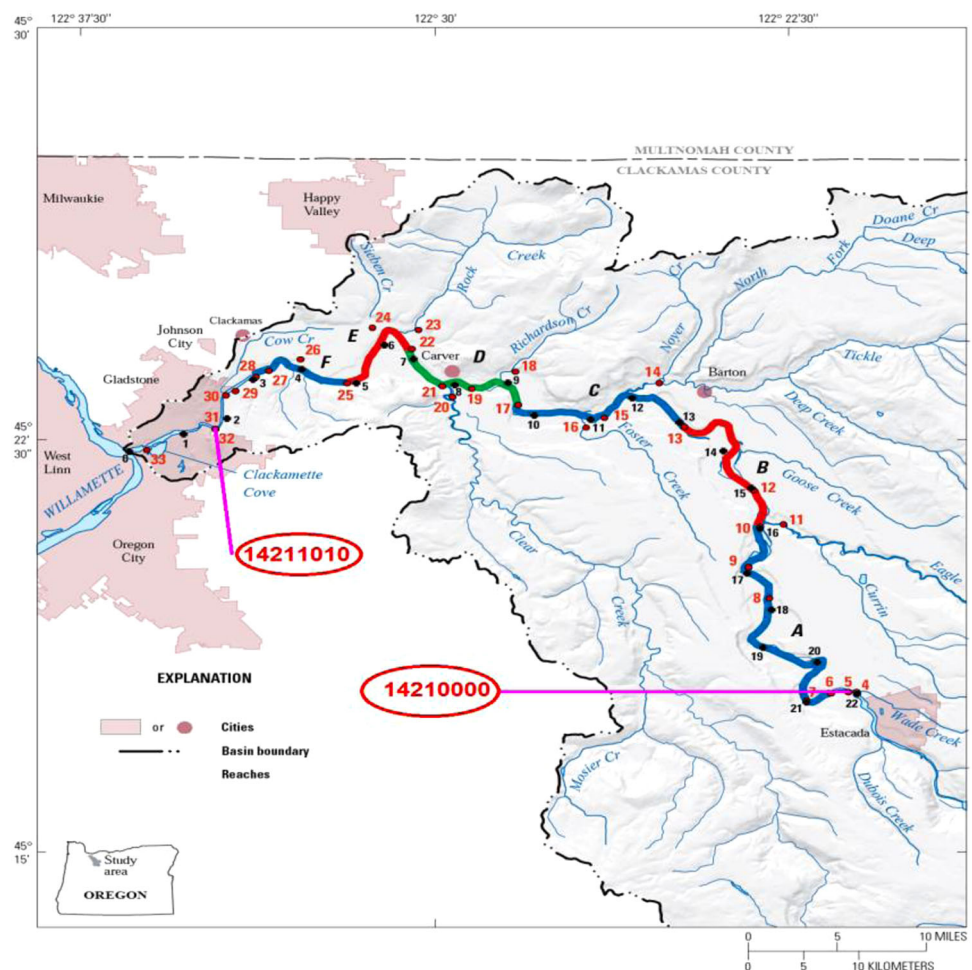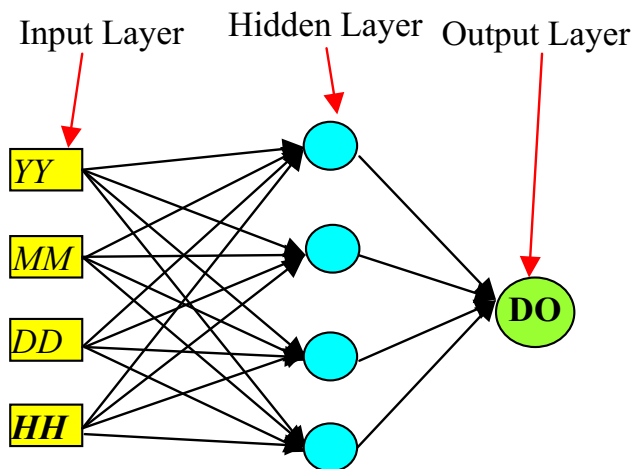
Fig. 2 Architecture of radial basis function neural network (RBFNN).*YY* year, *MM* month of year (1–12), *DD* day of month (1–31), *HH* hours of day (00–23)

neuron. To the mathematical point of view, the RBFNN structure shown in Fig. 2 can be presented as follow:

The RBFNN Gaussian function can be written as:

$$\varphi_i(x) = \exp\left(-\frac{\|x - \mu_i\|^2}{2\,\sigma_i^2}\right) \quad i = 1, 2, N \qquad (1)$$

where σ_i is the widths (or spread) of the hidden neuron.

The output of the RBFNN model can be calculated as follow

$$\Upsilon_i = \sum_{j=1}^{N} w_{ij}\, \varphi_j(x) + B_2 \qquad (2)$$

w_{ij} represents a weighted connections between the radial basis function neuron and output neuron; and N = number of hidden-layer neurons. The constant term B_2 represents a bias. Over the last few years, researchers successfully used RBFNN for many areas of science research (Pal et al. 2016; Bhunia et al. 2016; Parsaie 2016; Ehteshami et al. 2016; Barzegar and Moghaddam 2016; Heddam 2016a).

Multiple linear regression (MLR)

Multiple linear regression (MLR) models are used to examine the relationship between an ensemble of inputs variables and an output variable, using the flowing equation:

$$Y = \varphi(x_i) = \psi_0 + \psi_1 x_1 + \psi_2 x_2 + \psi_3 x_3 + \psi_4 x_4 + \psi_i x_i \qquad (3)$$

where Y is the dependent variables (DO), x_i are the independent variables (YY, MM, DD, and HH), and Ψ_i are the parameters of the models.

Model performance indices

For the evaluation of the model we use three indices: the coefficient of correlation (CC), the root mean squared error (RMSE) and the mean absolute error (MAE).

$$R = \left[\frac{\frac{1}{N}\sum (O_i - O_m)(P_i - P_m)}{\sqrt{\frac{1}{N}\sum_{i=1}^{n}(O_i - O_m)^2}\sqrt{\frac{1}{N}\sum_{i=1}^{n}(P_i - P_m)^2}}\right] \qquad (4)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \qquad (5)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|O_i - P_i| \qquad (6)$$

where N is the number of data points, O_i is the measured value and P_i is the corresponding model prediction. O_m and P_m are the average values of O_i and P_i.

Results and discussion

In this paper, two approaches were compared for modelling DO concentration without any water quality variables. The RBFNN and MLR models were implemented by using two different program codes written in MATLAB language. A comparison of the performance of the RBFNN model with that of the MLR model was carried out to study their efficacy in modelling DO concentration. The performances of two developed models, for the two stations are reported in Table 2. According to Table 2, the results using the RBFNN models for the two stations are very good with high level of accuracy in the all three phases. For the USGS 14211010 station, the calculated DO were correlated with the measured values with an R equal to 0.975, 0.973, and 0.974, in the training, validation and test phase respectively. The corresponding RMSE and MAE were (1.186 and 0.930 mg/L), (1.137 and 0.888 mg/L), and (1.184 and 0.934 mg/L), in the training, validation and testing phases, respectively. The results also show that the RBFNN is more accurate during all the three phase when compared to the MLR model. Model comparisons were made to see which model gave better results. The worst results were obtained in MLR models, with an R coefficient equal to 0.656 in the test phase. Figures 3 and 4 show scatter plots of the calculated against the corresponding measured DO for the RBFNN and MLR models, respectively, in the (a) training, (b) validation, (a) test, and (d) all data, for the USGS 14211010 station. According to Table 2, for the USGS 14210000 station, the results using RBFNN are always good with an R equal to 0.973 in the test phase, an RMSE equal to 0.377, and an MAE equal to 0.306. The

Table 2 Performances of the RBFNN and MLR models in different phases

Station	Models	Training			Validation			Testing		
		CC	RMSE	MAE	CC	RMSE	MAE	CC	RMSE	MAE
14211010	RBFNN	0.975	0.338	0.264	0.973	0.350	0.273	0.974	0.359	0.279
	MLR	0.629	1.186	0.930	0.658	1.137	0.888	0.656	1.184	0.934
14210000	RBFNN	0.975	0.354	0.283	0.970	0.393	0.316	0.973	0.377	0.306
	MLR	0.754	1.054	0.857	0.764	1.031	0.834	0.762	1.050	0.852
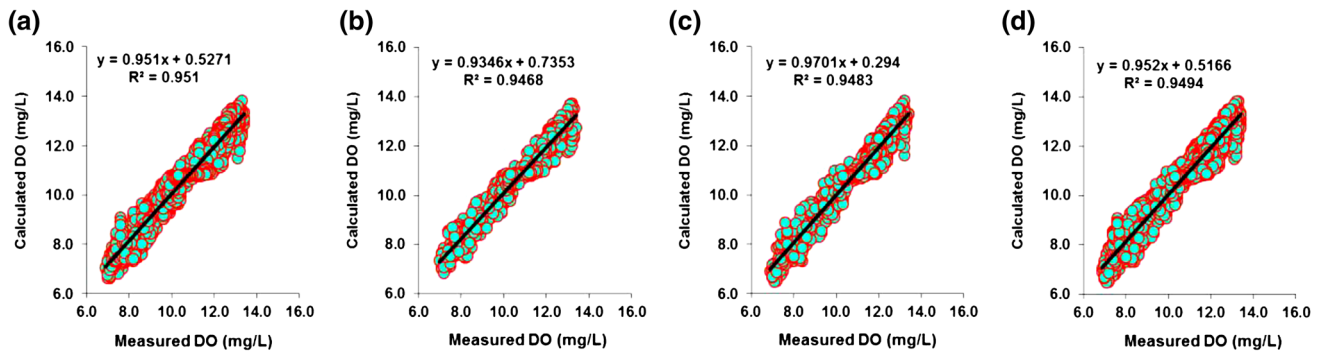
Fig. 3 Scatterplots of calculated versus measured values of dissolved oxygen (DO) for the USGS 14211010 station using RBFNN model: **a** training, **b** validation, **c** testing, and **d** all data
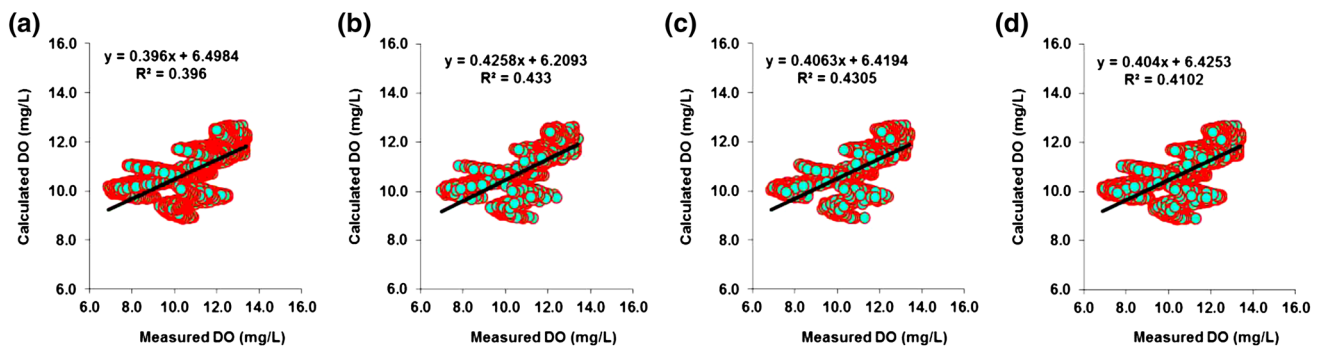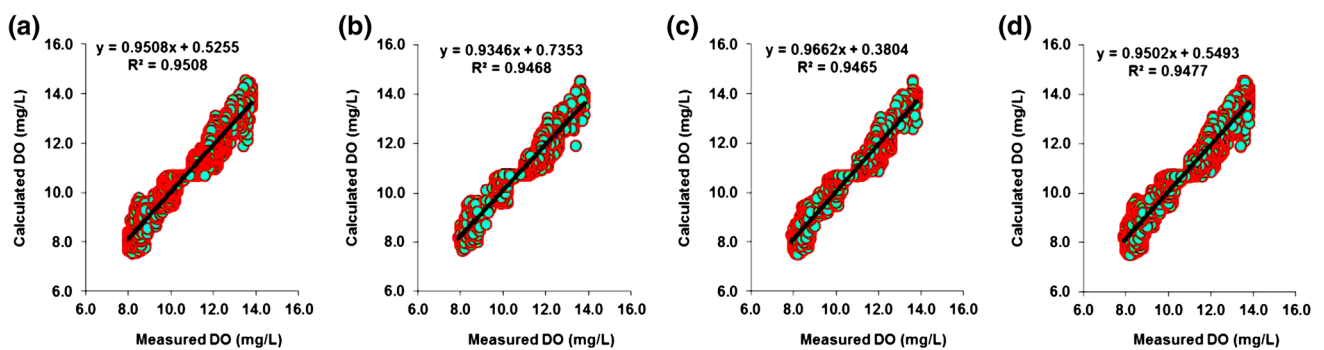

Fig. 4 Scatterplots of calculated versus measured values of dissolved oxygen (DO) for the USGS 14211010 station using MLR model: **a** training, **b** validation, **c** testing, and **d** all data

Fig. 5 Scatterplots of calculated versus measured values of dissolved oxygen (DO) for the USGS 14210000 station using RBFNN model: **a** training, **b** validation, **c** testing, and **d** all data
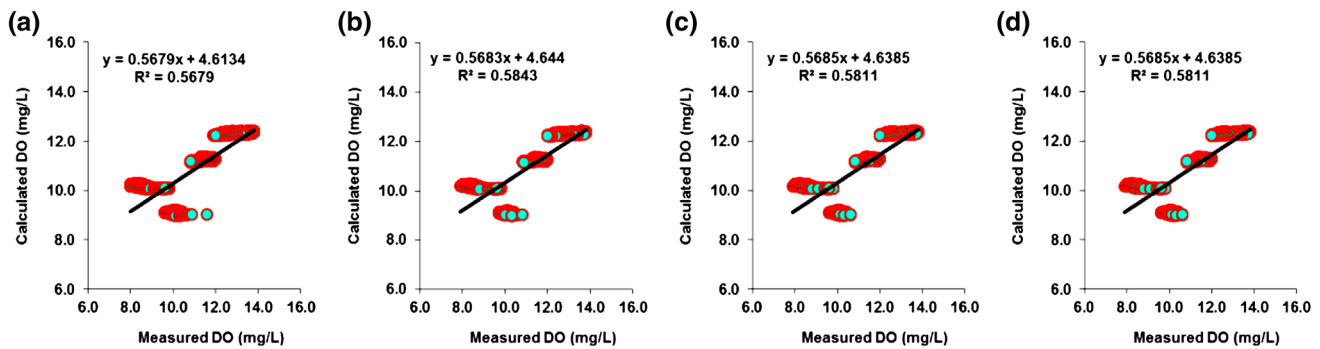
Fig. 6 Scatterplots of calculated versus measured values of dissolved oxygen (DO) for the USGS 14210000 station using MLR model: **a** training, **b** validation, **c** testing, and **d** all data

worst results were obtained in MLR models, with an R coefficient equal to 0.762 in the test phase. Figures 5 and 6 show scatter plots of the calculated against the corresponding measured DO for the RBFNN and MLR models, respectively, in the (a) training, (b) validation, (a) test, and (d) all data, for the USGS 14210000 station.

Conclusions

To deal to the lack of the available water quality variables generally used as input to the models developed for predicting DO concentration in river ecosystems, this study has attempted to proposes a new kind of model that can be used to predict hourly dissolved oxygen without the need to any water quality variables. Using only the four components of the Gregorian calendar, we have obtained very promising results. In the future, we need to elevate the promising results being achieved.

References

Barzegar R, Moghaddam AA (2016) Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. Model Earth Syst Environ 2:26. doi:10.1007/s40808-015-0072-8

Bhunia GS, Shit PK, Maiti R (2016) Spatial variability of soil organic carbon under different land use using radial basis function (RBF). Model Earth Syst Environ 2:17. doi:10.1007/s40808-015-0070-x

Ehteshami M, Farahani ND, Tavassoli S (2016) Simulation of nitrate contamination in groundwater using artificial neural networks. Model Earth Syst Environ 2:28. doi:10.1007/s40808-016-0080-3

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) Simultaneous modelling and forecasting of hourly dissolved oxygen concentration (DO) using radial basis function neural network (RBFNN) based approach: a case study from the Klamath River, Oregon, USA. Model Earth Syst Environ 2:135. doi:10.1007/s40808-016-0197-4

Heddam S (2016b) Fuzzy Neural Network (EFuNN) for modelling dissolved oxygen concentration (DO). In: Kahraman C, Sari IU (eds) Intelligence systems in environmental management: theory and applications, intelligent systems reference library 113, ("accepted"/"in press"). doi: 10.1007/978-3-319-42993-9_11

Heddam S (2016c) 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. Environ Process. doi:10.1007/s40710-016-0172-0

Lee KK (2011) Seepage investigations of the Clackamas River, Oregon: U.S. Geological Survey Scientific Investigations Report 2011–5191, 16 p. http://pubs.usgs.gov/sir/2011/5191/. Accessed 27 Jul 2016

Pal S, Manna S, Chattopadhyay B, Mukhopadhyay SK (2016) Carbon sequestration and its relation with some soil properties of East Kolkata Wetlands (a Ramsar Site): a spatio-temporal study using radial basis functions. Model Earth Syst Environ 2:80. doi:10.1007/s40808-016-0136-4

Parsaie A (2016) Predictive modeling the side weir discharge coefficient using neural network. Model Earth Syst Environ 2:63. doi:10.1007/s40808-016-0123-9