Global J. Environ. Sci. Manage.,4(4): 439-450, Autumn 2018
DOI: 10.22034/gjesm.2018.04.005

ORIGINAL RESEARCH PAPER

Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river

G. Elkiran[1,], V. Nourani[2], S.I. Abba[1], J. Abdullahi[1]*

[1]*Faculty of Civil and Environmental Engineering, Near East University, Near East Boulevard 99138, Nicosia, North Cyprus*

[2]*Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran*

Received 26 March 2018; revised 10 June 2018; accepted 12 August 2018; available online 1 October 2018

ABSTRACT: In this study, adaptive neuro-fuzzy inference system, and feed forward neural network as two artificial intelligence-based models along with conventional multiple linear regression model were used to predict the multi-station modelling of dissolve oxygen concentration at the downstream of Mathura City in India. The data used are dissolved oxygen, pH, biological oxygen demand and water temperature at upper, middle and downstream of the river. To predict outlet of dissolved oxygen of the river in each station, considering different input combinations as i) 11 inputs parameters for all three locations except, dissolved oxygenat the downstream ii) 7 inputs for middle and downstream except dissolved oxygen, at the target location and lastly iii) 3 inputs for downstream location. To determine the accuracy of the model, root mean square error and determination coefficient were employed. The simulated results of dissolved oxygen at three stations indicated that, multi-linear regression is found not to be efficient for predicting dissolved oxygen. In addition, both artificial intelligence models were found to be more capable and satisfactory for the prediction. Adaptive neuro fuzzy inference system model demonstrated high prediction ability as compared to feed forward neural network model. The results indicated that adaptive neuro fuzzy inference system model has a slight increment in performance than feed forward neural network model in validation step. Adaptive neuro fuzzy inference system proved high improvement in efficiency performance over multi-linear regression modeling up to 18% in calibration phase and 27% in validation phase for the best models.

KEYWORDS: *Adaptive neuro fuzzy inference system* (*ANFIS); Feed forward neural network (FFNN); Multi-linear regression (MLR); Dissolve oxygen (DO); Water quality; Yamuna river.*

INTRODUCTION

The issue of water quality (WQ) management has an essential role to play in respect to river basin planning and control of water pollution. The likeliness of industrial and municipal waste discharge into river is of immense concern particularly to those utilizing water diverted from rivers. Dissolve oxygen (DO) is the dissolved form of the amount of oxygen; it is among the best variables that indicate the health status and quality of the ecosystem (Sharma and Kansal, 2011). Ensuring DO concentration within the acceptance range which differ according to national and international standards is important. However, it can range from 0 up to 18 parts per million. With less DO, the lives of aquatic animals in the receiving environment are likely to be lose (Jain *et al.,* 2014). DO served as a major indicator of the river WQ and has been given several considerations in the literatures recently (Sharma and Kansal, 2011). At initial stage,

*Corresponding Author Email: *gozen.elkiran@neu.edu.tr*
Tel.: +90 548 8560681 Fax: +90 392 2236461

Note: Discussion period for this manuscript open until January 1, 2019 on GJESM website at the “Show Article”.

rivers are free from any form of impurities and are regarded as the cleanest in the entire globe but rapid increase in industries, urban and human development causes imposing pollutants into water bodies. For the maintenance of sustainable development, assessment of WQ is of immense significance (Jain *et al.,* 2014). Due to crucial role of WQ parameters in hydro-environmental studies, there is a need to develop reliable prediction methods for these parameters. In several studies, linear models have been applied to determine different variables of WQ by examining WQ characteristics (Mirbagheri *et al.,* 2010; Karbassi and Pazoki, 2015). However, complex and dynamic behaviours of system define the inability of the linear methods to withstand the interactions and processes in stream water body that is taking place. On the other hand, non-linear Artificial Intelligence (AI) models are crucial and play and essential role in simulation of complex and non-linear processes. As such AI based models for example, artificial neural network (ANN), Adaptive Neuro Fuzzy Inference System (ANFIS), could lead to accurate and reliable results in modeling and estimating the trend of non-linear hydrological processes (Quej *et al.,* 2017). Feed Forward Neural Network (FFNN), and ANFIS have their own advantages and limitations. Despite, ANN plays a crucial part of ANFIS, it can learn from the data but the results are not easily interpretable. On the other hand, ANFIS has the potential to derive the advantages of both fuzzy inference system (FIS) and ANN which could reduce their weaknesses, but it is more complex and only supports the Takagi-Sugeno-Kang (TSK) inference (Mamdani 1974). Recent works indicate that, the applications of the AI based models have been successful in modelling and prediction of WQ parameters. ASCE Task Committee, 2000 discussed in detail the FFNN-based modelling of hydrological phenomena. Nevertheless, the determination and prediction of DO have been carried out by some researchers. Areerachakul *et al.,* 2011 employed the application of ANN to estimate DO of a river the obtained result indicates that ANN provide high accuracy than experimental model. Feed forward neural network (FFNN) utilized using back propagation algorithm was applied by Singh *et al.,* 2009 to estimates biological oxygen demand (BOD) and DO level in the Gomti River of India, the results justified the robustness of ANN in the prediction of WQ parameters. Rankovic´ *et al.,* 2010 tested FFNN capabilities in predicting DO by applying many water quality variables; as revealed by the obtained results, pH and water temperature are the most effective variables in predicting DO. Ay and Kisi, 2011 applied and examined the performance accuracy of RBNN and MLP algorithms in modeling the DO concentration in Foundation Creek, Colorado, the obtained result indicates high merit of RBNN and also demonstrates better performance of two ANN than the linear multi-linear regression (MLR) model. Olyaie *et al.,* 2017 presented various computational intelligence techniques e.g., MLP, radial basic function (RBF), linear genetic programming (LGP) and support vector machine (SVM), to estimate DO Concentration. Highly encouraging results were obtained from the study which suggested that in modelling DO, SVM and ANNs approaches are promising. Wen *et al.,* 2013 examined the performance accuracy of the ANNs for DO modelling in the Heihe River. The proposed ANN model with limited information of WQ parameters being preferable choice for DO levels modelling. The literature survey shows that the multi-station prediction of DO has not been gotten proposal to investigate the impact of different segments of the river on the outputs. Such multi-station modelling is an important task in hydrological modelling (eg. in river flow modelling) (Turan and Yurdesev, 2009). There is a room to be applied for modelling WQ parameters of the rivers (e.g. DO as well). In this study, AI based multi-station WQ modelling is proposed and applied to the Yamuna River in India. For this propose, DO, biological oxygen demand (BOD), pH, and water temperature (WT) from 3 different stations of the river with different input combinations are imposed into FFNN and ANFIS models. In addition, the classical linear MLR are employ for comparison. The crux of selecting pH, DO, BOD and WT parameters to simulate the WQ based on the DO concentration is that, these variables are usually utilized to categorize the rivers according to the usage. pH is the concentration of hydrogen ions that indicates the acidic and basic level of solution. DO is a good indicator which comprises of temperature, volume and velocity of flowing water that allows water bodies to withstand aquatic life. WT is an essential factors for the survival of aquatic organism in which their lives is dependent on the vicinity temperature (Verma and Singh, 2013). BOD is the amount of oxygen required to oxidize organic matter presents in a sample by the actions of microorganisms (Jain *et al.,* 2014; Sharma and Kansal 2011). The aim of this study is to apply AI based and regression models to simulate dissolved oxygen at downstream of three Mathura sample stations in India during 1999-2012.

MATERIALS AND METHODS

Study region and data

The biggest tributary of River Ganga is Yamuna River, this river is as sacred and prominent as the immense River Ganga itself. It was being widely acknowledge as a holy river in various pilgrimage centres across India and in Indian mythology e.g. Allahabad, Baleshwar, Vrindavan and Mathura (all in Uttar Pradesh), Paonta Sahib (Himachal Pradesh), Yamunotri (Uttaranchal) are situated at the banks of this river. Covering 1,376 km, almost 57 million residents of North part of India rely upon it. A total catchment area of Yamuna is 366,223 km^2 which comprises of 42 percent of the Ganga basin area in the Indian Territory. Delhi as capital territory received almost 70 percent of its drinking water from Yamuna River while discharges almost 10,000 m^3/s yearly. But due to urbanization and inadequate water treatment plant, the River leaves Delhi as polluted water (Singh *et al.,* 2005). The upstream was monitored to evaluate the WQ of Yamuna before it enters Vrindavan – Mathura, while, the downstream site shows the effect of wastewater discharges from Mathura. The standard assessment and monitoring of the river administered by Central Pollution Control Board (CPCB) under the National WQ Monitoring Program (NWQMP) and National River Conservation Program (NRCR). Fig. 1 shows the location of the Yamuna River basin in India and the stations. The daily WQ data were obtained from the CPCB for years 1999 to 2012. In order to remove the noise from the obtained raw data, data processing was initially carried out for all the 12 parameters namely, DO, BOD, pH and WT for all 3 stations using the method of regression analysis. Table 1 shows the descriptive statistical analysis of each parameter.

Proposed methodology

In this study, FFNN, ANFIS and MLR models were proposed for multi-station modelling of DO in a river, data set were partitioned into two parts, 75% of the data were employed for the calibration phase and the 25% of the data for validation purposes from a total of 168 records. Selection of dominant inputs parameters is one of the important parts in any AI based modeling. As such, a key and preferable parameter combination for FFNN, ANFIS and MLR are selected using sensitivity and correlation analysis (Table 2). For development of the models, 11, 7, and 3 input combinations were considered for modelling in the upper, middle and downstream of the river, respectively. The mathematical expressions of the multi-station modelling are presented in Eqs. 1-5. MATLAB 9.3 (R2017b) was used for the analysis of FFNN and ANFIS while MLR model and correlation coefficient were developed using regression tool of

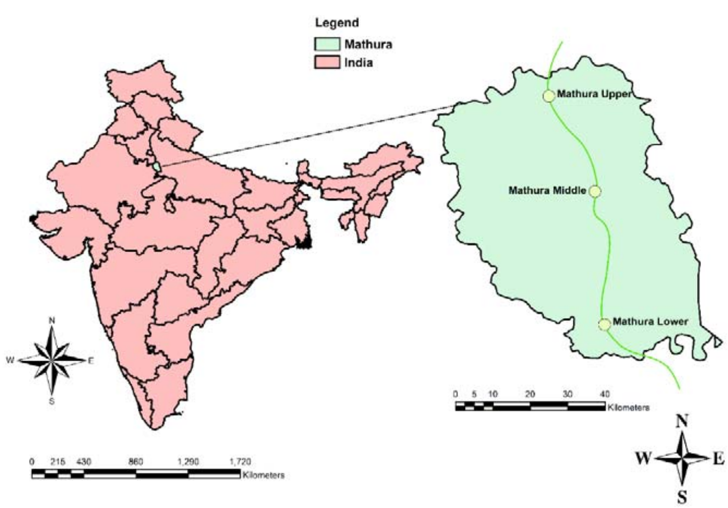

Fig. 1: Geographic location of the study area of Mathura along Yamuna River

EViews software 9.5 version.

$$DO_u = f_u(BOD_u, pH_u, WT_u) \tag{1}$$

$$DO_m = f_m(DO_u, BOD_u, BOD_m, pH_u, pH_m, WT_u, WT_m) \tag{2}$$

$$DO_d = f_d(DO_u, DO_m, BOD_u, BOD_m, BOD_d, pH_u, pH_m, pH_d, WT_u, WT_m, WT_d) \tag{3}$$

$$DO_d = f_d(DO_u, DO_m, BOD_u, BOD_m, pH_u, pH_m, WT_u, WT_m) \tag{4}$$

$$DO_d = f_d(DO_u, BOD_u, pH_u, WT_u) \tag{5}$$

The DO at upper and middle stream were modelled for comparison to DO at downstream, based on the functions of the parameters as given in Eqs. 1 and 2.

Where, DO_u, DO_m, DO_d are dissolve oxygen at upper, middle and downstream, f_u, f_m, f_d donate the functions of upper, middle and downstream, BOD_u, BOD_m, BOD_d are biological oxygen, WT_u, WT_m, WT_d are water temperature and pH_u, pH_u, pH_u are pH at upper, middle and downstream respectively.

From the Table 1, the range of pH of the upper stream is higher than both the middle and downstream which could be due to the discharge of industrial pollution located at the upper stream. The average WT is almost equal in all the 3 stations, also the minimum DO and BOD at the upper stream is higher than that of middle and downstream.

Multi-linear regression (MLR)

The variables input and output correlation level are generally estimated by regression models which determine their relationships form. The correlation coefficient ranges between -1 and +1 and quantifies the direction and strength of linear relationship between two variables. The most commonly used regression is MLR, which correlated the value of dependent variable with values of independent variables. Ordinarily, correlation level between two or more predictors (independent variables) and one response variable (dependent variable) are estimated by MLR. MLR model that correlates the given output (Y) to input variables (X) are defined in the Eq. 6. The assumption of Y as dependent and X as independent are considered in MLR (Dogan *et al.,* 2008).

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots b_ix_i \tag{6}$$

Where, x_i, is the value of the i^{th} predictor, b_0 is the regression constant, and b_i is the coefficient of the i^{th} predictor.

Feed forward neural network (FFNN)

An ANN is a model designed synaptic weight and learning process to resembles brain by processing information based on a mathematical model (Hsu *et al.,* 1995; Gaya *et al.,* 2014). FFNN is used to find the non-linear relationship among the residuals of the fitted linear model (Nourani *et al.,* 2011). ANN can be classified in terms of objective function, flow of information and learning method. Among different classes of ANN, FFNN with back propagation (BP) algorithm is the most common and widely used technique. In BP, each input training data flows via

Table 1. Descriptive statistics of each parameter

Station	Parameters	Min.	Max.	Median	$\bar{x}$	Variance	S.D.
Upper Stream (Mathura)	DO (mg/L)	1.0	17.2	6.5	6.92	9.14	3.02
	pH	6.9	9.3	7.8	7.81	0.16	0.41
	BOD (mg/L)	3.0	25.0	8.0	8.63	19.25	4.74
	WT (ºC)	10.0	36.0	28.0	26.06	35.44	4.38
Mid-Stream (Mathura)	DO (mg/L)	0.0	19.2	6.2	6.25	9.06	5.95
	pH	6.7	9.0	7.8	7.89	0.23	0.47
	BOD (mg/L)	2.0	27.0	8.0	9.02	19.78	4.45
	WT (ºC)	11.0	36.5	28.0	26.42	34.35	2.80
Down Stream (Mathura)	DO (mg/L)	0.0	19.6	6.2	6.26	9.05	3.01
	pH	6.7	9.1	7.9	7.89	0.23	0.47
	BOD (mg/L)	2.0	27.0	18.0	9.02	19.17	4.44
	WT (ºC)	11.0	36.5	28.0	26.4	34.35	5.86

*S.D.= Standard deviation; Min. = Minimum; Max.= Maximum; $\bar{x}$= Mean the system and passes to the output layer, the error of the training is generated and propagates backward until the output desired of the network is achieved (Nourani 2017; ASCE Task Committe, 2000). The primary aim of BPNN is to reduce the error in order for the network to learn the training data Fig. 2 shows a three-layer FFNN with BP (Nourani *et al.,* 2012)two artificial neural networks were developed to simulate outflow hydrograph from earthen dam breach. The required data for the modelling were collected from literature, laboratory experiments and a physically based model (i.e. BREACH. Sigmoid was the transfer function utilized in this study which was introduced to convert in each neuron linear to non-linear function ranging gradually between 0 and 1 (Nourani *et al.,* 2015; ASCE Task Committee 2000). However, every layer comprised of interconnected neurons by weight and activation function (Areerachakul *et al.,* 2011; Nourani *et al.,* 2015). Lavenberg-Marquardt (LM) is an algorithm used in training MLP model because of its outstanding performance as described in several hydrology literature (ASCE Task Committe, 2000). Before model training at the initial stage, the data for both input and output were normalized within a scale of 0 and 1 using the Eq. 7 which helps to increase integrity and reduce the redundancy of the data (Abba and Elkiran, 2017).

$$X_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{7}$$

Where "X_i" is the normalized quantity, "x_i" is un-normalized quantity, "x_{min}" is the minimum and "x_{max}" is the maximum quantity of the data set (Nourani *et al.,* 2012)two artificial neural networks were developed to simulate outflow hydrograph from earthen dam breach. The required data for the modelling were collected from literature, laboratory experiments and a physically based model (i.e. BREACH.

Adaptive neuro-fuzzy inference system (ANFIS)

Fuzzy logic as AI based technique deals with uncertainty, vagueness, and imprecision which was first introduced by Zadeh, 1996. A mathematical expression

Table 2: Correlation variables used in multi-station modelling

Parameters	Do_u	pH_u	BOD_u	WT_u	Do_m	pH_m	BOD_m	WT_m	DO_d	pH_d	BOD_d	WT_d
Do_u	1.0000											
pH_u	0.5152	1.0000										
BOD_u	0.0487	0.1021	1.0000									
WT_U	0.0772	0.2394	0.0771	1.0000								
Do_m	0.0526	-0.0613	0.0756	0.1004	1.0000							
pH_m	0.1001	0.1106	0.1596	-0.0231	0.2601	1.0000						
BOD_m	0.0056	0.0656	0.0543	0.1695	-0.1740	-0.0195	1.0000					
WT_m	-0.0117	0.0723	0.0951	0.1226	0.0295	0.2642	0.0137	1.0000				
DO_d	0.1465	0.0900	0.0458	0.0070	0.5768	0.2909	-0.0001	0.0664	1.0000			
pH_d	0.1521	0.2147	0.0965	-0.0036	0.2767	0.5154	0.0371	0.1244	0.3551	1.0000		
BOD_d	0.2503	0.1407	0.1708	0.2393	0.0012	-0.0582	0.4075	-0.1251	-0.1032	0.0843	1.0000	
WT_d	0.0362	0.0166	0.1639	-0.0686	0.0830	0.1343	-0.0772	0.4541	0.0641	0.3013	-0.0260	1.0000

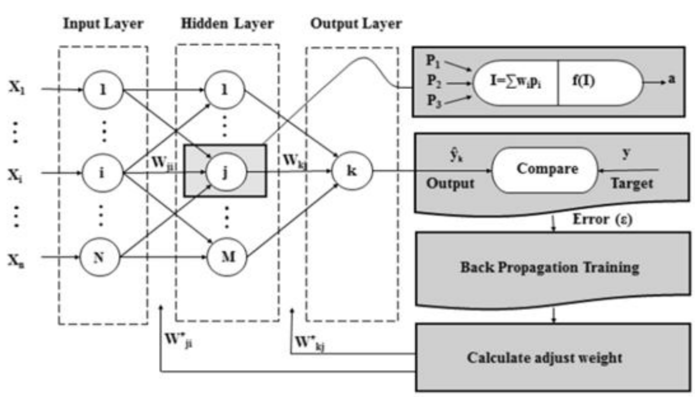

Fig. 2: Typical structure of three-layered FFNN (Nourani *et al.,* 2012)

called fuzzy rule input and output relationship of a system with respect to if -then statement and form of language variables (Yetilmezsoy *et al.,* 2011; Abba *et al.,* 2017). Fuzzifier, defuzzifier and fuzzy database are the main three parts of the system (Nourani *et al.,* 2011). Fuzzy rules and FIS knowledges are important aspects of fuzzy logic (Parmar and Bhardwaj, 2015). The combination of ANN with the fuzzy system creates a robust hybrid system that is able to solve a complex nature of relationship (Akrami *et al.,* 2014). ANFIS is a Multi-Layer Feed-Forward (MLFF) neural network using the integration of neural network and fuzzy logic algorithms in order to map inputs with outputs (Solgi *et al.,* 2017). ANFIS has several drawbacks like other soft computing tools as hybrid learning algorithm, however the approach is more suitable and complex for some inference systems like Takagi-Sugeno. The major importance of ANFIS rule systems basically are categorized into Mamdani and Takagi-Sugeno which are expressed normally into mathematical function and linguistic variable, respectively. Mamdani rule requires defuzzification process where as Sugeno requires no defuzzification (Takagi and Sugeno, 1993). As a universal approximator, ANFIS to any degree of accuracy on a compact set, has the capability of approximating any continuous real function. Functionally, ANFIS is equivalent to FIS (Takagi and Sugeno, 1993). Fig. 3 shows the general structure of the ANFIS.

Supposing a FIS containing two inputs and one output '*x*' '*y*' and '*f*'', a Sugeno fuzzy first order has the following rules (Eqs. 8 and 9).

Rule (1): if $\mu(x)$ is A_1 and $\mu(y)$ is B_1; then $f_1 = p_1x + q_1y + r_1$ (8)

Rule (2): if $\mu(x)$ is A_2 and $\mu(y)$ is B_2; then $f_2 = p_2x + q_2y + r_2$ (9)

Membership functions parameters for x and y inputs are A_1, B_1, A_2, B_2, outlet functions' parameters are p_1, q_1, r_1, p_2, q_2, r_2, a five-layer neural network arrangement followed the formulation and structure of ANFIS

Layer 1: Every node i is an adaptive node in this layer, which has a node function as in Eq. 10.

$$Q_i^1 = \mu_{Ai}(x) \text{ for } i = 1,2 \text{ or } Q_i^1 = \mu_{Bi}(x) \text{ for } i = 3,4 \quad (10)$$

Where ${Q^1}_i$ for input x or y is the membership grade. Gaussian membership function was chosen due to its lowest error in prediction.

Layer 2: T-Norm operator connects every rule in this layer between inputs that performs as 'AND' operator as Eq. 11.

$$Q^2{}_i = w_i = \mu_{Ai}(x).\mu_{Bi}(y) \text{ for } i = 1,2 \quad (11)$$

Layer 3: ''Normalized firing strength'' is the output in this layer and every neuron is labelled Norm as Eq. 12.

$$Q_i^3 = \overline{w}_i = \frac{w_i}{w_1 + w_2} \qquad i = 1, 2 \quad (12)$$

Layer 4: Every node *i* in this layer is an adaptive node

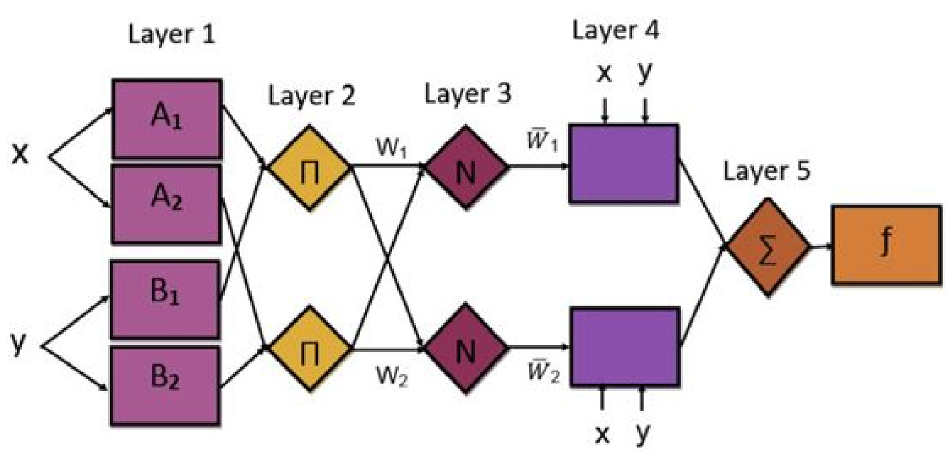

Fig. 3: First order type Sugeno FIS and ANFIS Structure Model

and performs the consequent of the rules as Eq. 13.

$$Q_i^4 = \overline{w}_i(p_i x + q_i y + r_i) = \overline{w}_i f_i \tag{13}$$

p_1, q_1, r_1 are irregular parameters referred to as consequent parameters.

Layer 5: In this layer the overall output is computed as the summation of all incoming signals as Eq. 14.

$$Q_i^5 = \overline{w}_i(p_i x + q_i y + r_i) = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{14}$$

Performance efficiency criteria

The performance efficiency of the model can be assessed through different statistical measures, in order to evaluate the predictive performance of the model. Determination coefficient (DC), Root mean square error (RMSE) and mean square error (MSE) were employed in this study as Eqs. 15, 16 and 17.

$$DC = 1 - \frac{\sum_{i=1}^{n}(o_{obs_i} - o_{com_i})^2}{\sum_{i=1}^{n}(o_{obs_i} - \overline{o}_{obs})^2} \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(o_{obs_i} - o_{com_i})^2}{n}} \tag{16}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_{obsi} - O_{comi})^2 \tag{17}$$

Where n, o_{obs_i}, $\overline{o}_{obs}$ *and* o_{com_i} are data number, observed data, averaged value of the observed data and calculated values, respectively. DC ranges between -∞ and 1, with perfect score of 1.

The RMSE and DC show the measured and computed values differences. Under normal circumstance, the best network is achieved with higher DC and lower RMSE (Legates and McCabe, 1999).

RESULTS AND DISCUSSION

For the multi-station modelling in this study, sensitivity and correlation analyses were examined between the variables. The analysis indicated that the most effective parameter to affect DO is pH,

Table 3: Performance evaluation of MLR model

Model type	Model structure	Calibration			Validation		
		RMSE[a]	MSE[a]	DC	RMSE[a]	MES[a]	DC
MLR I	(11-1)	1.02	1.04	0.78	1.11	1.23	0.69
MLR II	(10-1)	1.12	1.25	0.68	1.13	1.27	0.6
MLR III	(9-1)	1.68	2.82	0.54	1.21	1.46	0.49
MLR IV	(8-1)	1.23	1.51	0.62	1.34	1.79	0.57
MLR V	(7-1)	1.54	2.37	0.64	1.34	1.79	0.6
MLR VI	(6-1)	1.64	2.68	0.58	1.58	2.49	0.54
MLR VII	(5-1)	1.42	2.01	0.48	1.43	2.04	0.38
MLR VIII	(4-1)	1.68	2.82	0.59	1.23	1.51	0.50
MLR IX	(3-1)	1.23	1.51	0.57	1.35	1.82	0.54
MLR X	(2-1)	1.45	2.10	0.59	1.55	2.40	0.50

[a]RMSE has no unit since all the data were normalized
[a] MES has no unit since all the data were normalized

Table 4: Performance evaluation result for FFNN model

Model type	Model structure	Calibration			Validation		
		RMSE[a]	MES[a]	DC	RMSE[a]	MES[a]	DC
FFNN - I	(11 - 11- 1)	0.81	0.65	0.91	1.72	2.95	0.82
FFNN - II	(10- 10- 1)	0.76	0.57	0.89	1.38	1.90	0.83
FFNN - III	(9- 9 - 1)	1.29	1.66	0.86	1.50	2.25	0.77
FFNN - IV	(8 - 8- 1)	0.85	0.72	0.71	1.74	3.02	0.79
FFNN - V	(7- 7 - 1)	0.24	0.05	0.95	0.56	0.31	0.91
FFNN - VI	(6 - 6 - 1)	1.69	2.85	0.72	1.45	2.10	0.69
FFNN - VII	(5- 5- 1)	0.91	0.82	0.81	1.74	3.02	0.79
FFNN - VIII	(4- 4 - 1)	0.79	0.62	0.9	1.57	2.46	0.85
FFNN - IX	(3- 3 - 1)	1.59	2.52	0.78	1.60	2.56	0.76
FFNN - X	(2- 2 - 1)	1.45	2.10	0.67	1.50	2.25	0.65

[a]RMSE has no unit since all the data were normalized
[a] MES has no unit since all the data were normalized

which has the highest correlation in the middle and downstream of Mathura (Table 2). Data from Mathura upstream, Mathura middle and Mathura downstream (except DO) were considered. Ten different models were trained based on the number and types of input combinations, for all the methods the model types were defined as MLRI up to MLRX, FFNNI up to FFNNX and ANFISI up to ANFISX indicating the type of models from one to ten for MLR, FFNN, and ANFIS, respectively. MLR was applied for the estimation of DO of Mathura; the least square approach was used for fitting the model to data. In a model structure (Table 3), 11-1 stand for number of inputs and output variables, the best model was found to be MLR-I with the highest number of input parameters (Table 3). Fig. 4 shows scattered and time series plot for the best model for test data. As it is shown in Table 3, the DC values in MLR- I model were determined as 0.78, 0.69, the MSE were found to be 1.04, 1.23 and the RMSE were 1.02, 1.11 for calibration and validation, respectively. MLR-I model with all the input combinations show high value of DC that implies the accuracy of the model. In addition, MLR approaches desired results by using high number of input variables as shown in Table 3. To conclude, MLR-I model did not show prominent performance capability. This might be due to the nonlinear relationship and interactions within the system, while MLR model is mostly able to find out the linear relationship between the observed and predicted variables. Secondly, FFNN was applied to predict DO at three stations of Mathura, the FFNN models were trained using LM of BP algorithm. Ten different types of models were trained and the best structure of each network was obtained through a trial-error procedure. The model structure 11-11-1 indicates 11 number of input parameters, 11

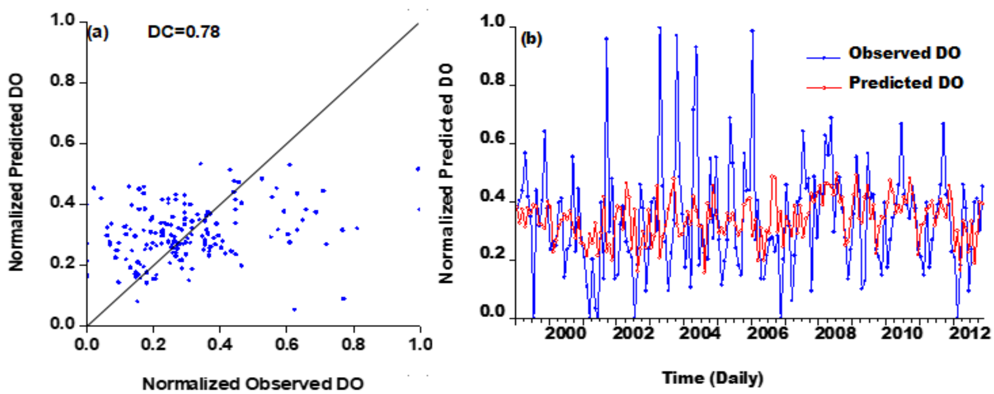

Fig. 4: Evaluation of MLR model a) scatter plot (b) observe versus computed time series of DO

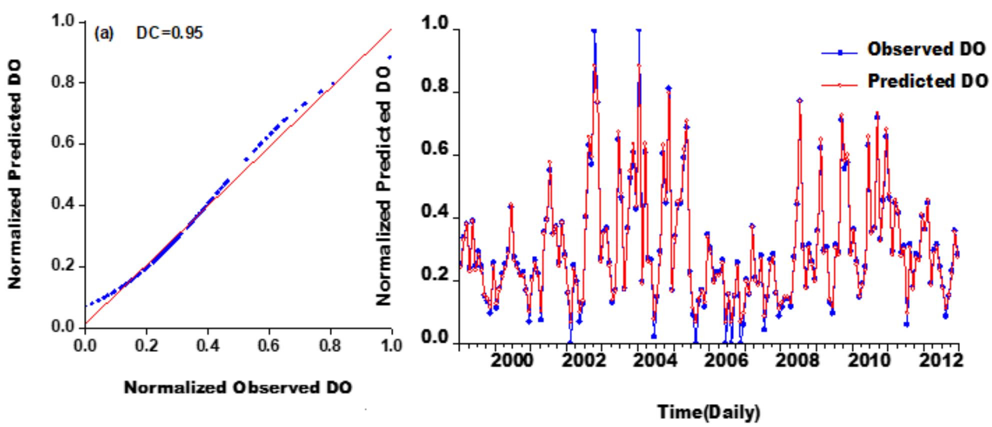

Fig. 5: Evaluation of FFNN model a) scatter plot (b) observe versus computed time series of DO

hidden layer neurons and 1 output layer. However, appropriate architecture selection of the number of neurons in hidden layer, model network, and optimal epoch number are important for calibration of the model to prevent the over-fitting. The obtained results for the best structure are presented in Table 4. FFNN-obtained result showed that high number of input variables could lead to the complex computations in calibration stage, which leads to over fitting of the model. The DC, MSE and RMSE values for calibration of the FFNN-V model were determined as 0.95, 0.91; 0.05, 0.31 and 0.24, 0.56 for calibration and validation, respectively. Fig. 5 depicts the scatter and time series plot of observed and predicted DO for the downstream. FFNN-V model shows that, the combination of seven input variables led to increase in accuracy of the model. This might be due to large input parameters that results to convolution of FFNN, in contrast to MLR and ANFIS models (Tables 3-5). To conclude, the FFNN model is efficient in predicting the DO at downstream of Mathura as compared to MLR. Lastly, ANFIS modelling was performed, and the proportions of calibration, validation were selected same as the ones selected for FFNN modelling. Fig. 6 shows the observed and predicted scatter and time series plot for the best model. Tables 5 and 6 shows the performance criteria of best model inputs structure and the results indicated that, ANFIS-I model with optimum input parameters led to the best outcomes. The DC and RMSE show the level of capability of the ANFIS model in prediction. Table 5 shows that, the DC, MSE and RMSE for ANFIS-I model were determined as 0.96, 0.91; 0.003, 1.58 and 0.018, 1.26, respectively. Thus, ANFIS model is found to be capable for prediction with satisfactory performance, the improvement of ANFIS could be due to it is capability to overcome both the limitation of fuzzy inference system and ANN. This finding corresponds with those of (Najah et al., 2014; Chen and Liu, 2014). In addition, low MSE for validation shows that the accuracy of ANFIS model proved high merit. For the purpose of this research, (11, trimf, 2) indicates that, a model with 11 input variables, trimf stands for triangular membership function and 2 as a number of membership function. To investigate the effect of different segment of river on DO, the DO of the others two stations was also predicted. Table 5 depicts the poor performance result of both DO_u and DO_m which is as a result of sewage treatment plant (STPs) located at upper stream of Mathura the discharge of these STPs have huge impact on middle stream and downstream WQ. However, the value of pH at the upstream of Mathura meets higher limit which indicates the significant industrial discharge (Table 1). Urban agglomeration like Mathura uses Yamuna water significantly for domestic and irrigational

Table 5: Performance evaluation result for ANFIS model

Models	DO_U			DO_m			DO_d		
	DC	RMSE[a]	MSE[a]	DC	RMSE[a]	MSE[a]	DC	RMSE[a]	MSE[a]
MLR	0.4314	0.1025	0.0105	0.5304	0.0916	0.0083	0.69	1.1	1.2100
ANN	0.4856	0.0975	0.0095	0.6818	0.0754	0.0056	0.91	0.56	0.3136
ANFIS	0.8013	0.0452	0.0020	0.854	0.0441	0.0019	0.91	1.26	0.1876

[a]RMSE has no unit since all the data were normalized
[a]MSE has no unit since all the data were normalized

Table 6: Performance efficiency for the three stations in validation steps

Model type	Model structure	Calibration			Validation		
		RMSE[a]	MES[a]	DC	RMSE[a]	MES[a]	DC
ANFIS - I	11, trimf, 2	0.018	0.0003	0.96	1.26	1.58	0.91
ANFIS - II	10, trimf, 2	0.140	0.01	0.79	1.64	2.68	0.72
ANFIS - III	9, trimf, 2	1.340	1.79	0.89	1.49	2.22	0.88
ANFIS - IV	8, trimf, 2	0.130	0.01	0.89	1.59	2.52	0.80
ANFIS - V	7, trimf, 2	1.440	2.07	0.79	1.50	2.25	0.85
ANFIS - VI	6, trimf, 2	0.240	0.05	0.88	1.59	2.52	0.69
ANFIS - VII	5, trimf, 2	1.460	2.13	0.78	1.59	2.52	0.67
ANFIS - VIII	4, trimf, 2	0.150	0.02	0.89	1.54	2.37	0.78
ANFIS - IX	3, trimf, 2	1.450	2.10	0.78	1.60	2.56	0.66
ANFIS - X	2, trimf, 2	1.540	2.37	0.83	1.48	2.19	0.78

[a]RMSE has no unit since all the data were normalized
[a]MES has no unit since all the data were normalized

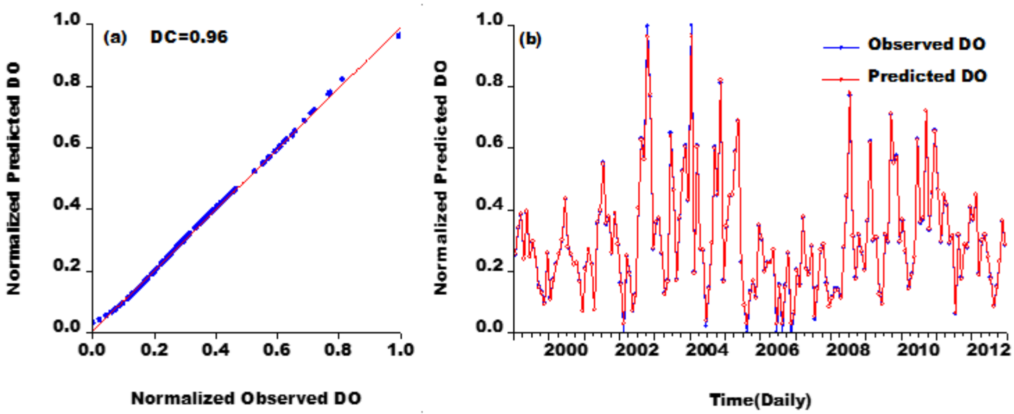

Fig. 6: Evaluation of ANFIS model a) scatter plot (b) observe versus computed time series of DO

supplies. There is a large cluster of industries established at Mathura which discharge the effluent into the catchment water bodies. It is scientifically proven that what happens to upstream based on WQ will benefit or adversely affect the downstream of the river depending on the catchment activities in the upstream region. Understanding the upper, middle and downstream linkage in hydrological process with regards to WQ based on interaction among parameters is essential for water resources management. The upper stream impact can be categorized into human influence or natural impact. In this study, the influence of Mathura downstream depicts the impact of human and industrial influence (wastewater discharge from Mathura upstream). As it is depicted in the Tables 3-5, ANFIS model could lead to the highest value of DC as compared to FFNN and MLR models. Hence for DO prediction of the river, the most efficient model turns out to be ANFIS. However, it can be stated that both the two models (ANFIS and FFNN) are dependable enough in predicting the $DO_{d,}$ as the difference between the DC values for both FFNN and ANFIS are negligible in both steps. However, for the same prediction purpose, ANFIS outperforms FFNN. Meanwhile, MLR model is found to be less reliable in the predictions. The outstanding performances of ANFIS and FFNN models over MLR model may result from nonlinear nature of the parameters which can be represented better by ANFIS and FFNN models.

As shown in Figs. 4, 5 and 6, the accuracy of the fitted computed values of FFNN and ANFIS are almost the same and are close to the observed values, this can be proven by the values of their DCs (Tables 4 and 5). In contrast, MLR computed values are deviated more from the observed values which may be due to the linear nature as justified by MLR performance efficiency shown in Table 3.

Where, DO_u, DO_m and DO_d are dissolve oxygen at upstream, middlestream and downstream of Mathura.

CONCLUSIONS

The river WQ modeling is paramount for preserving the life of aquatic animals. Modeling the water parameter with conventional classical method consumed more time and energy due to complex interactions. AI models can effectively handle nonlinearity and complexity of a system and overcome weaknesses of classic linear models. MLR, FFNN, ANFIS models were developed in modelling DO concentration at three stations of Mathura and the performance of models were computed and compared using DC and RMSE criteria. For all the models, the sensitivity analysis was carried out and MLR was found not to produce a considerable outcome due to its inability of handling nonlinear interactions. Even though, FFNN and ANFIS could handle the nonlinear interactions, it was found that ANFIS model performed better than FFNN model and outperformed MLR model. Generally, the results indicate that for predicting DO centration at Mathura, the input combination of middle and downstream parameters are satisfactory for better prediction of DO. Besides, the ANFIS model proved to have high accuracy when the data for all three stations are used in input layer. However, the best model of FFNN can also be considered in prediction of DO for all the three stations due to its capability. Comparing the three stations, the performance efficiency obtained from DO at downstream proved high merit than DO at upstream and middle stream for all the applied models. The obtained results also indicated that, for the application of these models in the real world, the uncertainty involved in process be addressed. As such, the AI tools should be combined in an ensemble approach in order to integrate a set of models so as to come up with a new model which could produce higher accuracy and more reliable estimates than the single models.

CONFLICT OF INTEREST

The author declares that there is no conflict of interests regarding the publication of this manuscript.

ACKNOWLEDGMENTS

The authors acknowledged the support of Central Pollution Control Board, India for providing the data used in this study.

ABBREVIATIONS

%	Percentage
^{0}C	Degree centigrade
AI	Artificial intelligence
ANFIS	Adaptive neuro fuzzy inference system
ANN	Artificial neural network
ASCE	American Society of Civil Engineer
BOD	Biological oxygen demand
BODu	Biological oxygen demand at upper stream
BOD_m	Biological oxygen demand at middle stream
BOD_d	Biological oxygen demand at downstream
BP	Back propagation
CPCB	Central Pollution Control Board
DC	Determination coefficient
DO	Dissolve oxygen
DO_u	Dissolve oxygen at upper stream
DO_m	Dissolve oxygen at middle stream
DO_d	Dissolve oxygen at downstream
FFNN	Feed forward neural network
FIS	Fuzzy Inference System
LM	Lavenberg-Marquardt
LPG	Linear genetic programming
mg/L	Milligram per litre
MLFF	Multi-layer feed forward
MLP	Multi-layer perceptron
MLR	Multi-linear regression
NRCR	National river conservation program
NWQMP	National water quality monitoring program
pH_u	pH at upper stream
pH_m	pH at middle stream
pH_d	pH at down stream
RBF	Radial Basis Function
RMSE	Root Mean Square Error
STPs	Sewage treatment plant
SVM	Support vector machine
TSK	Takagi-sugeno kang
WQ	Water quality
WT	Water temperature
WT_u	Water temperature at upper stream
WT_m	Water temperature at middle stream
WT_d	Water temperature at downstream

REFERENCES

Abba, S.I.; Elkiran, G. (2017). Effluent prediction of chemical oxygen demand from the wastewater treatment plant using artificial neural network application. Procedia Comput. Sci., 120: 156-163 (**8 pages**).

Abba, S.I.; Hadi, S.J.; Abdullahi, J., (2017). River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. Procedia Comput. Sci., 120: 75-82 (**8 pages**).

Akrami, S.A.; Nourani, V.; Hakim, S.J.S., (2014). Development of nonlinear model based on wavelet-ANFIS for rainfall forecasting at Klang Gates Dam. Water Resour. Manage., 28(10): 2999–3018 (**20 pages**).

Areerachakul, S.; Junsawang, P.; Pomsathit, A., (2011). Prediction of dissolved oxygen using artificial neural network. International Conference on Computer Communication and Management, 5: 524–528 (**5 pages**).

ASCE Task Committee, (2000). Application of artificial neural networks in hydrology. I: Preliminary concepts. J. Hydrol. Eng., 5(2): 115-123 (**9 pages**).

Ay, M.; Kisi, O., (2011). Modeling of dissolved oxygen concentration using different neural network techniques in Foundation Creek, El Paso County, Colorado. J. Environ. Eng., 138(6): 654-662 (**9 pages**).

Chen, W.B.; Liu, W.C., (2014). Artificial neural network modeling of dissolved oxygen in reservoir. Environ. Monit. Assess., 186(2): 1203-1217 (**15 pages**).

Dogan, E.; Ates, A.; Yilmaz, E.C.; Eren, B., (2008). Application of artificial neural networks to estimate wastewater treatment plant inlet biochemical oxygen demand. Environ. Prog. Sustainable Energy, 27(4): 439-446 (**8 pages**).

Gaya, M.S.; Wahab, N.A.; Sam, Y.M.; Samsudin, S.I., (2014). ANFIS Modelling of Carbon and Nitrogen Removal in Domestic. J. Technol., 67(5): 29-34 (**6 pages**).

Hsu, K.L.; Gupta, H.V.; Sorooshian, S., (1995). Artificial neural network modeling of the rainfall runoff process. Water Resour. Res., 31(10): 2517-2530 (**14 pages**).

Jain, A.K.; Jain, K.A.; Punmia, B.C., (2014). Wastewater engineering (including air pollution). (B. Pumia, Ed.) (Second). India: Laxmi Publications (P) LTD 180-222 (**43 pages**).

Karbassi, A.R.; Pazoki, M., (2015). Environmental qualitative assessment of rivers sediments. Global J. Environ. Sci. Manage., 1(2): 109-116 (8 pages).

Legates, D.R.; McCabe, G.J., (1999). Evaluating the use of "goodness□of□fit" measures in hydrologic and hydroclimatic model validation. Water Resour. Res., 35(1): 233-241 (**9 pages**).

Mamdani, E.H., (1974). Application of fuzzy algorithms for control of simple dynamic plant. In Proceedings of the Institution of Electrical Engineers., 121(12): 1585-1588 **(4 pages)**.

Mirbagheri, S.A.; Nourani, V.; Rajaee, T.; Alikhani, A., (2010). Neuro-fuzzy models employing wavelet analysis for suspended sediment concentration prediction in rivers. Hydrol. Sci. J., 55(7): 1175–1189 (**15 pages**).

Najah, A.; El-Shafie, A.; Karim, O.A.; El-Shafie, A.H., (2014). Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. Environ. Sci. Pollut., Res., 21(3): 1658-1670 (**13 pages**).

Nourani, V; Kisi, O.; Komasi, M., (2011). Two hybrid Artificial Intelligence approaches for modeling rainfall – runoff process. J. Hydrol., 402(1–2): 41–59 **(19 pages).**

Nourani, V.; Hakimzadeh, H.; Amini, A.B., (2012). Implementation of artificial neural network technique in the simulation of dam breach hydrograph. J. Hydroinfo., 14(2): 478–496 (**19 pages**).

Nourani, V.; Alami, M.T.; Vousoughi, F.D., (2015). Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. J. Hydrol., 524: 255–269 (**15 pages**).

Nourani, V., (2017). An emotional ANN (EANN) approach to modeling rainfall-runoff process. J. Hydrol. 544: 267–277 (**11 pages**).

Olyaie, E.; Abyaneh, H.Z.; Mehr, A.D., (2017). A comparative analysis among computational intelligence techniques for dissolved oxygen prediction in Delaware River. Geosci. Front. 8(3): 517-527 (**11 page**s).

Parmar, K.S.; Bhardwaj, R., (2015). River water prediction modeling using neural networks, fuzzy and wavelet coupled model. Water Resour. Manage., 29(1): 17–33 (**17 pages**).

Quej, V.H.; Almorox, J.; Arnaldo, J.A.; Saito, L., (2017). ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J. Atmos. Solar-Terrestrial Physics, 155: 62-70 (**9 pages**).

Ranković, V.; Radulović, J.; Radojević, I.; Ostojić, A.; Čomić, L., (2010). Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. Ecol. Model., 221(8): 1239-1244 (**6 pages**).

Sharma, D.; Kansal, A., (2011). Water quality analysis of River Yamuna using water quality index in the national capital territory, India (2000–2009). Appl. Water Sci., 1(3-4): 147-157 (**11 pages**).

Singh, K.P.; Mohan, D.; Singh, V.K.; Malik, A., (2005). Studies on distribution and fractionation of heavy metals in Gomti River sediments—A tributary of the Ganges, India J. Hytrol., 312(1-4): 14-27 (**14 pages**).

Singh, K.P.; Basant, A.; Malik, A.; Jain, G., (2009). Artificial neural network modeling of the river water quality-A case study. Ecol. Modell. 220: 888-895 (**8 pages**).

Solgi, A.; Zarei, H.; Nourani, V.; Bahmani, R., (2017). A new approach to flow simulation using hybrid models. Appl. Water Sci., 7(7): 3691-3706 (**16 pages**).

Takagi, T.; Sugeno, M., (1993). Fuzzy identification of systems and its applications to modeling and control. In readings fuzzy sets for intelligent systems, 387-403 (**17 pages**).

Turan, M.E.; Yurdusev, M.A., (2009). River flow estimation from upstream flow records by artificial intelligence methods. J. Hydrol., 369(1-2): 71-77 (**7 pages**).

Verma, A.K.; Singh, T.N., (2013). Prediction of water quality from simple field parameters. Environ. Earth Sci., 69(3): 821-829 (**9 pages**).

Wen, X.; Fang, J.; Diao, M.; Zhang, C., (2013). Artificial neural network modeling of dissolved oxygen in the Heihe River, Northwestern China. Environ. Monit. Assess., 185(5): 4361-4371 (**11 pages**).

Yetilmezsoy, K.; Ozkaya, B.; Cakmakci, M., (2011). Artificial intelligence-based prediction models for environmental engineering. Neural Network World, 21(3): 193-218 (**26 pages**).

Zadeh, L.A. (1996). Fuzzy sets. In fuzzy sets, fuzzy logic and fuzzy systems: Selected papers by Lotfi A. Zadeh., 394-432 (**39 pages**).

AUTHOR (S) BIOSKETCHES

Elkiran, G., Ph.D. Associate Professor, Faculty of Civil and Environmental Engineering, Near East University, Near East Boulevard 99138, Nicosia, North Cyprus. Email: *gozen.elkiran@neu.edu.tr*

Nourani, V., Ph.D., Professor, Department of Water Resources Engineering, Faculty of Civil Engineering, University of Tabriz, Tabriz, Iran. Email: *nourani@tabrizu.ac.ir*

Abba, S.I., Ph.D. Candidate, Faculty of Civil and Environmental Engineering, Near East University, Near East Boulevard 99138, Nicosia, North Cyprus. Email: *saniisaabba86@gmail.com*

Abdullahi, J., M.Sc., Faculty of Civil and Environmental Engineering, Near East University, Near East Boulevard 99138, Nicosia, North Cyprus. Email: *jazulibinabdallah@gmail.com*

HOW TO CITE THIS ARTICLE

Elkiran, G.; Nourani, V.; Abba, S.I.; Abdullahi, J., (2018). Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river. Global J. Environ. Sci. Manage., 4(4): 439-450.

DOI: 10.22034/gjesm.2018.04.005

url: http://www.gjesm.net/article_32056.html