Dissolved oxygen modeling of an urban stream using grid partitioning and subtractive clustering fuzzy techniques

Sameer Arora
Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India
Corresponding author: E-mail: sameer_arora01@yahoo.co.in, Tel No. +91-98733-96503

Abstract
Dissolved oxygen is one of the prime parameters for assessing the water quality of any stream and the health status of aquatic life. The dissolved oxygen present in the water body plays an essential role in deciding water treatment processes to enhance water quality up to the design standards for the specified water use. Thus, the accurate estimation of dissolved oxygen concentration is necessary to evolve measures for maintaining the riverine ecosystem and designing the appropriate water treatment plants. Machine learning techniques are becoming useful tools for the prediction and simulation of water quality parameters. With these viewpoints, a study was carried out in the Delhi stretch of Yamuna River, India, and physiochemical parameters were examined for five years to simulate the dissolved oxygen using different machine learning techniques. Simulation and prediction competencies of ANFIS grid partitioning (ANFIS-GP) and ANFIS subtractive clustering (ANFIS-SC) were tested on various water quality parameters. Variation in dissolved oxygen was examined on various combinations of parameters. ANFIS-GP has been designed using the Gaussian function, and ANFIS-SC works on the likelihood of cluster centers. Results obtained from the models were evaluated using root mean square error (RMSE) and coefficient of determination (R2) to identify the optimum solution and appropriate combination of parameters that simulate the observed dissolved oxygen. Results of ANFIS-GP and ANFIS-SC indicate that both the models produce suitable solutions for the prediction; however, ANFIS-GP outperforms the ANFIS-SC and could act as a useful tool for defining, planning, and management of water quality parameters.

Keywords: Grid partitioning, Subtractive clustering, Water quality, Dissolved oxygen

1 INTRODUCTION

The production and consumption of dissolved oxygen (DO) in rivers are dynamic (Zahraeifard and Deng, 2012). Dissolve oxygen remains in water as free oxygen, and its concentration varies due to diffusion. The concentration of DO in water depends upon several sources, sinks, and solubility rates. The atmosphere is the most significant external source of oxygen to stream, and photosynthesis plays a significant role as an internal source of oxygen. Photosynthesis contributes more oxygen to water, as the oxygen generated from the algae contains pure oxygen, whereas the atmospheric diffusion contains only 20% oxygen in overall gas transfer at the air-water interface. All the microorganisms, aquatic plants, and aquatic animals consume the oxygen through respiration, known as sinks. All the sinks consume oxygen throughout the day and night.

In contrast, photosynthesis generates oxygen during the daytime only, whereas the algae act as both sources and the sink of oxygen (Arora and Keshari, 2020). There should be synchronization in the source and sinks of oxygen to maintain a healthy riverine ecosystem. Another factor is solubility that further depends upon the pressure, temperature, and salinity of the water. The higher pressure increases the solubility of gas, whereas higher salinity and temperature reduce the solubility rate. All these factors affect the DO concentration in the river along with the depth of the water body. Mathematically, the concentration of DO can be expressed as:

$$DO = DO_{so} S - DO_{si}$$

Where, DO_{so} is the source of DO and DO_{si} is the sink of DO at S solubility. The low concentration of dissolved oxygen in the river for a longer duration increases the inception of several environmental problems (Kisi and Murat, 2012). The river system's Biota starts getting affected if the oxygen content reduces below 30% to the saturation limit. Variation in DO concentration occurs rapidly based on flow available in rivers, velocity, turbulence, the number of organics, and atmospheric reactions involved in the riverine system (Cox, 2003). Anthropogenic activities are becoming the significant sinks of oxygen that consume the available DO through partially or untreated wastewater from the domestic, industrial, commercial and agricultural sectors (Arora and Keshari, 2018). It is mandatory to maintain the equilibrium between the sources and sinks for the aquatic ecosystem's sustainability.

River systems are greatly affected by water abstraction for the municipal supplies and urban wastewater discharge through drains/tributaries during the low flow period. Estimating DO concentration variation for the heavily polluted rivers based on the statistical methods is not the appropriate approach nowadays due to the complex and nonlinear water quality parameters (Cox, 2003; Parmar and Keshari, 2012). Various researchers have profoundly used machine learning techniques to predict the variation in various water quality parameters. For the simulation and forecasting of water quality parameters, artificial neural networks (Singh et al., 2009; Heddam, 2016) and ANFIS (Chen and Liu, 2013; Ay and Kisi, 2016) have been successfully utilized by several researchers. Cox (2003) illustrated various mathematical models available for predicting DO concentration in a lowland river; however, such models require a significant amount of data for the excellent prediction (Kannel et al., 2011).

1.1 Literature Review

Fuzzy logic has several advantages in classification, data mining, interpretation, and optimization of time series data of various fields (Nguyen et al., 2013; Wijayasekara and Manic, 2014). The derivation of fuzzy models significantly depends on linguistic terms designed via membership functions (MFs) and delivers input parameters to the optimization model (Cordon, 2011). The fuzzy theory has been widely used to model the nonlinear behavior for various hydrological applications (Altunkaynak et al., 2005; Keskin et al., 2006; Chang et al., 2015; Khan and Valeo, 2015; Ay and Kisi, 2017; Arora and Keshari, 2020). The fuzzy system can remove the uncertainties from the data and develop the model structure through the rule-based system (Guyonnet et al., 2003; Huang et al., 2010). The identification, validation, optimization, and interpretation can be applied before decision making using fuzzy to manage water resources. Altunkaynak et al. (2005) used the Takagi–Sugeno fuzzy logic approach to model fluctuations in DO at Golder Horn and compared the results with ARMA models. The results reveal that the fuzzy models are more superior to ARMA in the prediction of DO fluctuations. Guldal and Tongal (2010) identified the variation in the depth of water in the lake and compared the accuracy of RNN, ANFIS, and stochastic models using the coefficient of determination. Authors found that RNN and ANFIS perform exceptionally over stochastic models. Moosavi et al. (2013) compared different data-driven models to predict the reservoir's groundwater level at two distinct basins. The author used ANN, ANFIS and ANN-ANFIS coupled models and found that the ANFIS and various combinations of ANFIS perform better over the ANN due to the errors involved in selecting the adequate number of neurons for the ANN model. ANFIS is also advantageous over ANN with the capability of former of analyzing uncertainties in input parameters. Parmar and Bhardwaj (2014) compared the regression, ANN, Wavelet, and ANFIS, to predict the COD in India's

Yamuna River. They also compared the conventional techniques with the wavelet coupled model. Khan and Valeo (2015) applied fuzzy regression and compared it with the Tanaka and Diamond method of fuzzy modeling to predict the DO and found that the ability to record water quality parameters' uncertainty makes the fuzzy regression technique a substantial approach for the prediction of DO. The literature review reflects that the fuzzy modeling techniques can be applied to a wide area with high accuracy. However, detailed studies over the differences between the two approaches (subtractive clustering and grid partitioning) of fuzzy modeling are not available. The application of a correct approach for the prediction of the parameter can improve the results significantly.

As the DO acts as the health indicator of the riverine system, accurate prediction of DO for assessing the state of the water body, designing policies for the water resource management, and appropriate allocation of available water keeping the sufficient amount of flow in rivers are the predominant task. In this study, the grid partitioning and subtractive clustering approaches are analyzed to model the stream passing through a highly urbanized area, discharging tremendous wastewater from domestic, industrial, and agricultural sources through multiple drains.

2 MATERIAL AND METHODOLOGY

ANFIS models are developed for the simulation and prediction of DO. A hybrid algorithm combining the least-squares method and the gradient descent method is used to conserve the search space and minimize the model's operational time. The model's structure is designed using subtractive clustering and grid portioning methods, and various combinations of input parameters are tested out using both methods.

2.1 Adaptive Neuro-Fuzzy Inference System

Adaptive neuro-fuzzy inference system (ANFIS) is the combined structure of the neural network and fuzzy logic. This composite structure allows neurons to record the input data and fuzzy rules to optimize the solution. The fuzzy sets in the model define the fuzzy rule base and make the ANFIS capable of simulating the nonlinear behavior of input parameters. The rule base of the network increases with the number of input parameters. However, it also increases the computational time of the model (Chang and Chang 2006). The ANFIS structure is designed using five different layers that include input, fuzzification, normalization, defuzzification, and output layers. The number of input parameters is defined in the first layer. Fuzzification includes the distribution of membership function to each input parameter and allocation of type of membership function.

If-then rules bases are formed based on the number and type of membership functions defined in the previous step. The fuzzification covers the input into breakable fuzzy sets, and defuzzification again converts the fuzzy sets into output after applying inference processes, normalization, and optimization (Chang and Chang, 2006). However, the FIS rule base can be altered by understanding the relationship between input parameters and reducing the computational time with optimized output. The alteration of the rule base and the modified structure of FIS make it worthwhile for wide application over neural networks (Arora and Keshari, 2020). FIS is designed using Gaussian type membership functions with a hybrid learning algorithm to optimize the model. The structure of FIS dominantly depends upon the type and number membership function selected for modeling. (Babuska and Verbruggen 2003; Sonmez et al. 2017). The overall architecture of the FIS model is shown in figure 1.

[Insert Figure 1]

2.2 Grid Partitioning

Grid partitioning is commonly used to design the FIS, which is a fuzzy portioning method. The minimum distance between two variables is required for each input parameter. The problem region is divided into sub-regions, and input space is further divided into sub-regions to refine the space depending upon the type and number of membership functions selected for designing the model. The rule base of grid partitioned FIS is defined as:

$$x_1 = A_1^{k1}, x_2 = A_2^{k2}, \ldots, x_n = A_m^{kn}$$

$$y_m = y^{k_1 k_2, \ldots, k_n}, k_i = 0,1$$

If, k_i=0, then $A_m^{kn} = a_i$, where a_i is the minimal value of the input parameter. If, k_i=1, then $A_n^{kn} = b_i$, where b_i is the maximal value of the input parameter, and both the values would be computed using the least square method. The input sub-region is divided into m^{th} sub-regions, where x = $x_1, x_2, x_3, \ldots, x_m$. The membership function for the fuzzy term A_m^{ki} would be:

$$\mu_m^0(x_i) = \frac{b_m - x_i}{b_m - a_m}$$

For k_i=0

$$\mu_m^1(x_i)=\frac{x_i-a_m}{b_m-a_m}$$

For k_i=1

The output corresponding to m^th^ sub-region is written as:

$$O=\sum_{k_1k_2,\ldots,k_n} A_m^{k_1k_2,\ldots,k_n}(x_i)*y_m^{k_1k_2,\ldots,k_n}$$

The sub-regions are divided on the maximum value of error from the training samples. Once the maximum errors of every sub-region are achieved, the region splits into two regions, and the new approximation error is the minimum of the new sub-region. The sub-region splitting continues until the two regions' errors become constant, as shown in Figure 2. The maximum error obtained from the sub-region at which split occur is written as:

$$E_m=\frac{1}{N_m}\sum_{x_m}\left[\frac{1}{2}(x_j-t_j)^2\right]$$

Where, E_m is the error obtained from m^{th} sub-region from the N_m training samples and x_j and t_j are the output generated from the model and targeted respectively from j^{th} training samples.

2.3 Subtractive Clustering

The partitioning method is preferred when the knowledge about the center's distribution is not adequate (Benmouiza and Cheknane, 2018). In subtractive clustering, the rule base formed is equivalent to the membership function formed. In this method, each data point is considered the center, and the importance of each center is identified through the data point in the center's neighborhood. The process runs through several iterations and allocates the center by identifying the most influential center with the highest number of data points in its surrounding. The radius of the cluster of points is identified using the center of neighboring points. The process repeats until all the data points fall within the radius of every cluster. The potential of the data point is written as:

$$P_i=\sum_{j=1}^{n}\exp\left(\frac{-\|x_i-x_j\|^2}{0.5r^2}\right)$$

Pi is the potential index of xi data points, r is the radius where all the neighborhood's data points fall. The second iteration is calculated as:

$$\acute{P}_i=P_i-P_{c1}\exp\left(\frac{-\|x_i-x_{c1}\|^2}{0.5{r_a}^2}\right)$$

Where, P_{c1} represents the potential of cluster 1 and r_a is the K_r, where K_r is the positive integer usually 1.5. The process is repeated, and a radius of cluster is recalculated until a sufficient number of cluster centers are not generated.

[Insert Figure 2]

2.4 Model Development

The sampling data is divided into two parts of 70:30 to design the model, where 70 percent of sampling data points are used for model training, and 30 percent are used for testing the model. Different sets of parameters are selected for designing the model. The four models are designed using temperature, BOD, COD, conductivity, and ammonia. To estimate the DO of any stream, the concentration of organic matter present in water plays a significant role as for the decomposition of organic content, DO acts as a source of energy for aerobic bacteria. The type of bacteria that would develop depends upon the temperature of the stream. Hence, the first model (M1) is developed with the temperature, BOD, and COD as input parameters. The addition parameter selected in the second model (M2) is conductivity. The amount of DO also varies with dissolved solids due to the oxygen demand of solids. The presence of ammonia reflects the generation of algae in the water, which acts as both the source and sink of oxygen. The third model (M3) is designed by combining the base parameters with ammonia. Moreover, the fourth model (M4) is designed to look at the combined effect of conductivity and ammonia and the base parameters on the river's dissolved oxygen.

Both the fuzzy portioning methods are used to design the model with similar parameters. The input parameters selected to design the FIS models, as shown in Table 1, are used for both the grid partitioning and subtractive clustering method.

[Insert Table 1]

The coefficient of determination (R2) and root mean square error (RMSE) are evaluated and compared with the observed dissolved oxygen to analyze models' performance. The lowest RMSE nears to zero value indicates the adequate model and when R^2 nears to 1 represents a better correlation between the observed and the predicted values obtained from the FIS model. Formulas used to identify the performance of models are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \left(p_i - o_i \right)^2}{n}}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(p_i - o_i \right)^2}{\sum_{i=1}^{n} \left(p_i - \bar{o}_i \right)^2}$$

3 STUDY AREA

Delhi is one of India's biggest and dense cities, and all the wastewater generated from various sectors (domestic, commercial, industrial, and agricultural) of Delhi joins the Yamuna River. Some part of the wastewater before its confluence with the river passes through the treatment processes. However, the percentage of wastewater that gets treatment is too less than the wastewater directly discharged into the drain through irregular means, ultimately joining the river. The samples were collected from Nizamuddin, Delhi, 16 km downstream of the Wazirabad barrage in Delhi. Yamuna river travels around 375 km before reaching Delhi, and the flow of the river gets obstructed at the Wazirabad Barrage for water supply to Delhi. The fresh water remains low in the river throughout the year, and only wastewater from Delhi flows in the river except during the monsoon period. In between the Wazirabad barrage and the Nizamuddin, Yamuna River receives effluents from several drains, out of which maximum effluents are discharged from the Najafgarh drain, which contains the flows 2.5 times to the water available in the river. As the Najafgarh drain joins the river just after the Wazirabad barrage (0.5 km downstream), it causes maximum damage to the river's water quality (CPCB, 2006).

The stretch of river Yamuna in Delhi is one of the most polluted sections of the river, having a total length of 1376 km. The water quality falls in category E, which indicates that the river's water is not designated for any use. The BOD load increases in the river up to 80 tons/day after the Najafgarh drain's confluence. The sampling location also receives the effluents from a thermal power plant located at the river's right bank. The river observes one road bridge, one railway bridge, and one metro railway bridge that further obstructs the flow and causes silting near the bridges' piers. The dissolved oxygen content of the river falls to zero in this stretch and causes the significant degradation of aquatic plants and animals. The generation of anaerobic conditions has also been observed in the river due to the decomposition of organic matter in the absence of oxygen.

The sampling location is selected after considering the distance from the Wazirabad barrage, upstream of which water quality is much clear, time taken by the flow to provide sufficient mixing of wastewater with the river water to collect the homogeneous sample. The river receives the effluent load from the right bank only and has heavy habitation close to the right bank, as shown in Figure 3. On the left bank, the river has a flood plain to cater to the excess water during the flood; however, the encroachment of the flood plain is another problem of Yamuna River that causes the frequent release of flow during the monsoon period and the rest of the time river flow with low discharge.

[Insert Figure 3]

The water samples were collected for five years every month, and physiochemical analyses have been carried out as per the standard procedure (APHA, 2005) after preserving the samples with the recommended reagent. During summers, the average temperature remains around 32^0C, affecting the saturation rate of dissolved oxygen in the river. Delhi receives the precipitation in the form of rainfall from July to September, and sufficient water flows in the river during this period, BOD level falls below 50 mg/l, and DO rises to 2 mg/l. However, these conditions still indicate the riverine ecosystem's terrible health, but somewhat better than the river's non-monsoon state.

4 RESULTS AND DISCUSSION

An appropriate combination of parameters is selected for the development of ANFIS models after several trial and error. Takagi-Sugeno (TS) algorithm is used for the development of the model. Three MFs are selected for each input parameter with Gaussian function type membership function, and constant type MF is selected for the output generation. To generate the FIS, both the methods, grid portioning and subtractive clustering, are used and compared. The structures of the models are shown in Table 2. In all the models developed using grid partitioning, three membership functions are used for each input, whereas the number of membership functions varies with the number of input parameters in each model in the case of subtractive clustering. In M1, there are three input parameters, and five membership functions are used; however, in M2 and M3, have four input parameters in each model. Therefore six membership functions are used in both models. However, M4 contains five input parameters, and nine membership functions are used to design the model.

Along with the increase in input parameters, the rule base also elevates exponentially in Grid partitioning whereas, and this does not hold the truth in the case of subtractive clustering. The optimization is carried out for both the partitioning method using hybrid learning for all the models. The varying numbers of epochs for each model are used until the observed error gets constant or reduced to the minimum.

[Insert Table 2]

The performance of the models is identified using RMSE and R^2. Results of ANFIS-GP and ANFIS-SC indicate that both the models produce suitable solutions for the prediction. The M1 of both the ANFIS-GP and ANFIS-SC produces considerable but high RMSE compared to other models. It indicates that the input parameters used for modeling are insufficient to explain the phenomenon of dissolved oxygen variation in the river. However, R^2 of more than 0.75 indicates that input parameters are substantial factors that affect the variability in dissolved oxygen concentration. The performance of M2 increases considerably over M1, where M2 includes conductivity as an extra parameter other than temperature, BOD, and COD as considered in M1. The M3 model contains ammonia as an extra parameter, including parameters considered in M1; however, the RMSE in M3 increases compared to M2 for the grid partitioning method. Comparing M2 and M3 of ANFIS-GP indicates that conductivity is more significant than ammonia to predict dissolved oxygen. Simultaneously, the combination of both the conductivity and ammonia and other parameters is considered in M4 and produces magnificent results over other methods. The RMSE of M4 is found only 0.049, and R^2 is 0.953 for ANFIS-GP. The correlation between the observed and predicted dissolved oxygen from all the ANFIS models obtained using both the partitioning methods are shown in Figure 4.

[Insert Figure 4]

The performance of ANFIS-SC shows similar results for models. The highest RMSE is found in M1 and lowest in M4 as similar to ANFIS-GP. The M2 and M3 deliver approximately similar results as both the models contain a similar number of input parameters and membership functions, which indicates that the output of ANFIS-SC essentially depends upon the number of membership functions rather than the characteristics of input parameters. The M4 produces the RMSE of 0.150, which is the lowest among all the

ANFIS-SC models, however higher than the M4 of ANFIS-GP. The ANFIS-GP classifies the model based on the rule base developed for each model, and that is the necessary condition that results in better performance. Therefore, from the results, it is evident that the ANFIS-GP outperforms the ANFIS-SC and could act as an effective tool for defining, planning, and managing water quality parameters.

[Insert Table 3]

5 Conclusion

The study was carried out for the simulation of dissolved oxygen using ANFIS. The models are developed using two proportioning methods, grid partitioning and subtractive clustering, and obtained results were compared to simulate the dissolved oxygen in a river. Various combinations of input parameters are used to develop the models using both the ANFIS method (ANFIS-GP and ANFIS-SC), and the applicability of ANFIS models are tested using water quality parameters of the Yamuna River. The M4 model of ANFIS-GP is found with the lowest RMSE and the maximum R^2 of 0.953. However, all the models of the ANFIS-GP worked well over the ANFIS-SC and have shown a good correlation with the observed values of the dissolved oxygen. The extensive formulation of the rule base helps identify vital parameters and improves the accuracy of the model. However, it is expected that the accuracy of the model can be further improved with the large data set.

References

1. Abba SI, Hadi SJ, Abdullahi J, (2017) River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. Procedia Computer Science 120:75–82
2. Altunkaynak A, Ozger M, Cakmakcı M, (2005) Fuzzy logic modeling of the dissolved oxygen fluctuations in Golden Horn. Ecological Modelling 189: 436–446
3. APHA (2005) Standard method to the examination of water and wastewater. 21st ed. Washington, DC: American Public Health Association, American Water-Works Association, Water Environment Federation.
4. Arora S, Keshari AK (2018) Estimation of re-aeration coefficient using MLR for modelling water quality of rivers in urban environment. Groundwater for sustainable development 7: 430-435

5. Arora S, Keshari AK (2020) Monte Carlo simulation and fuzzy modelling of river water quality for multiple reaches using QUAL2Kw. Environmental Processes and Management (Water Science and Technology Library), Springer 91:3-24
6. Ay M, Kisi O, (2012) Modeling of Dissolved Oxygen Concentration Using Different Neural Network Techniques in Foundation Creek, El Paso County, Colorado. Journal of Environmental Engineering, 138(6): 654–662.
7. Ay M, Kişi Ö, (2016) Estimation of dissolved oxygen by using neural networks and neuro fuzzy computing techniques. KSCE Journal of Civil Engineering, 21(5): 1631–1639.
8. Babuška R, Verbruggen H, (2003). Neuro-fuzzy methods for nonlinear system identification. Annual Reviews in Control, 27(1): 73–85.
9. Benmouiza K, Cheknane A, (2018) Clustered ANFIS network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. Theoretical and Applied Climatology.
10. Chang F-J, Chang Y-T, (2006) Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. Advances in Water Resources, 29(1): 1–10.
11. Chang FJ, Tsai YH, Chen PA, Coynel A, Vachaud G, (2015) Modeling water quality in an urban river using hydrological factors-Data driven approaches. Journal of Environmental Management, 151: 87-96
12. Chen W-B, Liu W-C, (2013) Artificial neural network modeling of dissolved oxygen in reservoir. Environmental Monitoring and Assessment 186(2):1203–1217.
13. Cordón O, (2011) A historical review of evolutionary learning methods for Mamdani-type fuzzy rule-based systems: designing interpretable genetic fuzzy systems International Journal of Approximate Reasoning, 52: 894–913
14. Cox BA, (2003) A review of dissolved oxygen modelling techniques for lowland rivers. Science of total Environment 314 –316: 303-334
15. CPCB (2006) Water Quality Status of Yamuna River (1999–2005): Central Pollution Control Board, Ministry of Environment & Forests, Assessment and Development of River Basin Series: ADSORBS/41/2006-07.
16. El-Shafie A, Taha MR, Noureldin A, (2006) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. Water resource Management, 21:533-556
17. Guldal V, Tongal H, (2010) Comparison of Recurrent Neural Network, Adaptive Neuro-Fuzzy Inference System and Stochastic Models in Egirdir Lake Level Forecasting. Water resource Management, 24:105-128

18. Guyonnet D, Bourgine B, Dubois D, Fargier H, Come B, Chilès J-P, (2003). Hybrid Approach for Addressing Uncertainty in Risk Assessments. Journal of Environmental Engineering, 129(1): 68–78.
19. Huang Y, Chen X, Li YP, Huang, GH, Liu T, (2010). A fuzzy-based simulation method for modelling hydrological processes under uncertainty. Hydrological Processes, 24(25): 3718–3732.
20. Kannel PR, Kanel SR, Lee S, Lee Y-S, Gan TY, (2011). A Review of Public Domain Water Quality Models for Simulating Dissolved Oxygen in Rivers and Streams. Environmental Modeling and Assessment 16:183-204
21. Keshtegar B, Heddam S, (2018) Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. Neural Computing Application 30(10): 2995-3006.
22. Keskin ME, Taylan D, Terzi O, (2006) Adaptive neural-based fuzzy inference system (ANFIS) approach for modelling hydrological time series. Hydrological Sciences Journal, 51(4): 588-598
23. Khan UT, Valeo C, (2015) A new fuzzy linear regression approach for dissolved oxygen prediction. Hydrological Sciences Journal, 60(6): 1096-1119
24. Li YP, Huang GH, Huang YF, Zhou HD, (2009) A multistage fuzzy-stochastic programming model for supporting sustainable water-resources allocation and management. Environmental Modelling and Software, 24:786-797
25. Moosavi V, Vafakhah M, Shirmohammadi B, Behnia N, (2013) A Wavelet-ANFIS Hybrid Model for Groundwater Level Forecasting for Different Prediction Periods. Water resource Management, 27:1301-1321
26. Parmar D, Keshari AK, (2012) Sensitivity analysis of water quality for Delhi stretch of river Yamuna, India. Environmental monitoring and assessment 184: 1487–1508
27. Parmar KS, Bhardwaj R, (2015) River Water Prediction Modeling Using Neural Networks, Fuzzy and Wavelet Coupled Model. Water resource Management, 29:17-33
28. Singh KP, Basant A, Malik A, Jain G, (2009) Artificial neural network modeling of the river water quality—A case study. Ecological modelling 220: 888-895.
29. Sonmez AY, Kale S, Ozdemir RC, Kadak AE, (2018) An Adaptive Neuro-Fuzzy Inference System (ANFIS) to Predict of Cadmium (Cd) Concentrations in the Filyos River, Turkey. Turkish Journal of Fisheries and Aquatic Sciences 18:1333-1343

30. Wijayasekara D, Manic M, (2014) Data driven fuzzy membership function generation for increased understandability. 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). doi:10.1109/fuzz-ieee.2014.6891547
31. Zahraeifard V, Deng Z-Q, (2012) VART Model-Based Method for Estimation of Instream Dissolved Oxygen and Reaeration Coefficient. ASCE Journal of Environmental Engineering 138: 518-24.

Table 1: Development of ANFIS models

Model	Parameters	Output
M1	Temp., BOD, COD	DO
M2	Temp., BOD, COD, Cond.	DO
M3	Temp., BOD, COD, Ammonia	DO
M4	Temp., BOD, COD, Cond., Ammonia	DO

Table2: ANFIS model structure

	Grid Partitioning	**Subtractive Clustering**
No. of MF	3 for each input in each model	5 for M1 6 for M2 and M3 9 for M4
Type of MF	Gaussian	Gaussian
Optimization model	Hybrid learning	Hybrid learning
Fuzzy rules	M1 – 27 M2 – 81 M3 – 81 M4 – 243	M1 – 5 M2 – 6 M3 – 6 M4 – 9

Table 3: Performance of developed models

	Grid Partitioning		**Subtractive Clustering**	
	RMSE	R^2	RMSE	R^2
M1	0.642	0.758	0.458	0.824
M2	0.181	0.908	0.287	0.872
M3	0.308	0.861	0.284	0.871
M4	0.049	0.953	0.150	0.911