Simulation of BOD-DO Modeling in Mahanadi River System lying in Odisha using ANN, India

Nibedita Guru[1] and Ramakar Jha[2]

[1]*(Research Scholar,Department of Civil Engineering, National Institute of Technology, Rourkela, Odisha, India)*
[2]*(Professor, Department of Civil Engineering, National Institute of Technology, Rourkela,Odisha, India)*

Abstract: *Most of the water resources have been utilized for the disposal of municipal and industrial wastes. To address the non-linearity, subjectivity, transfer and transformation rule of the pollutants and complexity of the cause-effect relationships between water quality variables and water quality status, development and use of water quality model is of utmost importance. Owing to the random discharge of point pollution from various sources has not only rendered such water bodies eutrophic but also their advantageous uses such as water supply, irrigation, recharge of ground water, recreation and habitat for flora and fauna have been adversely affected. Oxygen- demanding substances are major contaminants in domestic and municipal wastewater. The main indicators of river pollution which deals with the oxygen domestic conditions of the river are Biochemical oxygen demand (BOD) and dissolved oxygen (DO). The Multi-layer Perceptron (MLP) neural network techniques was used to estimate for the analysis of point source pollution in terms of BOD and DO concentration and the neural network model is developed using the data collected from the upstream and downstream stations on Mahanadi river system lying in Odisha. The accuracy performance of training, validation and prediction of seasonal water quality parameters has been tested. To test the validity of ANN model, correlation coefficient (R) statistical model was used.*
Keywords: *Pollution; BOD; DO; ANN; Correlation Coefficient.*

I. Introduction

The quality of water can be negatively influenced by natural phenomena, but the main reason for impaired water quality is contamination caused by human activities. Urban and industrial development, use of chemical and fertilizers in farming, mining activities, combustion of fossil fuels, stream-channel alteration, animal feeding operations, and other human activities has changed the quality of natural waters. It has been found that the global freshwater consumption raised by six times at above twice the rate of population growth from the literature during 1900 and 1995 (WMO, 1997). In the models biodegradable organic matter is taken into consideration by a parameter termed "Biochemical oxygen demand, BOD" and is defined as the amount of oxygen consumed by microorganisms from a unit volume of water, while they decompose organic matter, during a specified period of time. The aquatic ecosystem life is generally, required dissolved oxygen for their metabolism process (i.e. breathing). Indian rivers are polluted due to discharge of organic sewage and industrial effluents. The water quality monitoring of major rivers indicates that organic pollution and almost all the water sources from surface are infected to some extent. The rate at which dissolved oxygen is used will depend on the quantity of the organics, the ease with which they are biodegraded and the dilution capacity of the stream. Mainly, the dissolved oxygen in water bodies is dependent upon temperature, salinity, turbulence and atmospheric pressure. Bio-depletion and re-aeration processes also control dissolved oxygen contents. If the dissolved oxygen concentration drops below that required by certain organisms living in the water, these organisms will die. In recent year, Artificial Neural Network (ANN) models with various numerical analysis techniques are being used to develop non-linear functions for water quality modeling.

II. Materials and Methods

2.1 Study Area

In Odisha River Mahanadi is prime river and largest one (Fig. 1). The geographical co-ordinates of Mahanadi basin lying in Odisha lies between $85^0 30'$ to $85^0 50'$ East longitudes and $20^0 09'$ to $20^0 15'$ North latitudes. River Mahanadi raises from a small pool located at about 6 km from Pharsiya village in the Amarkantak hills of Bastar Plateau, which lies on the extreme south east of Raipur district of Chhattisgarh state. Out of total length of 851 km, it covers 494 km. in Odisha state. Ib,Ong, Tel, Hariharjore and Jeera are the main tributaries and Kathajodi, Kuakhai, Devi and Birupa are the major distributaries of Mahanadi in Odisha. Major towns located on the bank of this river are Sambalpur, Sonepur, Cuttack and Paradeep. The catchment area of Mahanadi spreads over 141600 square km (65,628 square km in Odisha).

The basin having soil types are red and yellow soils, laterite soils. The river and Tel sub-basin is the most densely inhabited and agriculturally affluent part of the area with compact settlements. Forest is dominated some of the lower parts of river having types of tropical moist deciduous, dry deciduous and the coastal forests. The important industries in river basin lying Odisha are aluminum factories at Hirakund and Korba, paper mill near Cuttack and cement industries at Sundargarh.
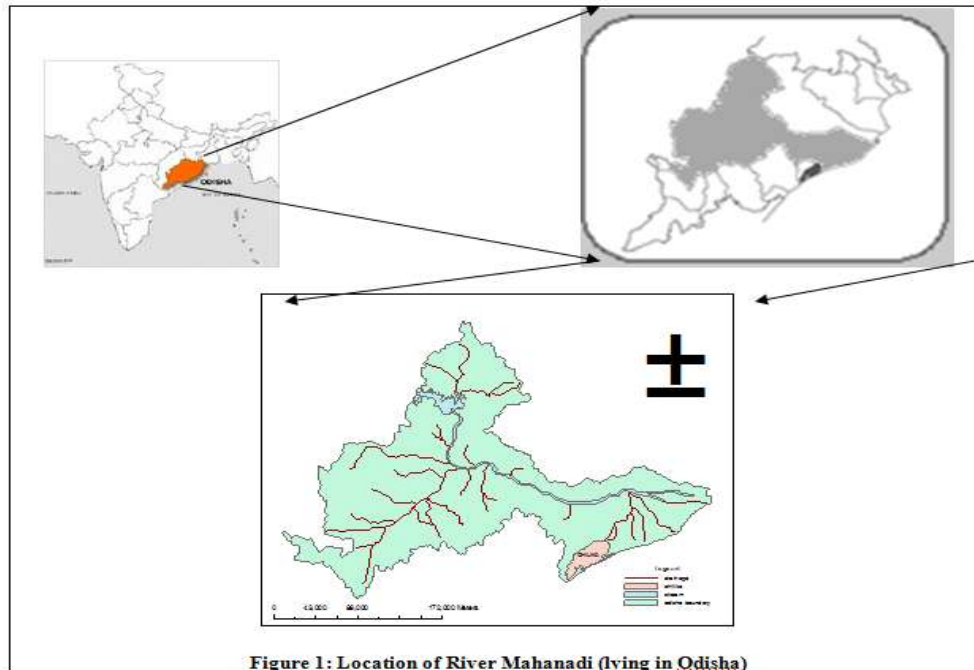

Figure 1: Location of River Mahanadi (lying in Odisha)

2.2 Artificial Neural Networks for BOD-DO Modeling

An ANN is a simplified mathematical and computational that inspired by the structural/ functional aspects of biological neural network. The application of ANNs to water resources problems has become very popular due to their power and potential in modeling nonlinear systems. ANN has a number of data processing elements called neurons or nodes. The neurons in the input layer receive the input vector and transmit the values subsequently to the next layer across connections. This process is continued until the output layer is reached. A three-layer feed forward ANN, shown in Fig. 2, has an input layer, an output layer and a hidden middle layer. The solution to the input problem is emanating from the output nodes. When the interconnection weights are modified, the ANN output changes.
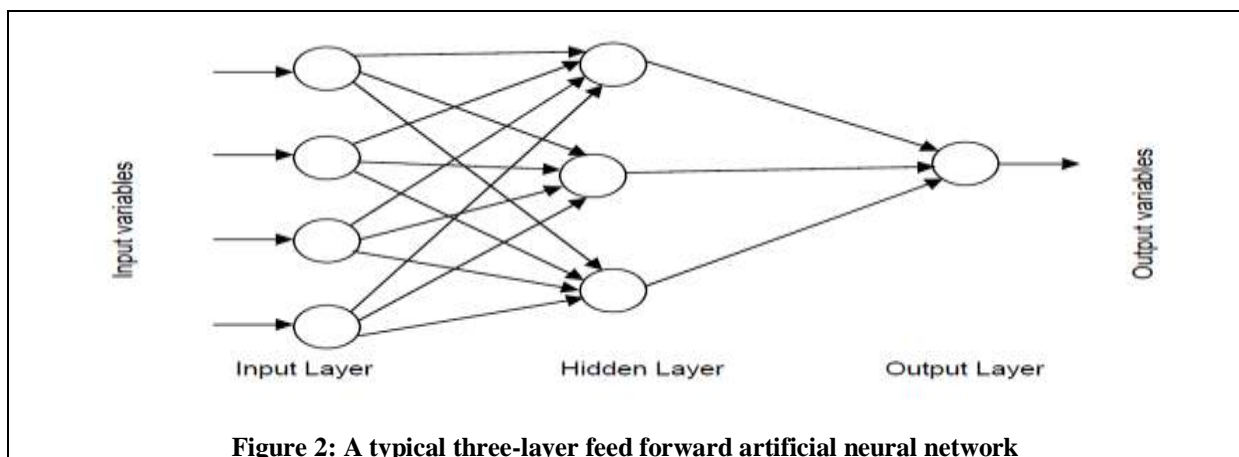

Figure 2: A typical three-layer feed forward artificial neural network

2.2.1 Training of an ANN

An ANN stores the knowledge about the problem in terms of weights of interconnections. Training is generally, refers as the method of formative ANN weights and it is trained with a input and output training datasets. At the beginning of training, the initial value of weights can be assigned arbitrarily or based on

Simulation of BOD-DO Modeling in Mahanadi River System lying in Odisha using ANN, India

experience and is changed systematically by the learning algorithm. The difference between the ANN computed output and the actual output is small, hence considered ANN is trained. By multiplying each neuron with every input by its inter-connection weight an output will be produce, sums the product, and then passes the sum through a transfer function having S-shaped curve which is increasing steadily, called a sigmoid function. This function is continuous, differentiable everywhere, and is monotonically increasing. The input to the function can vary between $\pm\infty$ and the output is always bounded between 0 and 1.

A popular algorithm to adjust the inter-connection weights during training and is based upon the generalized delta rule popularized by Rumelhart *et al.* (1986) is the back-propagation (BP) algorithm error. The actual result is subtracted from the target result to find the output-layer errors and is "back-propagated" through the network and is used to adjust weights. Sometimes the training will take a long time due to high number of nodes in the hidden layer and the network may sometimes over fit the data (Karunanithi *et al.*, 1994). An over fit network typically has very small error for the training data set, but large error for the validation data set. After training is complete, the weights are frozen and the ANN's performance is to be validated and the performance measures for training data sets are better than for validation data sets. However, if the performance of an ANN on the validation or test data set is poor, it shows that either the ANN has not been able to successfully learn the underlying behavior of the process, or there might be a discontinuity between the training and validation data sets. In many books such as Vemuri (1992) and Yegnanarayana (1999), theory of ANN has been described properly.

In the present study, the available data are divided into two parts. The first part is used to calibrate the model and the second to validate it and the optimal length of calibration data depends upon the number of estimated parameters. The general practice is to use about two-thirds of the data for calibration and the remainder for validation. Out of the data sets, 2/3 of the data sets were randomly selected and used for training. Remain data were used for testing and validating the ANN model functions developed during calibration. Upstream BOD values have been used to estimate the BOD values of downstream stations. However for DO estimation of downstream station various combinations of BOD and DO values are used. During training, the number of nodes in the hidden layer was systematically changed and the value that gave the best result for a data set was finally adopted. For the hidden layer, a tan-sigmoid transfer function was used and, for the output layer, a linear transfer function was chosen. The multilayer perceptron technique was used to train the ANNs. It requires iterative training, which may be quite slow for large number of hidden units and datasets, but the networks are quite compact, execute quickly once trained, and in most problems yield better results than the other types of networks. The ANN training epoch consisted of 1000 cycles.

A statistical criteria is available to compare the goodness/adequacy of a given model. Results of ANN models have been compared with the measured data on the basis of following correlation coefficient (R), as given below.

$$R = \frac{\sum(Q_0-M_0)(Q_P-M_P)}{\sqrt{\sum(Q_0-M_0)^2\sum(Q_P-M_P)^2}} \qquad (2.1)$$

Where Q_0 and Q_P are the observed and estimated concentrations at the time step, M_0 and M_P are the mean of the observed and estimated concentrations respectively, and N is the total number of observations of the data set.

III.   Results and Discussion

In order to determine the number of nodes in the hidden layer and transfer functions different ANN models were constructed and tested. Selection of an appropriate number of nodes in the hidden layer is very important aspect as a larger number of these may result in over-fitting, while a smaller number of nodes may not capture the information adequately. Subsequently, two different ANN models were constructed for the computation of DO and BOD in the river water. For BOD modeling various combinations of data structures were tried for different reaches and for reaches 1 to 5, data structures showing the highest correlation coefficient were considered during training of the data sets. It has been found that the developed data structures show very good results during validation with very high correlation coefficients ranging between 0.81 to 0.97. Fig. 3.1 shows the plots between measured and computed values of BOD for training and validation data sets at different river reaches.

www.iosrjournals.org   54 | Page
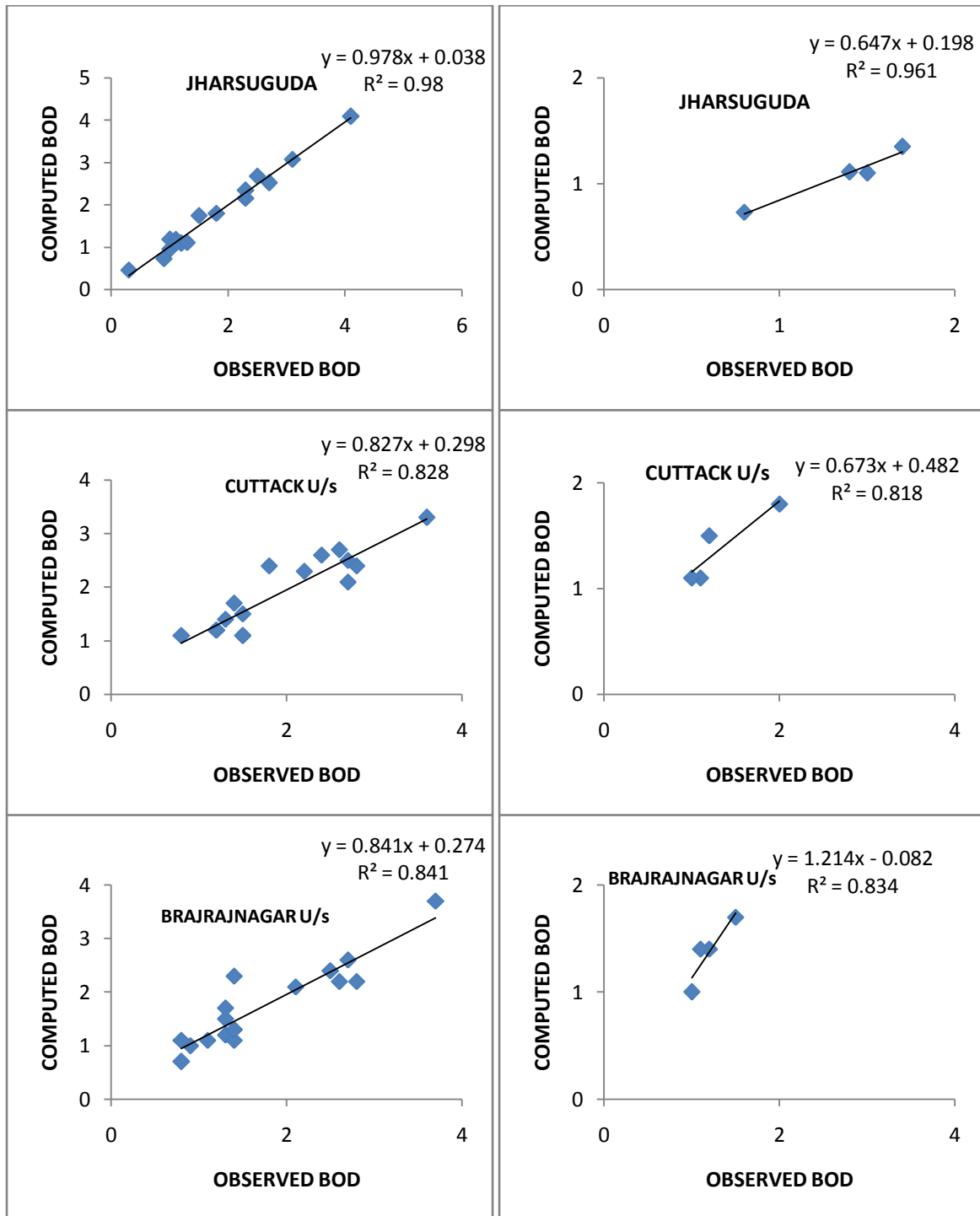
(a) Training data set   **(b) Validation data set**

Figure 3.1 Comparison of the model computed and Observed BOD levels in the river water

Similar to BOD modeling, various combination of data structures were tried for different reaches for DO modeling and for reaches 1 to 5, data structures showing the highest correlation coefficient were considered during training of the data sets. It has been found again that the developed data structures show very good results during validation with very high correlation coefficients ranging between0.98 to 0.99. Fig. 3.2 shows the plots between measured and computed values of DO for training and validation data sets at Brijrajnagar D/s and Cuttack U/s sampling station.
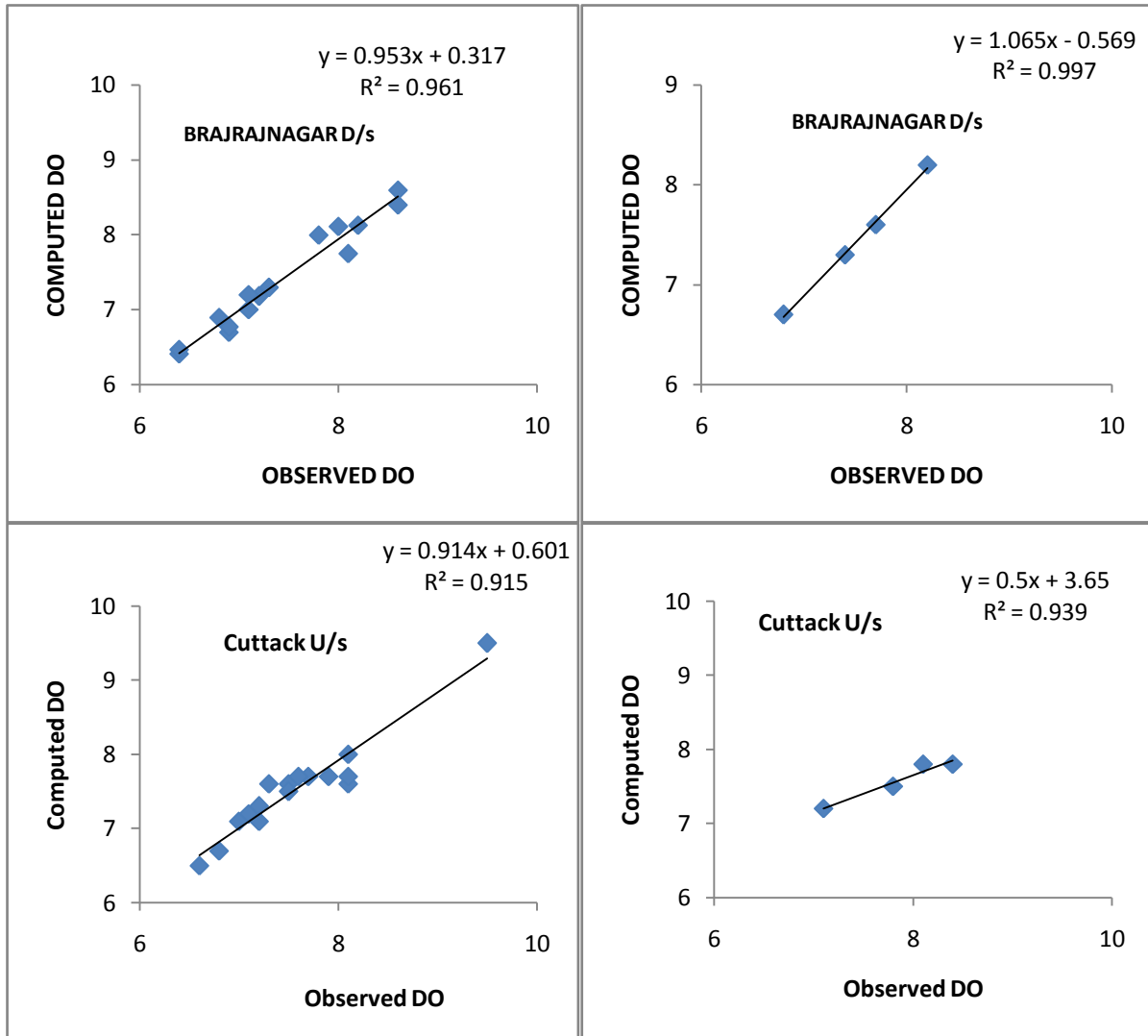
(a) Training data set   **(b) Validation data set**
Figure 3.2 Comparison of the model computed and Observed DO levels in the river water

The correlation coefficient (R) as computed for the training and validation data sets used for the two models (DO and BOD) are presented in TABLE 3.1. The results suggest for a good-fit of the BOD and DO models to the data set using proper data structure in ANN. The results indicate the applicability of neural network model to recognize the pattern of the water quality variable and to provide good predictions of the monthly variations of BOD and DO in Mahanadi River lying in Odisha.

Table 3.1: Data structure and their error statistics for training and validation data sets

MODEL	ANN- Structure	DATASETS	R
BOD	1-8-1	TRAINING	0.98
		VALIDATION	0.98
	1-18-1	TRAINING	0.9
		VALIDATION	0.88
	1-3-1	TRAINING	0.9
		VALIDATION	0.92
	2-7-1	TRAINING	0.91
		VALIDATION	0.92
DO	6-3-1	TRAINING	0.97
		VALIDATION	0.99
	4-3-1	TRAINING	0.93
		VALIDATION	0.99

The results obtained are comparable with the BOD-DO model used in the previous section. However, no extensive data sets are required to be monitored to simulate the BOD and/or DO values in the downstream sites.

IV.   Conclusion

The prospective of an artificial neural network technique (ANN) was examined by comparing the results of observed and estimated BOD and DO in the Mahanadi River and from above discussion it was found that for prediction of the BOD and DO in the Mahanadi River lying in Odisha an ANN model appears to be a useful tool. The results are comparable and in some cases better. To test the validity of these models, performance evaluation was done using error statistic.

Acknowledgement

References

[1]   Butcher, J.B. & Covington, S., Dissolved Oxygen analysis with temperature dependence, J. Environ. Eng., 1995, pp. 756-759.
[2]   Chapman, D (1992) Water Quality Assessments, 1st Edition, Chapman and Hall Ltd., London, 80-81.
[3]   Cox, B. A., A review of dissolved oxygen modeling techniques for lowland rivers, The Science of the Total Environment, 2003, pp.   303-334.
[4]   E.Dogan, B.Sengorur and R.Koklu, "Modeling biological oxygen demand of the Melen river in Turkeyusing an artificial neural network technique," Journal of Environmental Management, 2008, pp1229-1235.
[5]   Garcia, A., Revilla, J.A, Medina, R., Alvarez, C. & Juanes, J. A., A model for predicting the tempoeral evolution of dissolved oxygen concentration in shallow estuaries, Hydrobiologia,2002, pp 205-211.
[6]   G.Civelekoglu, N.O. Yigit, E.Diamadopoulos and M.Kitis, "Prediction of Bromate Formation Using Multi-Linear Regression and Artificial Neural Networks," Journal of Ozone Science and Engineer, Taylor&Francis, vol.5, no.5, 2007, pp.353-362.
[7]   Lopes JF, Dias JM, Cardoso AC, Silva CIV (2005) The water quality of the Ria de Aveiro lagoon, Portugal: From the observations to the implementation of a numerical model. Marine Environmental Research 60: 594–628.McBean, E. A., & Al-Nassri, S., 1988, Uncertainty in Suspended Sediment Transport Curves. ASCE, 114(1), 63-73.
[8]   Radwan M, Willems P, El-Sadek A, Berlamont J (2003) Modelling of Dissolved Oxygen and Biochemical Oxygen Demand in River Water Using a Detailed and Simplified Model.   Int. J. River Basin Management, 1 (2) 97-103. Jang J. S. R., 1993 ANFIS: adaptivenetwork-based fuzzy inference system IEEE Trans. on Syst. Man Cybern. 23, 65−85
[9]   Simon Haykin, "Neural Networks:A Comprehensive foundation second edition", Pearson Prentice Hall, Delhi India, 2005.
[10]   S.Palani, S.Liong, P. Tkalich and J.Palanichamy. Development of a neural network for dissolved oxygen in seawater. Indian Journal of Marine Science. 2009, pp.151-159.
[11]   Suen JP, Eheart JW, Asce M (2003) Evaluation of Neural Networks for Modelling Nitrate Concentration in Rivers. Journal of Wat. Res. Plan. And Man.129: 505-510.
[12]   Zaheer I and Bai CG (2003) Application of Artificial Neural Network for Water Quality Management. Lowland Technology International 5 (2) 10-15.