Environmental Research Journal 8 (2): 55-63, 2014
ISSN: 1994-5396

Comparison of Artificial Neural Network Algorithm for Water Quality Prediction of River Ganga

Aradhana Giri and N.B. Singh
Department of Civil Engineering, Institute of Engineering and Technology,
Sitapur Road, 226021 Lucknow, Uttar Pradesh, India

Abstract: The development of any region depends greatly on the availability of appropriate water supplies. The quality of water can be judged based on a variety of parameters among which the most important is the temperature. In this study, Artificial Neural Network algorithms, Lavenberg Marquardt (LM) and Gradient Descent Adaptive (GDA) have been used to predict the quality of water. Using the data of temperature for the year 2008 to 12, researchers have measured Biochemical Oxygen Demand (BOD) and Dissolved Oxygen (DO) along River Ganga. Both the algorithms, mentioned above, have been compared for their performance. The results show that the algorithm LM gives a better performance as compared to that of GDA. Hence, simulated values for the desired locations at which measured data are unavailable can be efficiently provided by a trained ANN Model.

Key words: Artificial Neural Networks (ANN), Lavenberg Marquardt (LM), Gradient Descent Adaptive (GDA), River Ganga, water

INTRODUCTION

Rapidly increasing population, rising standards of living and exponential growth of industrialization and urbanization have exposed the water resources, in general and rivers, in particular, to various forms of degradation. The availability and the quality of the fresh water resources is the most pressing of the many environmental challenges on the national horizon in India. The stress on water resources is from multiple sources and the impacts can take diverse forms. The human actions in urban areas surrounding the Ganga River such as the Kanpur, Allahabad, Varanasi, etc. always generate the long term severe environmental impact in the form of critically alarming conditions.

The Ganga, the most sacred and worshipped river of the Hindus, is now one of the most polluted rivers of the country. Twenty five big cities located along its bank generated 1,340 mid sewage over 95% of the same entered the river without being treated prior to the Ganga Action Plan (GAP). Out of the total length of the river (2,525 km) for Gangotri to Gangasagar about 600 km long stretch is highly polluted. This pollution is due to the dumping of city garbage, industrial effluents, human and animal excreta, agricultural wastes, pesticides, burning of human bodies, community bathing and faulty social and religious practices.

To evaluate the present situation and to predict the effects of measures taken to improve river water quality, models are used. Before making a model of a river, data is gathered: hydro meteorological data, quality measurements in the river, land use, management practices on land, point pollution measurements, flow and water level data. People start making assumptions, extrapolations and make use of statistical relationships.

The targeting of the required pollutant load reductions and the finding of technical solutions for their implementation are the challenging key ingredients of the river basin planning and all the existing science, technology, mathematics and practical experience in this field will be needed to achieve compliance with the water quality standards with regard to chemical substances and ecological status. Prediction models are however considered useful for river basin management and are used to predict the behavior of water quality with respect to changes in pollutant loads and hydrological conditions. They are therefore used to evaluate target pollutant loads and management actions which will achieve compliance with water quality standards. The target pollutant loads are then used to set up regulatory rules and to plan waste water treatment plants, agricultural practices and general land use.

In this study, the measurement of BOD and DO has been calculated in order to evaluate the performance of River Ganga using Artificial Neural Network algorithm. The simulations have been carried out in MATLAB, for the monthly data, of previous 5 years (2008 to 2012), of temperature, flow rate, Biochemical Oxygen Demand

Corresponding Author: Aradhana Giri, Department of Civil Engineering, Institute of Engineering and Technology, Sitapur Road, 226021 Lucknow, Uttar Pradesh, India

(BOD) and Dissolve Oxygen (DO) of River Ganga. It can be seen that the results follow the same pattern as those of previous years data, infact they give more accurate values and so are highly reliable.

MATERIALS AND METHODS

Study area and water quality data: Ganga basin is the largest river basin in India in terms of catchment area, constituting 26% of the country's land mass (8,61,404 km^2) and supporting about 43% of its population (448.3 million as per 2001 census). The drainage area covered by Uttrakhand and Uttar Pradesh 294,364 km^2. The stretch from Kanpur to Allahabad has been taken in this study. The monthly data of Dissolved Oxygen (DO) and Biochemical Oxygen Demand (BOD), time period 2008 to 2012 was taken from the Uttar Pradesh Pollution Control Board (UPPCB), Pickup Bhawan, Gomti Nagar, Lucknow (UP).

Introduction of neural network: The concept of artificial neurons was first introduced in 1943 (McCulloch and Pitts, 1943) and applications of ANNs in research areas began with the introduction of the Back-Propagation Training (BP) algorithm for feedforward ANNs in 1986 (Rumelhart *et al.*, 1986). An Artificial Neural Network is an advanced technology for modelling the arrangements and working purpose of the brain. ANN being a great effort to simulate with particular hardware or software the information processing abilities of neurons associated in multiple layers. Each neuron get input from a different neuron or an external motivation, processes the input gesture using an activation or transfer function and form a transformed output signal. The output signal would be the last output from the network or the input to another neuron. The input to particular node is weighted sum of the inputs from all nodes to which it is associated. Numerous ANN architectures are being present but multilayer networks are commonly used for predicting (Zhang *et al.*, 1998; Maier and Dandy, 2000). An ANN gets used to learn the connection or plotting between input and outputs during the training development (Mas and Ahlfeld, 2007).

Division of data: In this project, 70% of the available data are used for training and remaining data is used for testing; the data is divided into the following ways: Input_train = 1×42 (input value is given), output_Train = 1×42 (output value), Input_validation = 1×18 (remaining input value), Output_validation = 1×18 (remaining output value).

Training: During the training development, these helpful aspects are steadily cut off. In this project, training and testing of ANN Model for the water quality parameters prediction is conducted by using neural network toolbox in the MATLAB. The MLP network has been trained by using the back propagation integrated with Levenberg-Marquardt algorithm. The tangent hyperbolic function has been used as activation function in the hidden layer neurons. The linear activation function has been used in the output layer neurons. The Gaussian radial basis function has been used as activation function in the hidden layer and the linear activation function is used in the output layer. The other important point is the selection of transfer functions. For network training two different transfer function were tested. These were tansig (Hyperbolic tangent sigmoid transfer function) and logsig (Log-sigmoid transfer function). Activation functions "Logsig" and "Tansig" (MATLAB ANN Tutorial).

Testing: When the network training is completed, the trained network performance must be tested. The testing data set have to be not used as a part during data sets training session. After testing the model with unknown data sets and then if there will be a big variation in the error attained after using the testing data set in comparison with the trained data set, it means that may both data sets are not belonging to identical population or the network is over fitted (Master, 1993). Deprived testing may occurred due to the network architecture, poor data reprocessing and rescaling of training and testing data sets. In this research, the network performance has been tested with different unknown data sets.

Simulation method: After training each network configuration, performance evaluation model for the artificial neural networks with different topologies and multiple regression to determine the optimal number of repetitions of the statistics, Correlation Coefficient®, Mean Square Error (MSE) and Root Mean Square Error (RMSE) was used:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

$$MSE = \frac{1}{n} \sum \left[\hat{z}(x_i) - z(x_i) \right]^2$$

$$RMSE = \sqrt{MSE}$$

the above formula, X_i and Y_i, ith real data and the estimates and and the average data, X_i and Y_i and value observed in the ith point, the estimated amount of the ith point, n the number of X_i and Y_i are examples.

RESULTS AND DISCUSSION

Temperature and DO model results: In this study, neural network, i.e., Levenberg-Marquardt back propogation (LM) and Gradient Descent with adaptive learning rate back propogation (GDA) were developed to predict dissolved oxygen along Ganga River. The Levenberg-Marquardt back propogation (LM) prediction results compared with gradient descent with Adaptive Learning Rate back propogation (GDA). Ganga River was found to be at lower levels in some location like upstream Kanpur, Down stream Kanpur and Down stream Allahabad. This is due to several factors that have influence in the dissolved oxygen contents on the surface water. These factors might be the use of oxygen for respiration by aquatic life, transport and mixing of oxygen within surface water and the incoming sewage flows (Table 1).

The results of LM and GDA from table show that LM it shows that the best trained R = 0.9860 whereas the best validation R is 0.9800. The best result of output versus target is 0.9919. From the results of LM network, it is observed that the MSE (Mean Square Error) performance range is between 0.087-0.8552 and for GDA, performance range is between 0.1508-0.7839. This could be because of the strong correlations between the network inputs and the output at this location as compared with the other locations. Whereas, in comparison to LM network the result of GDA is not more accurate and the process of GDA is slower than the LM Model. The methodology used for the development of MLP Prognostic Model may be utilized for other water quality variables along Ganga River. The comparison between LM and GDA neural networks was used to investigate the most proper method for predicting dissolved oxygen concentrations of the water quality in the Ganga River (Fig. 1).

Table 1: Results summary of LM and GDA neural networks models

Locations	Algorithm	Training R	Validation R	Testing R (output vs. target)	MSE
Upstream Kanpur	LM	0.9828	0.9800	0.9719	0.6772
	GDA	0.9803	0.9060	0.9251	0.7839
Downstream Kanpur	LM	0.9860	0.9917	0.9813	0.8552
	GDA	0.8836	0.9270	0.9666	0.7690
Upstream Allahabad	LM	0.9046	0.9633	0.9696	0.1085
	GDA	0.9561	0.9402	0.9439	0.1240
Downstream Allahabad	LM	0.9717	0.9515	0.9916	0.0877
	GDA	0.8977	0.8845	0.9646	0.1508

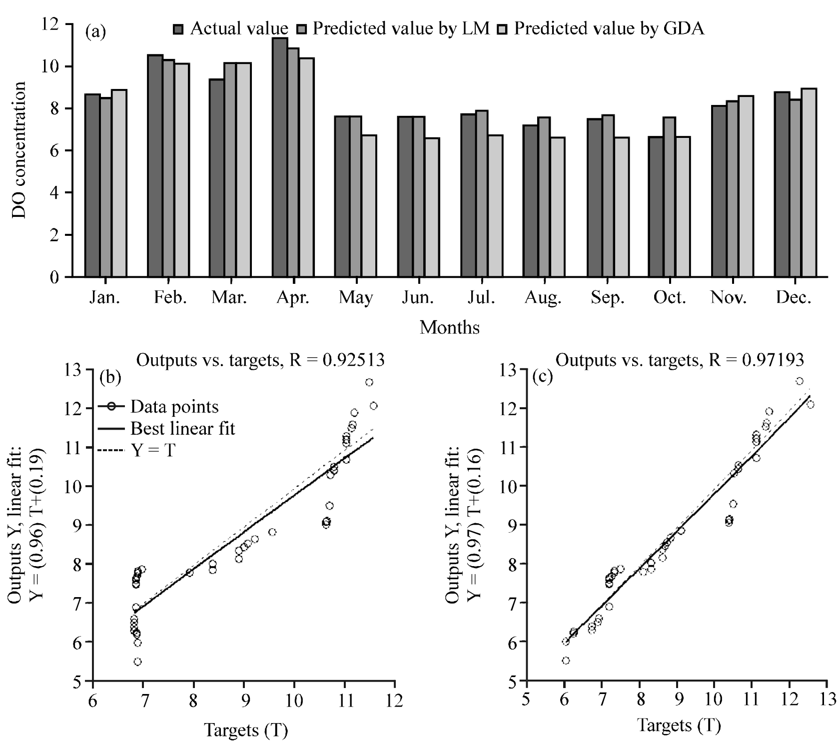

Fig. 1: a) UPS Kanpur DO Model testing resultl; b) Ganga DO-linear regression graph of GDA Neural Network Model and c) Ganga DO-linear regression graph of LM Neural Network Model

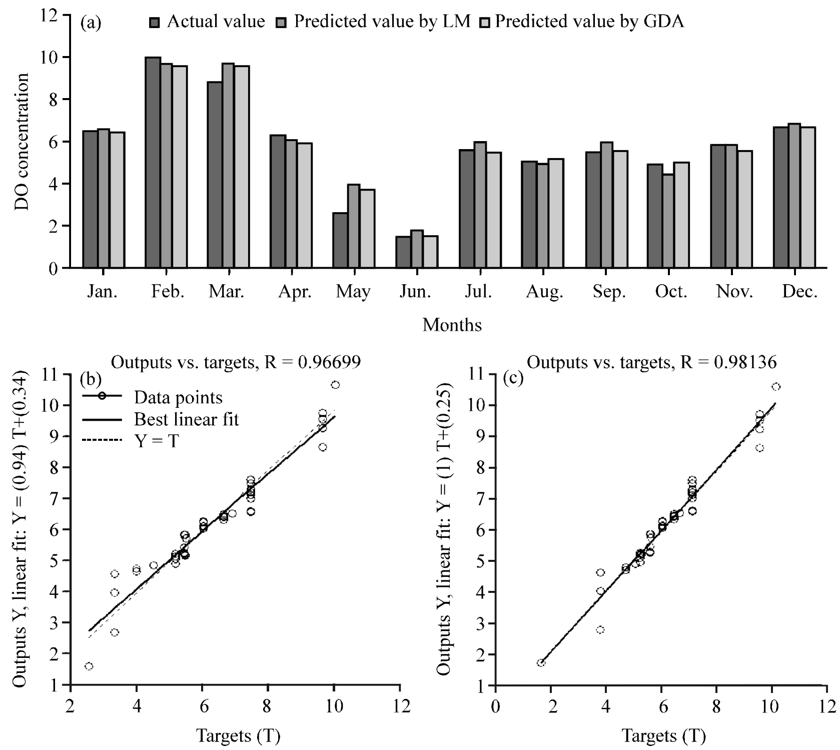

Fig. 2: a) DS Kanpur DO Model testing result; b) Ganga DO-linear regression graph of GDA Neural Network Model and c) Ganga DO-linear regression graph of LM Neural Network Model

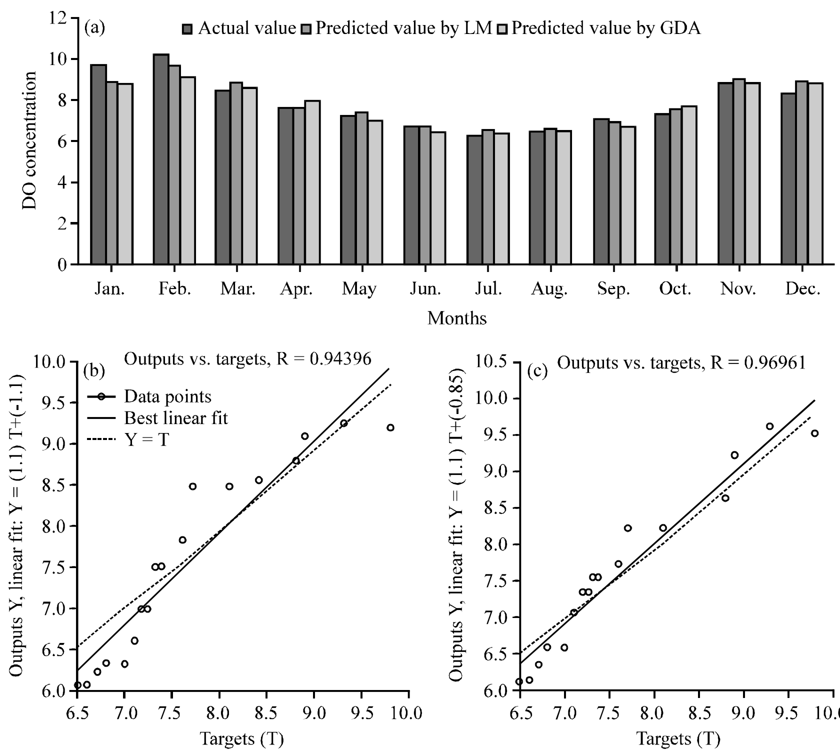

Fig. 3: a) UPS Allahabad DO Model testing result; b) Ganga DO-linear regression graph of GDA Neural Network Model and c) Ganga DO-linear regression graph of LM Neural Network Model

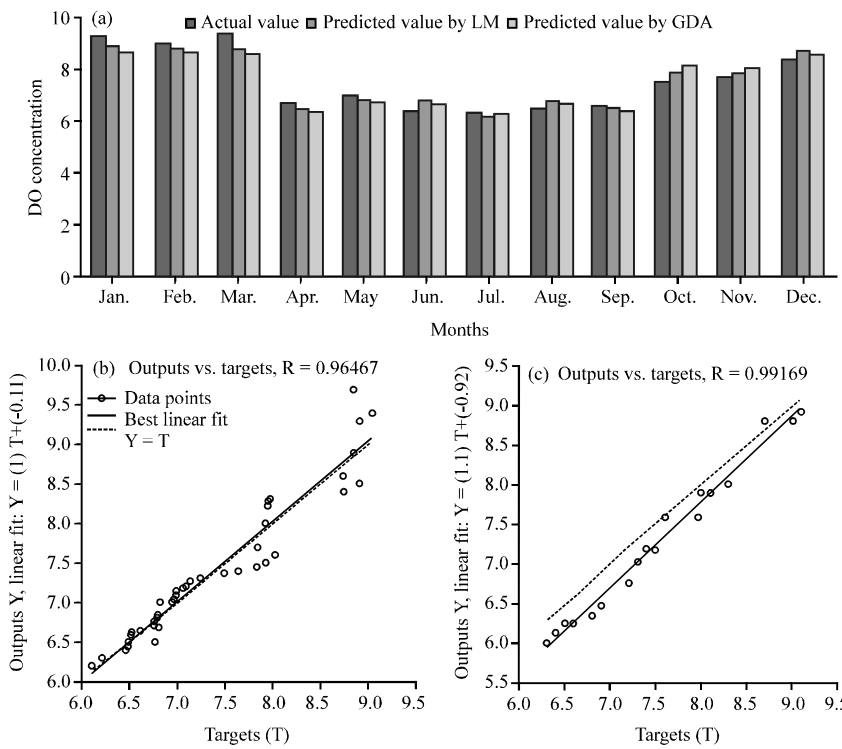

Fig. 4: a) DS Allahabad DO Model testing result; b) Ganga DO-linear regression graph of GDA Neural Network Model and c) Ganga DO-linear regression graph of LM Neural Network Model

Comparative analysis between actual and predicted values of dissolved oxygen with temperature of Ganga River by using artificial neural network models are shown in Fig. 1-4.

Effect of temperature on DO: Surface water temperature is one of the main significant parameters in the river surroundings, for the reason that nearly all the physicochemical and biological belongings are ruled and controlled by the temperature. Temperature restricts the diffusion values of solid substances and gases that are dissolved into river waters. Rate of the chemical reactions and biological actions including: BOD, rust photosynthesis, expansion and loss of organisms are all correlated with temperature. Temperature is negatively related to Dissolved Oxygen (DO). This difference in temperature might owe the quantity of untreated wastewater contribution and runoff. The water temperature is wide-ranging from 14.8-32°C. About 54% of temperature data were below 25°C, 74% below 28°C and 3% below 30°C. Temperature higher than 30°C was comparatively unusual and was recorded in summer. The temperature demonstrates ordinary seasonal differences by the higher temperatures in dry season and inferior temperatures in wet season.

Dissolved oxygen is the most important and most critical parameter, requiring continuous monitoring in intensive production systems. The saturation concentration of dissolved oxygen would be highest at low temperature and lowest at high temperatures. This condition is exactly the opposite of what fish require for basic metabolism and food conversion which is highest at high temperatures and lowest at low temperatures. Although, the air we breathe contains 21% oxygen, oxygen is only slightly soluble in water. As a result, aquatic species must spend a great deal of energy to remove the dissolved oxygen from water as compared to the energy that land dwelling species expend to obtain oxygen from the air. Oxygen solubility decreases as temperature and salinity increase. Both barometric pressure and altitude directly affect oxygen concentration.

Thermal regime influences aquatic organisms in terms of growth rate, metabolism, reproduction and life history, distribution, behaviour and tolerance to parasites/diseases and pollution (Alabaster and Lloyd, 1980; Crisp, 1996; Webb, 1996; Caissie, 2006). Most communities and species in freshwater ecosystems are cold-blooded and will therefore be sensitive to changes in the water temperature regime (Conlan *et al*., 2007). The effects of temperature change on the distribution,

Table 2: Results summary of LM and GDA neural networks models

Locations	Algorithm	Training R	Validation R	Testing R (output vs. target)	MSE
Upstream Kanpur	LM	0.9868	0.9497	0.9643	0.7821
	GDA	0.9633	0.9615	0.9459	0.8930
Downstream Kanpur	LM	0.9357	0.9774	0.9136	0.1356
	GDA	0.8437	0.9776	0.8630	0.3384
Upstream Allahabad	LM	0.8928	0.8803	0.9622	0.8756
	GDA	0.9411	0.9279	0.9782	0.6784
Downstream Allahabad	LM	0.9720	0.9394	0.9465	0.3465
	GDA	0.9368	0.9679	0.9496	0.4687

abundance and diversity, growth and reproduction of freshwater fishes have been particularly well documented. Davidson and Hazelwood (2005) predict that future temperature increases are likely to have significant effects on the growth rate of freshwater fish such as trout and salmon, in UK rivers. Similarly, Webb and Walsh (2004) have predicted that higher river temperatures as a result of climate change will be detrimental to the habitat of cold water fish species such as Atlantic salmon, Brown trout and Grayling.

Changes in water temperature are therefore linked to changes in water quality (e.g., dissolved oxygen concentrations and nitrogen levels). Statistical analysis of the effects of air temperature on river2 Science Report Climate change impacts and water temperature water quality have shown that biological oxygen demand and suspended solids increase and dissolved oxygen concentrations decrease in response to an increase in air temperature (Ozaki *et al*., 2003).

Temperature and BOD model results: From the results of LM network, it is observed that the MSE (Mean Square Error) performance range was between 0.1356-0.8930 and from GDA network, it is observed that the MSE (Mean Square Error) performance range was between 0.3384-0.8930. The results from Table 2 show that LM it shows that the best trained R = 0.9868 whereas the best validation R is 0.9776. The best result of output versus target is 0.9782. Biochemical Oxygen Demand concentrations are found to be at higher levels in some locations. This is due to several factors that have influence in the Biochemical Oxygen Demand (BOD) contents on the surface water. These factors might be the use of oxygen for respiration by aquatic life, transport and mixing of oxygen within seawater and the incoming sewage flows. The values of BOD ranged between 1.7- 8.4 mg L^{-1}. The mean values among all locations do not vary greatly among all locations while values range is found to be (5.04-5.60 mg L^{-1}). Approximately 27% of BOD values were <5.0, 91<7.0 mg L^{-1}. Concentrations of BOD >7.0 mg L^{-1} were about 9%.

Effect of temperature on BOD: Water temperature has a strong influence on the physical characteristics of streams and rivers such as surface tension, density and viscosity, solubility of gases and chemical reaction rates (Webb, 1996; Webb and Nobilis, 2007). Biological Oxygen Demand (BOD) is a measure of the oxygen used by microorganisms to decompose this waste. If there is a large quantity of organic waste in the water supply, there will also be a lot of bacteria present working to decompose this waste. In this case, the demand for oxygen will be high (due to all the bacteria) so the BOD level will be high. As the waste is consumed or dispersed through the water, BOD levels will begin to decline. The main factor that contributes to a high BOD is the presence of high level of organic matter or food in the waste water. Other organic matters that contribute to high BOD in a body of water are dead plants, leaves, grass clippings, manure and sewage. As mentioned above, when organic matter level in the water supply is high, the bacteria will begin the process of breaking down this waste. When this happens, much of the available dissolved oxygen is consumed by aerobic bacteria by increasing their metabolic activity which enhances temperature of river water.

The temperature of the water contributes to high BOD levels. Temperature controls the growth rates of phytoplankton, macrophytes and epiphytes, making freshwater ecosystems sensitive to rising temperatures (Whitehead and Hornberger, 1984; Wade *et al*., 2002). Water temperatures also regulate the behaviour of aquatic organisms such as fish migration and the timing of emergence and abundance of insect population at different life-cycle stages (Davidson and Hazelwood, 2005).

Warmer water usually will have a higher BOD level than colder water. As water temperature increases, the rate of photosynthesis by algae and other plant life in the water also increases. When this happens, plants grow faster and also die faster. When the plants die, they fall to the bottom where they are decomposed by bacteria. The bacteria require oxygen for this process so the BOD is high at this location. Therefore, increased water temperatures will speed up bacterial decomposition and result in higher BOD levels.

Comparative analysis between actual and predicted values of Biochemical Oxygen Demand (BOD) with temperature of Ganga River by using Artificial Neural Network Models are shown in Fig. 5-8.

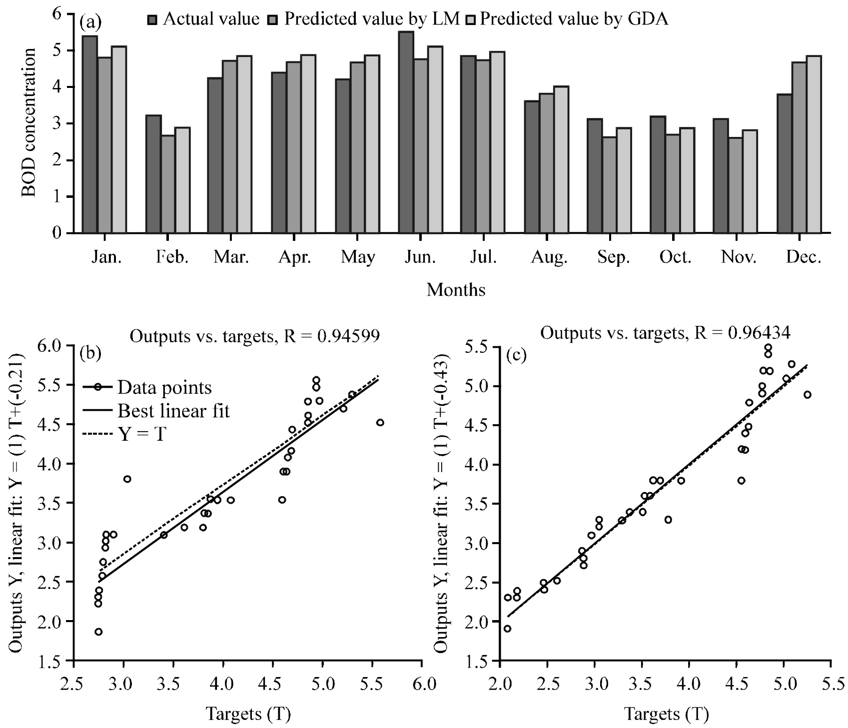

Fig. 5: a) UPS Kanpur BOD Model testing result; b) Ganga BOD-linear regression graph of GDA Neural Network Model and c) Ganga BOD-linear regression graph of LM Neural Network Model

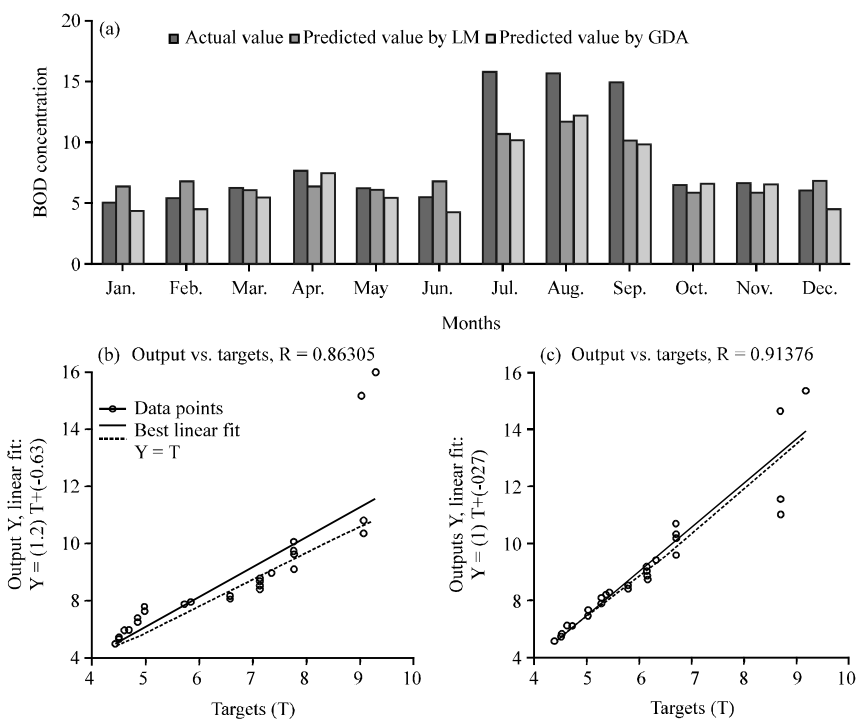

Fig. 6: a) DS Kanpur BOD Model testing result; b) Ganga BOD-linear regression graph of GDA Neural Network Model and c) Ganga BOD-linear regression graph of LM Neural Network Model

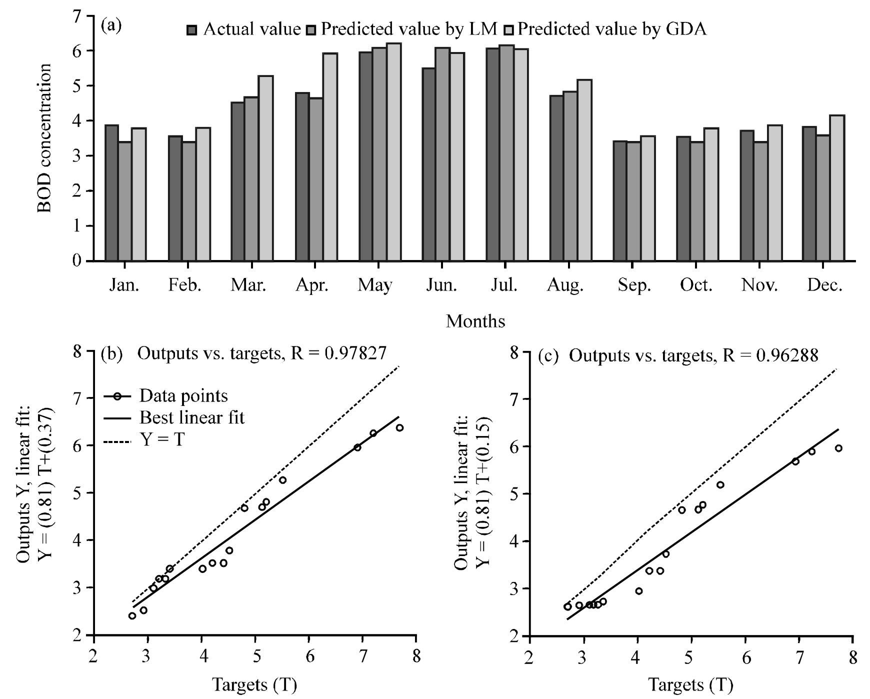

Fig. 7: a) UPS Allahabad BOD Model testing result; b) Ganga BOD-linear regression graph of GDA Neural Network Model and c) Ganga BOD-linear regression graph of LM Neural Network Model

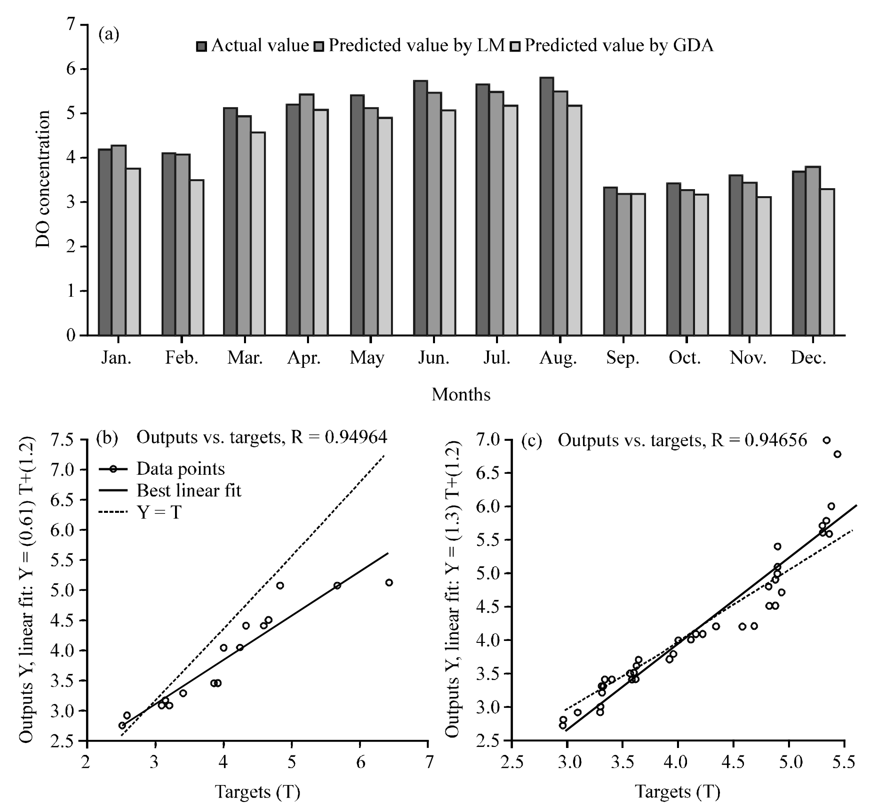

Fig. 8: a) DS Allahabad BOD Model testing result; b) Ganga BOD1-linear regression graph of GDA Neural Network Model; c) Ganga BOD3-linear regression graph of LM Neural Network Model

CONCLUSION

This study adopts the GDA and LM Neural Network algorithm to evaluate the water quality in Ganga River. Some conclusions can be got by the research: LM and GDA Neural Network algorithm can be used to evaluate the water quality. The result of evaluation through LM Neural Network algorithm has high precision. LM convergence is fast than GDA. The modelling result can provide reference to the water environment protection and plan kind of pagination.

It was observed that for the given data of DO and BOD, the predicted data from the regression equations are almost closer to the actual value. It is recommended that the ANN Model can be used to predict the BOD and DO based on available temperature and flow rate values. It is possible to represent the prediction with in an environment of an Optimal Neural Network Model and such a network presented with experimental data can learn the relationship between the functional parameters involved quite well when compared with Non-Optimal Models. Hence, with the proposed model applications it is possible to manage water quality parameters such as DO and BOD in a more cost-effective and easier way. Consequently, it has been demonstrated that DO and BOD in the River Ganga can be predicted with acceptable accuracy from a small set of physical and meteorological measurements. The chemical, physical and biological components of aquatic ecosystems are very complex and nonlinear. So, this result may be applied to automate DO and BOD estimations which utilized in water management and treatment systems corresponding to water management. Modeling of water quality variables is a very important aspect in the analysis of any aquatic systems.

REFERENCES

Alabaster, J.S. and R.L. Lloyd, 1980. Water Quality Criteria for Freshwater Fish. Butterworths, London, Pages: 297.

Caissie, D., 2006. The thermal regime of rivers: A review. Freshwater Biol., 51: 1389-1406.

Conlan, K., S. Lane, S. Ormerod and T. Wade, 2007. Preparing for climate change impacts on freshwater ecosystems (PRINCE): Results. Science Report to Environment Agency No SC030300/SR.

Crisp, D.T., 1996. Environmental requirements of common riverine European salmonid fish species in fresh water with particular reference to physical and chemical aspects. Hydrobiologia, 323: 201-221.

Davidson, I.C. and M.S. Hazelwood, 2005. Effect of climate change on salmon fisheries. Environment Agency Science Report W2-047/SR, Bristol, pp: 52.

Maier, H.R. and G.C. Dandy, 2000. Neural networks for the prediction and forecasting of water resources variables: A review of modeling issues and applications. Environ. Modeling Software, 15: 101-124.

Mas, D.M.L. and D.P. Ahlfeld, 2007. Comparing artificial neural networks and regression models for predicting faecal coliform concentrations. Hydrol. Sci. J., 52: 713-731.

Master, T., 1993. Practical Neural Network Recipes in C++. Academic Press, San Diego, California.

McCulloch, W.S. and W. Pitts, 1943. A logical calculus of the ideas immanent in nervous activity. Bull. Math. Biophys., 5: 115-133.

Ozaki, N., T. Fukushima, H. Harasawa, T. Kojiri, K. Kawashima and M. Ono, 2003. Statistical analyses on the effects of air temperature fluctuations on river water qualities. Hydrol. Processes, 17: 2837-2853.

Rumelhart, D.E., G.E. Hinton and R.J. Williams, 1986. Learning Internal Representations by Error Propagation. In: Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Rumelhart, D.E., J.L. McClelland and PDP Research Group (Eds.). MIT Press, Cambrige, UK., pp: 318-362.

Wade, A.J., P.G. Whitehead, G.M. Hornberger and D.L. Snook, 2002. On modelling the flow controls on macrophyte and epiphyte dynamics in a lowland permeable catchment: The River Kennet, Southern England. Sci. Total Environ., 282-283: 375-393.

Webb, B.W. and A.J. Walsh, 2004. Changing UK River Temperatures and their Impact on Fish Populations. In: Hydrology: Science and Practice for the 21st Century, Webb, B. (Ed.). Vol. 2, British Hydrological Society, USA., pp: 177-191.

Webb, B.W. and F. Nobilis, 2007. Long-term changes in river temperature and the influence of climatic and hydrological factors. Hydrol. Sci. J., 52: 74-85.

Webb, B.W., 1996. Trends in stream and river temperature. Hydrol. Processes, 10: 205-226.

Whitehead, P.G. and G.M. Hornberger, 1984. Modelling algal behaviour in the river Thames. Water Res., 18: 945-953.

Zhang, G., B.E. Patuwo and M.Y. Hu, 1998. Forecasting with artificial neural networks: The state of the art. Int. J. Forecast., 14: 35-62.