RESEARCH PAPER International Journal of Agricultural Engineering | Volume 12 | Issue 1 | April, 2019 | 10-17
ISSN–0974–2662 ■ Visit us : *www.researchjournal.co.in* ■ **DOI: 10.15740/HAS/IJAE/12.1/10-17**

Prediction of TDS in groundwater by using BP-NN modeling

Debaditya Gupta and Alivia Chowdhury

Received : 21.01.2019; **Revised :** 02.02.2019; **Accepted :** 19.02.2019

See end of the Paper for authors' affiliation

Correspondence to :

Debaditya Gupta
Department of Agricultural and Food Engineering, Indian Institute of Technology,
Kharagpur (W.B.) India
Email : debadityagupta24@gmail.com

■ **ABSTRACT :** Total dissolved solids (TDS) comprise inorganic salts (principally calcium, magnesium, potassium, sodium, bicarbonates, chlorides, and sulfates) and some small amounts of organic matter that are dissolved in water. TDS in drinking-water originate from natural sources, sewage, urban run-off, industrial wastewater, and chemicals used in the water treatment process, and the nature of the piping or hardware used to convey the water. The present study deals with the prediction of TDS in Nadia district, West Bengal using back propagation neural network approach with gradient descent training method and the performance evaluation was done using RMSE, NSE, IOA, MAE and R^2. It is found that the best result was obtained by M-6-10-1 (Input-Hidden-Output). The effectiveness of total hardness, chloride, potassium is also explained by the result of this study.

■ **KEY WORDS :** Back propagation neural network, TDS, Gradient descent method, Hidden nodes, Correlation matrix, Matlab

■ **HOW TO CITE THIS PAPER :** Gupta, Debaditya and Chowdhury, Alivia (2019). Prediction of TDS in groundwater by using BP-NN modeling. *Internat. J. Agric. Engg.*, **12**(1) : 10-17, **DOI: 10.15740/HAS/IJAE/12.1/10-17.**

Total dissolved solids (TDS) is a measure of the dissolved combined content of all inorganic and organic substances present in a liquid in molecular, ionized or micro-granular (colloidal sol) suspended form. Water quality is one of the most important factors contributing to a healthy life. From the water quality management point of view, TDS is the most important factor and many water developing plans have been implemented in recognition of this factor (Salmani and Jajaei, 2016).Anthropogenic activities can alter the relative contributions of the natural causes of variations and also introduce the effects of pollution (Whittemore *et al.,* 1989). India which is agriculture based country where groundwater is an important source of agricultural as well as domestic water supply. Indiscriminate and unplanned use of this natural resource has put a threat on the sustainability of this major freshwater reserve in various parts of our country.It is very difficult to express water quality in terms of TDS by the help of any single parameter due to the no linearity of nature as the water quality function of the number of different dissolved inorganic particles *viz.*, calcium, magnesium, sodium and potassium, chloride, sulfate, carbonate and bicarbonate and silica. These major constituents constitute the bulk of the mineral matter contributing to TDS. TDS is directly associated with sodium absorption ratio, the salinity of water and drinking water quality (Asghari *et al*., 2006 and Mehrdadi *et al.,* 2012). Artificial Neural network approach hasbeen successfully applied in many cases of groundwater studies in order to overcome the nonlinear behaviour of the dependent variable over the independent variable. In recent years, ANN models have been successfully employed to the water quality studies, ANNs can efficiently comprehend and model the nonlinear relationship among the different variables which affect water quality (Lek *et al*.,1996).Variation of groundwater quality was shown using ANN(Kuo *et al*., 2004) in the area of blackfoot disease. A neural network approach was done for the nitrate prediction of groundwater (Yesilnacar *et al.*, 2008) usingLevenberg-Marquardt algorithm with 25 hidden neurons and groundwater salinity forecasting (Banerjee *et al.*, 2011). ANN is used to predict water parameters froma few known parameters (Kumar *et al*., 2010). An attempt was made to predict the TDS parameter with the neural network in Fajr Purification Center in the south of Iran in 2012 (Mehrdadi *et al.*, 2012) with successful results. Similar kind modeling was done by Zare *et al.* (2011) to predict nitrate parameter have been successful. A groundwater quality assessment study (Kheradpisheha *et al.*, 2015) using ANN in was carried by Bahabad plain, Yazd, Iran and it was concluded that ANN could predict accurately the various groundwater parameters. Application of Neuro-Fuzzy system is also now-a-days applied for the assessment of groundwater quality (Khaki *et al*., 2014). The present study deals with the modeling of TDS using back propagation neural network approach for efficient prediction of this major groundwater quality parameter for Nadia district, West Bengal.

METHODOLOGY

The present study area in Nadia district, West Bengal (22º53'-24º11'North) and (88º09' and 88º48'East) having a geographical area of 390027 km^2 and consists of 17 blocks. The average winter (January and February) temperature of 14°C and the average summer (April and May) temperature of 35° C. The mean annual rainfall is 1500 mm, precipitating more than 80 per cent during June through September.Groundwater quality data on seven water quality parameters namely chloride, specific conductivity, TDS, total hardness, bicarbonate, potassium and sodium for the year of 2011 and Geographical coordinates of sampling location of 166 sites of Nadia district has been collected from Soil and Water Investigation Directorate, Kolkata.

Neural network modeling:

A neural network is a massively parallel distributed processor made up of simple processingunits that has a natural propensity for storing experiential knowledge and making it available for use) (Haykin, 1994). The procedure used to perform the learning process is called a learning algorithm. The basic structure of an artificial neural network is illustrated in Fig. A.

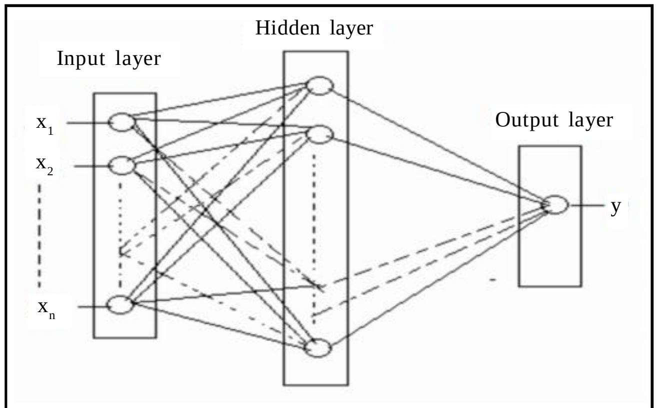


Fig. A : Basic structure of an artificial neural network

In the present study six post-monsoon groundwater quality parameters *viz.*, specific conductivity, bicarbonate, chloride, total hardness, sodium and potassium were used in different combinations as inputs to train and test multi-layered perceptron (MLP) ANN models with one input layer, one hidden and one output layer in order to simulate the TDS. The training algorithm was back propagation (Back Propagation Neural Network, BP-NN). The steps are enlisted below in Fig. B.

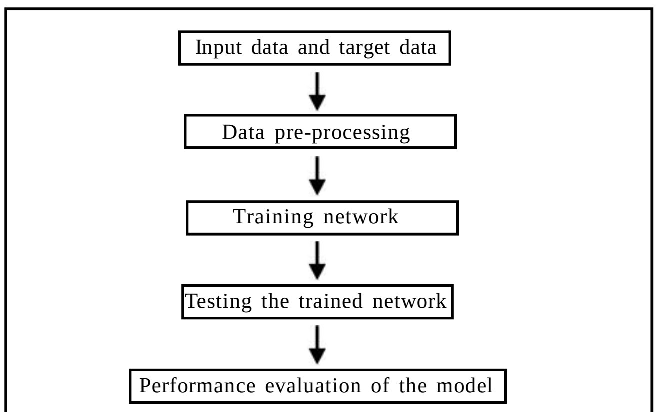


Fig. B : Flow chart of the ANN modeling process

The data preprocessing involves normalization of the input and output data between 0 and 1.

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

70 per cent of the dataset was used for training and

remaining 30 per cent dataset were used for testing purpose. The activation function selected for the hidden layer is logistic sigmoid ('logsig') which is given by the following equation and linear function is set to the output layer.

$$\mathbf{f\,(x)} = \frac{\mathbf{1}}{\mathbf{1+e^{-x}}}$$

The set of basic equation for the BP-NN model are given below.

$$\mathbf{h_i = \sum w_i x_i + b_i}$$

$$\mathbf{h_o = f\,(h_i)}$$

$$\mathbf{y_{out} = \sum\ v_o\, h_o + b_o}$$

$$\mathbf{y = f\,(y_{out})}$$

where,

$\mathbf{x_i, h_i, h_o, w_i, v_o, b_i, b_o, y_{out}, y}$ are inputs of the input layer, inputs of the hidden layer, outputs of the hidden layer, weights between the input and hidden layer, weights between the hidden and output layer, bias at input-hidden layer, bias at hidden-output layer, input of the output layer and final output, respectively.

The update of the weights depends upon the error between the actual and the desired output. Finally, the weights are updated in the training phase using back propagation method. In this study, correlation matrix was performed to understand the relative importance between the dependent variables and the set of independent variables. From correlation plot, good (r = 0.8 to 1), medium (r = 0.5 to 0.8) and poor (r < 0.5) correlation between selected variables was found out and accordingly the different input parameters for training different ANN models were selected. The training algorithm for the back-propagation neural network select is "Gradient Descent" ('traigd') algorithm and the with MATLAB 2013a software package.

Hidden node selection:

The number of hidden nodes is calculated using the trial and error method by specifying the range of hidden neurons between 1 to$2^n/3$ (Arai,1993), 2n-1(Rajasekaran and Pai, 2004) and for sigmoidal activation function 2n+1 (Hecht-Nielsen,1987) is suggested, where n is the number of nodes in the input layer.

Model performance evaluation:

Model performance was evaluated by the root mean square error (RMSE), mean absolute error (MAE), co-efficient of determination (R^2), nash sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) and index of agreement (IOA) (Willmott, 1981).

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{pi}-y_{ai}\right)^2}$$

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{ai}-y_{pi}\right|$$

$$\mathbf{R^2} = \left\{\frac{\sum_{i=1}^{n}\left[\left(y_{pi}-\bar{y}_p\right)\left(y_{ai}-\bar{y}_a\right)\right]}{\sqrt{\left[\sum_{i=1}^{n}\left(y_{pi}-\bar{y}_p\right)^2\sum_{i=1}^{n}\left(y_{ai}-\bar{y}_a\right)^2\right]}}\right\}^2$$

where,

$0 \le R^2 \le 1$, where R^2 nearer to 1 stands for better prediction.

$$\mathbf{NSE} = 1-\frac{\sum_{i=1}^{n}\left(y_{ai}-y_{pi}\right)^2}{\sum_{i=1}^{n}\left(y_{ai}-\bar{y}_a\right)^2}$$

where,

$-\infty < \mathbf{NSE} \le 1$, where NSE 1 stands for perfect fit and $-\infty$ means worst fit and NSE =1 stands for perthefect fit.

$$\mathbf{IOA} = 1-\frac{\sum_{i=1}^{n}\left(y_{pi}-y_{ai}\right)^2}{\sum_{i=1}^{n}\left(\left|y_{pi}-\bar{y}_a\right|+\left|y_{ai}-\bar{y}_a\right|\right)^2}$$

where,

$0 \le \mathbf{IOA} \le 1$, where IOA close to 1 stands for a better degree of agreement between the observed and predicted data.

where, $n, y_{ai}, y_{pi}, \bar{y}_a, \bar{y}_p$ are total number of data points, actual TDS data, predicted TDS data, actual mean and predicted mean, respectively.

RESULTS AND DISCUSSION

This part discusses the selection of different architecture of ANN models to estimate the total dissolved solids (TDS) for the study area. Six groundwater quality inputs namely specific conductivity, bicarbonate, chloride, total hardness, sodium, and potassium for the 166 well locations for the post-monsoon period were used in different combinations to the different ANN models trained and tested for the study and post-monsoon TDS values are selected as output for the study area. In order to understand the relative importance and dependence of different input variables on TDS, a correlation matrix between different inputs and output was constructed and the correlation coefficient values are given in Table 1 and the summary of the different ANN- architecture is given in Table 2. The line plot and the scatter diagram between actual and ANN predicted TDS are shown in Fig. 1. (a)-(f) and Fig. 2. (a)-(f), respectively.

Comparison of different ANN structures:

The actual vs. predicted TDS is shown for different BP-ANN structures are shown in Fig. 1. (a)-(f)

The scatter plot (with 1:1 line) between actual vs. predicted TDS for different BP-ANN structures are shown in Fig. 2. (a)-(f).

The ANN structure M-6-10-1(6 nodes in the input layer, 10 nodes in the hidden layer and one node in output layer) (Table 2) gives highest R^2 (0.99), NSE (0.8), IOA (0.95) and lowest RMSE (39.02) and MAE (37.56). The lowest R^2 (0.57), NSE (0.37), IOA (0.84) and highest RMSE (68.21) and MAE (59.00). Fromthe model results and the correlation plot,it is clear that specific conductivity plays an important role in predicting total dissolved solid in the study area. The study also gives an idea of successful prediction of TDS using back propagation neural network (BP-NN) for the study area with the proper selection of input variables.

Conclusion:

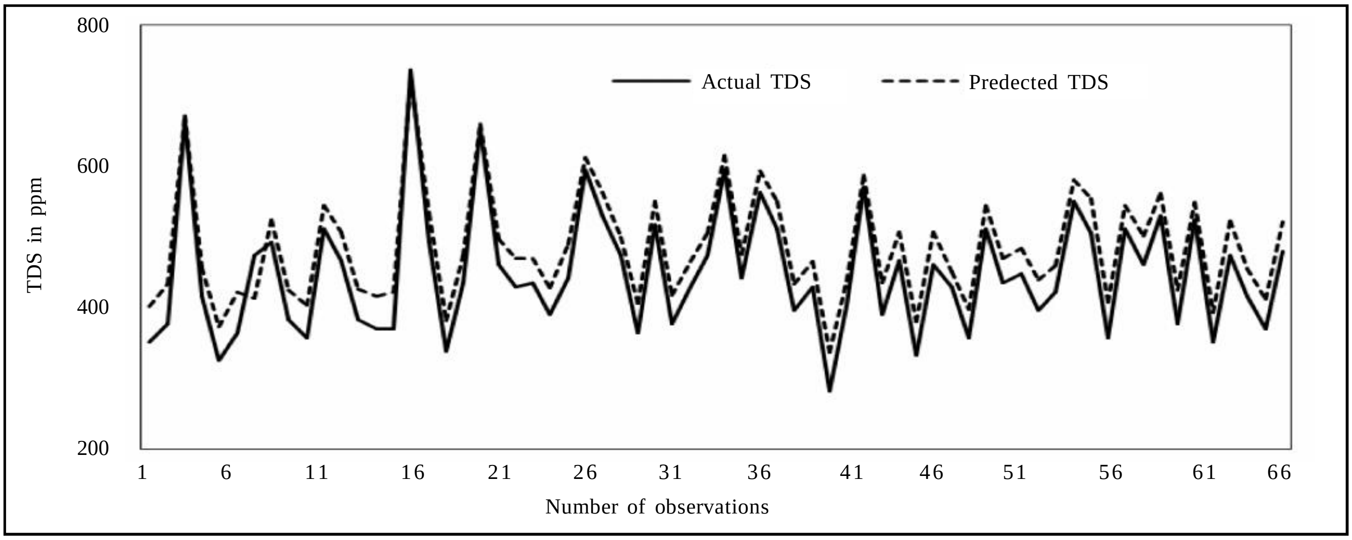
Total dissolved solids (TDS) plays one of the most important water quality parameter for both agricultural as well as domestic water. In order to predict the TDS concentration in groundwater, a back propagation neural network (BP-NN) was carried out using Matlab 2013a software package with six different groundwater quality parameters as input variables. The results shows that the effectiveness of specific conductivity in predicting TDS in the study area. It can be concluded that TDS can be successfully predicted by BP-NN with only few water quality parameters as input which can further be improved by taking more number of water quality parameters into account as input and this study can also help in groundwater management in the study area more

Table 1: Correlation matrix between the TDS and the input variables

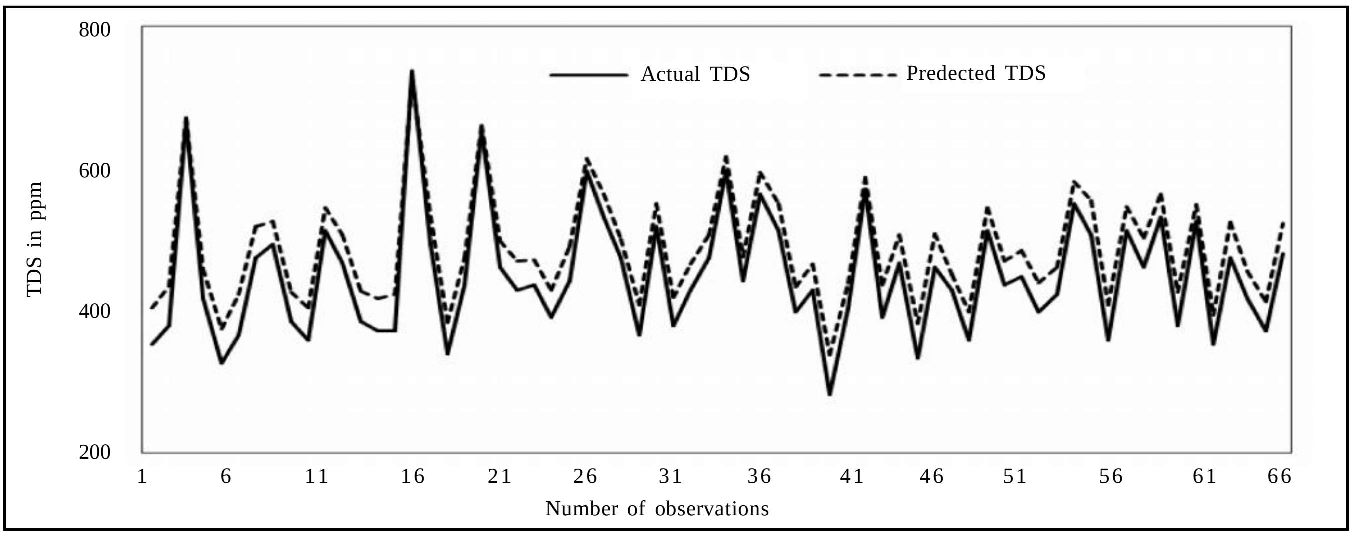
	Ph	Sp. conductivity	Bicarbonate	T.H	Chloride	Total iron	Total as	TDS
Ph	1							
Sp. conductivity	-0.231	1						
Bicarbonate, ppm	0.012	0.216	1					
T.H, ppm	-0.066	0.774	0.168	1	Symmetric matrix			
Chloride, ppm	-0.299	0.911	0.262	0.797	1			
Total Iron, ppm	-0.045	0.124	0.131	0.138	0.122	1		
Total As, ppm	-0.001	0.092	0.183	0.133	0.089	0.274	1	
TDS, ppm	-0.176	0.907	0.251	0.707	0.83	0.098	0.076	1

Table 2 : Summary of the model statistics

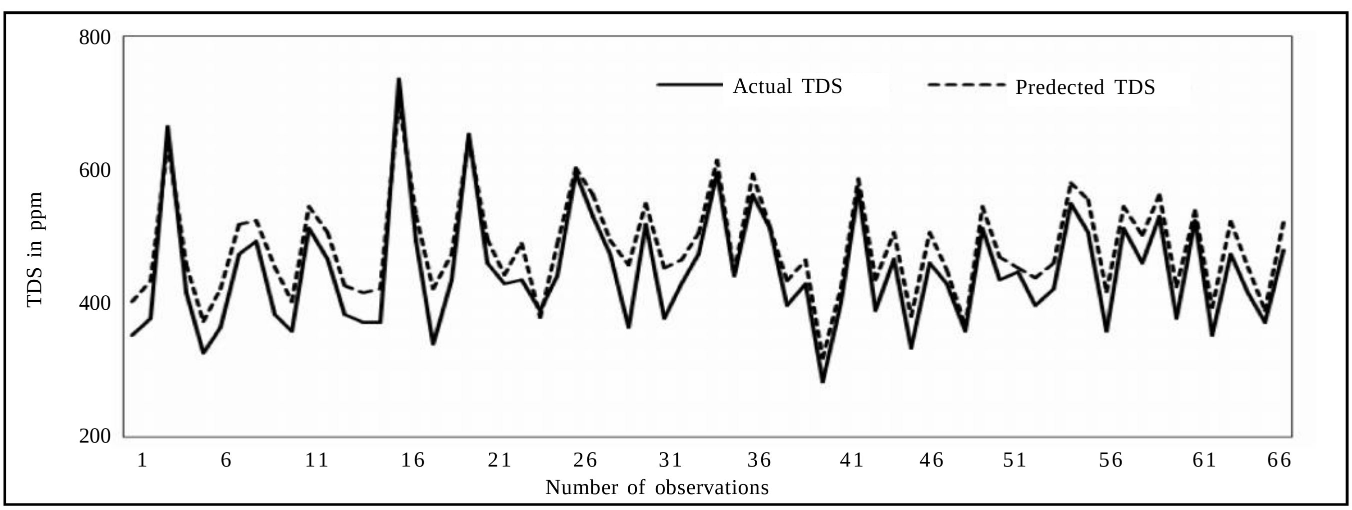
ANN structure	Input variables	RMSE	MAE	R^2	NSE	IOA
M-3-8-1	Sp. conductivity, chloride,TH	39.35	37.79	0.97	0.79	0.94
M-6-10-1	Sp.conductivity, chloride,TH,sodium, bicarbonate,potassium	39.02	37.56	0.99	0.80	0.95
M-4-7-1	Sp.conductivity, chloride,TH, sodium	41.03	36.96	0.95	0.77	0.94
M-5-10-1	Sp.conductivity, chloride,TH, sodium, bicarbonate	46.84	39.90	0.84	0.71	0.92
M-4-6-1	Chloride,sodium,bicarbonate, potassium	68.21	59.00	0.57	0.37	0.84
M-5-8-1	Chloride, sodium, bicarbonate,TH, potassium	62.98	55.60	0.67	0.47	0.84



(a)

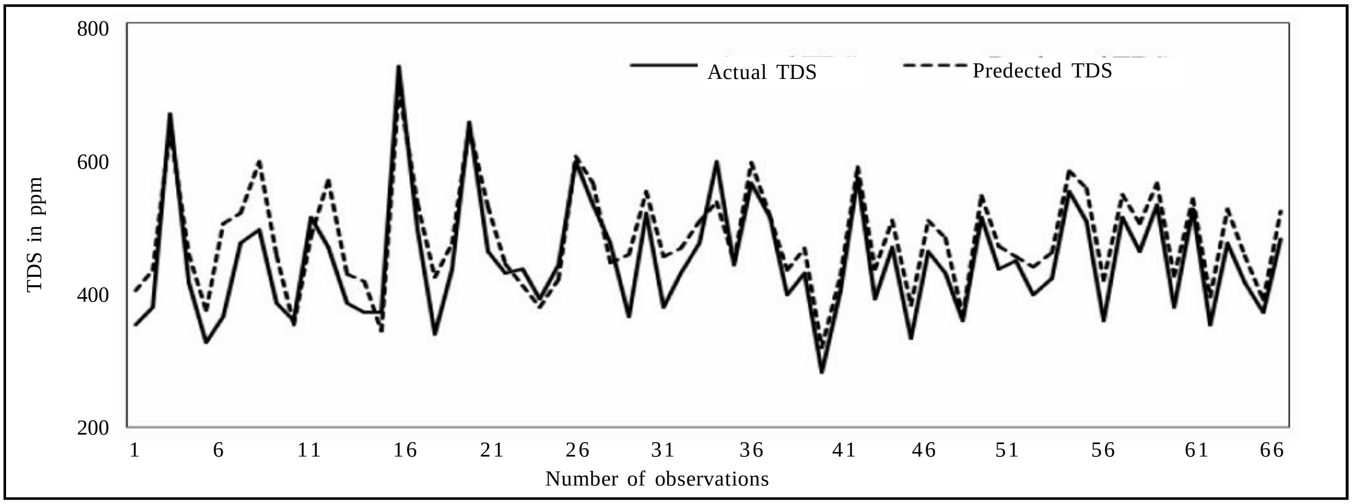


(b)

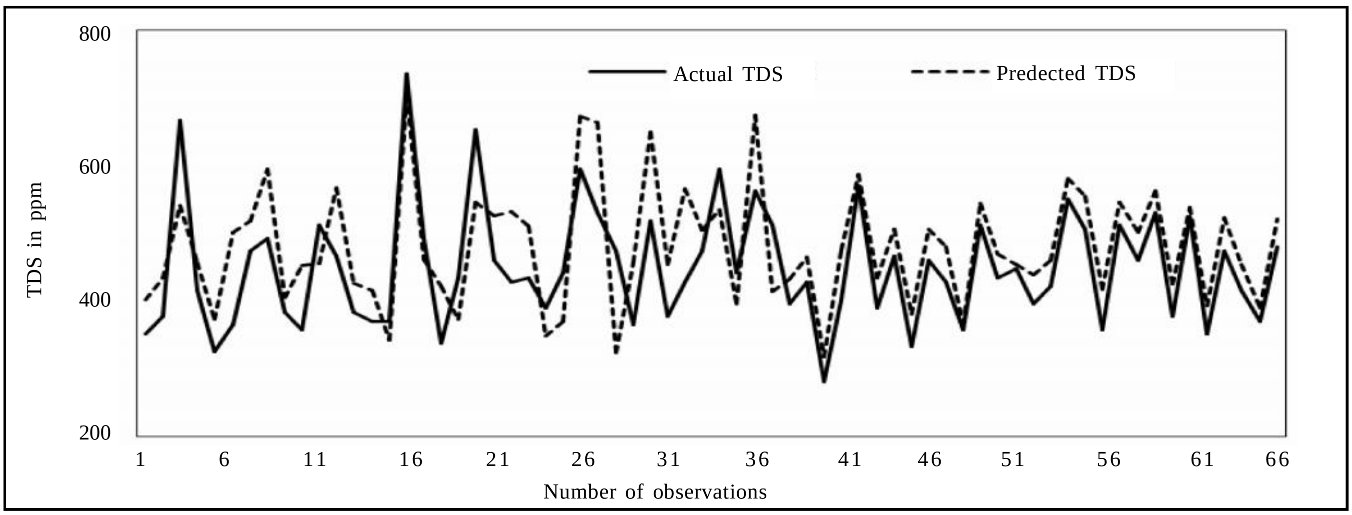


(c)

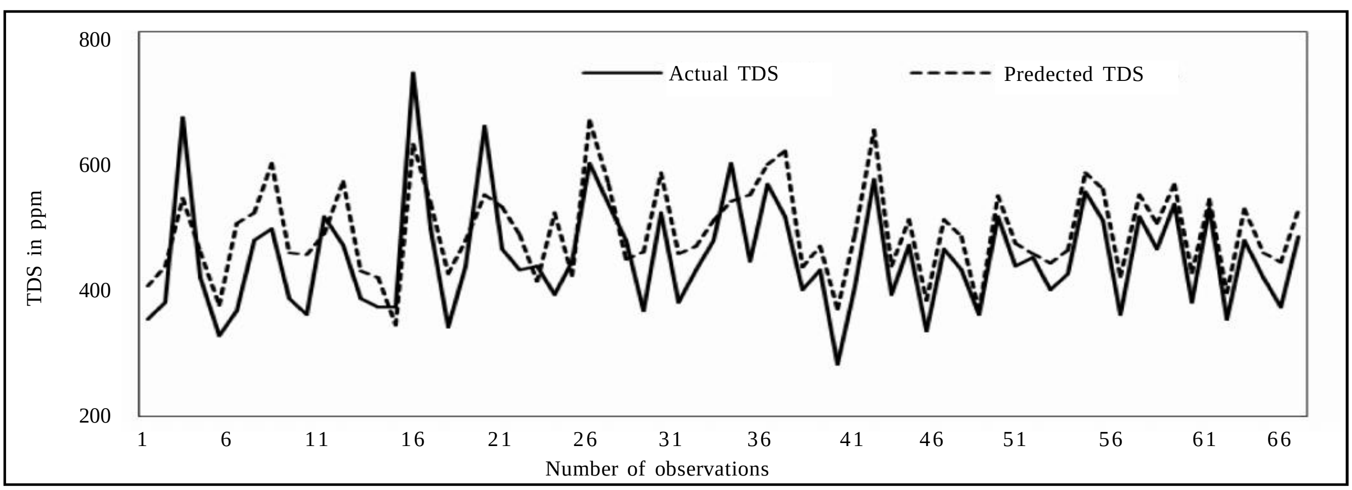
Fig. 1: Contd...........



(d)



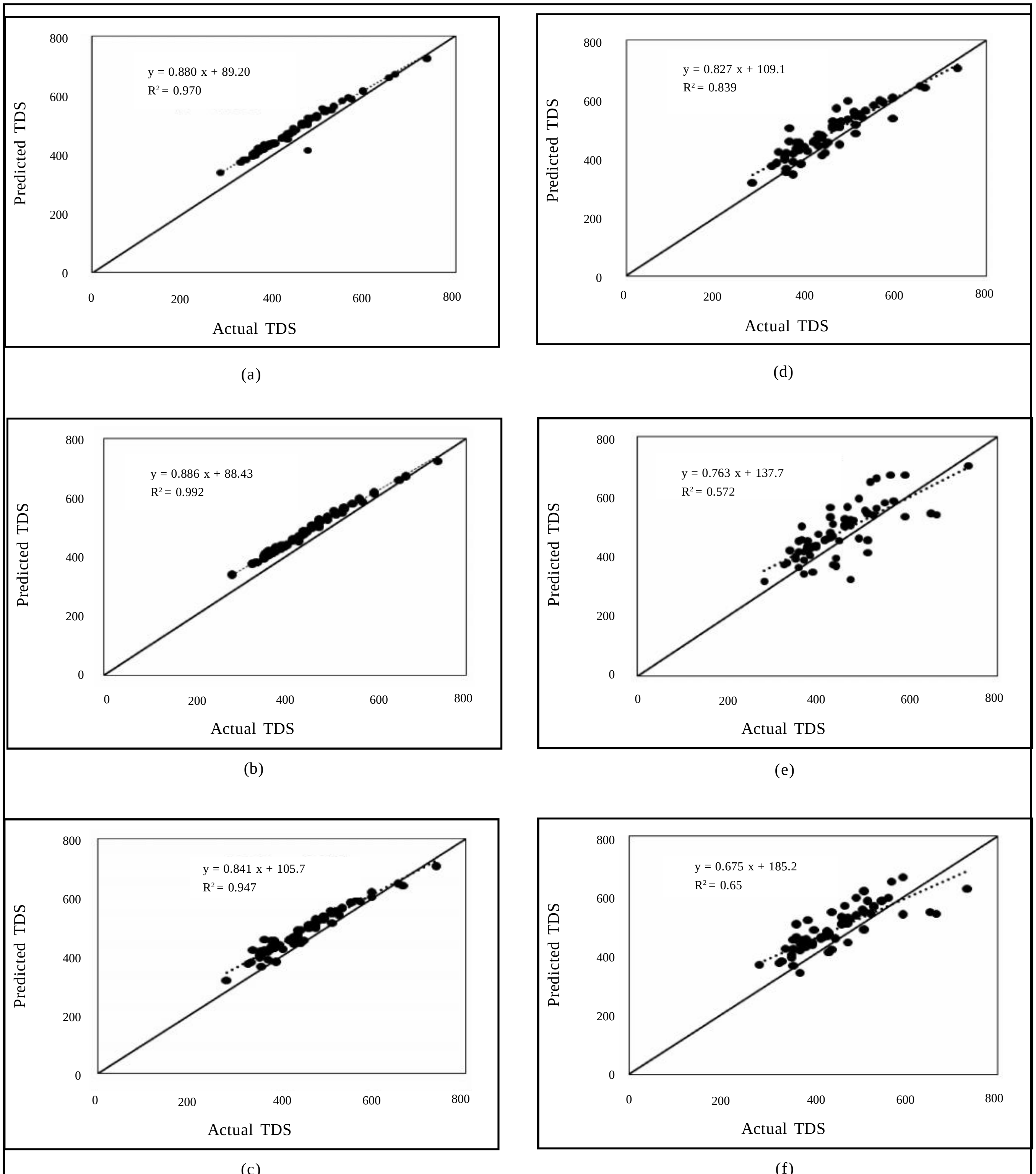
(e)



(f)

(a) M-3-8-1 (b) M-6-10-1 (c) M-4-7-1 (d) M-5-10-1 (e) M-4-6-1 (f) M-5-8-1

Fig. 1 : Actual vs Predicted TDS data for the ANN outputs



(a) M-3-8-1 (b) M-6-10-1 (c) M-4-7-1 (d) M-5-10-1 (e) M-4-6-1 (f) M-5-8-1

Fig. 2 : Scatter plot between actual vs. predicted TDS

systematically and efficiently.

Acknowledgment:

Authors are grateful the State Water Investigation Directorate (SWID), Kolkata for providing the necessary data for the completion of the study.

Authors' affiliations:
Alivia Chowdhury, Department of Agricultural and Food Engineering, Indian Institute of Technology, **Kharagpur (W.B.) India**

REFERENCES

Arai, M. (1993). Bounds on the number of hidden units in binary-valued three layered neuralnetworks. *Neural Networks*, **6** : 855-860.

Asghari, M. A., Nadiri, A. and Fijani, E. (2006). Ability to study different models of artificial neural networks to evaluate groundwater level in the hard formation. Tenth Conference of Geological Society, Teheran.

Banerjee, P., Singh, V. S., Chattopadhyay, K., Chandra, P. C., Bhoop, **Singh (2011).** Artificial Neural network model as a potential alternative for groundwater salinity forecasting. *J. Hydrology, Elsevier*, **398** (3-4): 212-220.

Haykin, S. (1994). *Neural networks : A comprehensive foundation*. Macmillan, New York, U.S.A.

Hecht-Nielsen, R. (1987). Kolmogorov's mapping neural network existence theorem. IEEE first International Conference on Neural Networks, **3** : 11-13.

Khaki, M., Yousoff, I. and Islami, N. (2014). Application of neural network and neuro-fuzzy system for assessment of groundwater quality. C*LEAN Soil Air Water, WileyOnline Library*, **43** (4) : 551-560.

Kheradpisheha, I. Z., Talebib, A., Rafatia, L., Ghaneeiana, M.T. and Ehrampousha, M.H. (2015). Groundwater quality assessment using artificial neural network: A case study of Bahabadplain, *Yazd, Iran Z.*, **20** (1): 65-71.

Kumar, V., Methew, N. S. and Swaminathan, G. (2010). Analysis of groundwater from potability from Tiruchirappalli city using back propagation ANN Model and GIS. *J. Environ. Protec.*, I : 136-142.

Kuo, Y. M., Liu C. W. and Lin, K. H. (2004). Evaluation of the ability of anartificial neural network model to access the variation of groundwater quality in an area of blackfoot disease in Taiwan. *Water Res. Elsevier.*, **38** (1): 148-158.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S. (1996). Application of neural networks to modeling non-linear relationships in ecology. *Ecological Modeling*, **90** : 39-52.

Mehrdadi, N., Hasanlou, H., Jafarzadeh, M.T., Hasanlou, H. and Abodolabadi, H. (2012). Simulation of low TDS and biological units of Fajr industrial wastewater treatment plant usingartificial neural network and principal component analysis hybrid method. *J. Waterresource & Protection*, **4**: 370-376.

Nash, J. E. and Sutcliffe, J.V. (1970). River flow through conceptual models part I – Adiscussion principles. Index of Agreement. *J. Hydrology, Elsevier,* **10** (3): 282-290.

Rajasekaran, S. and Pai, G. A. V. (2004). *Neural networks, fuzzy logic and geneticalgorithms- Synthesis and applications*, 3rd Ed. Prentice-Hall of India Private Limited.

Salmani, M. H. and Jajaei, E. S. (2016). Forecasting models for flow and total dissolved solids in Karoun river-Iran. *J. Hydrology, Elsevier,* **535**: 148-159.

Whittemore, D. O., McGergor, K. M. and Marotz, G. A. (1989). Effects of variations in recharge on groundwater quality. *J. Hydrology, Elsevier,* **106**(1-2): 131-145.

Willmott C. J. (1981). On the validation of models. Physical Geography, *Taylor &Francis Online,* **2** : 184-194.

Yesilnacar, M. I., Sahinkaya, E., Naz, M. and Ozkaya, B. **(2008).** Neural network prediction of nitrate in groundwater of Harran Plain, Turkey. *Environmental Geology,* **56**(1): 19-25.

Zare, A.H., Bayat, V. M. and Daneshkare, A. P. (2011). Forecasting nitrate concentration in groundwater using artificial neural network and linear regression models. *International Agro physics*, **25** (2) : 187 – 192.