International Journal of Engineering Research
Volume No.5, Issue No.2, pp : 149- 152
ISSN:2319-6890)(online),2347-5013(print)
1 Feb. 2016

Prediction of SAR for Groundwater along the Kham River in Aurangabad District, Maharashtra using ANN

Kazi F.S., Shinde S.D. and Sadgir P.A.
Govt. College of Engineering, Aurangabad, Maharashtra, India
Email: farhankazi555@gmail.com

Abstract: Monitoring groundwater quality is important as the aquifers are vulnerable to contamination due to point sources and non-point sources. This study presents an artificial neural network model for predicting Sodium Adsorption Ratio (SAR) values for well samples. The data for 3 years from 27 wells to the left bank and 27 wells to the right bank of the Kham River in Aurangabad district, India is used for this investigation. A Correlation analysis is performed to select the input parameters that show a strong relation between Sodium and SAR (0.8412). Electrical Conductivity, Sodium, Magnesium and Calcium are used as input parameters in the prediction model. The Levenberg–Marquardt algorithm is selected as the best out of the 12 Back-propagation algorithms and optimal neuron number is determined as 10 for the model. The model tracked the experimental data closely giving a correlation coefficient of R=0.9380.The results obtained from the model shows that Artificial Neural Network could be used as an applied tool for the prediction of irrigation water parameters, especially, SAR.

Keywords: Artificial Neural Network (ANN), Correlation, MAE, RMSE, Sodium Adsorption Ratio (SAR)

I. Introduction

Sub-surface water is an important source used in rural areas for irrigation purpose. A lot of parameters contribute in variation in groundwater quality. The SAR constitutes one of the fundamental parameters as regards water used for agriculture. In general, higher the SAR, less suitable is the water for irrigation. Use of irrigation water with high SAR for years can cause a decrease in the ability of the soil to form a stable aggregate and a loss of soil structure. High sodium ions in water can also affect the permeability of the soil and is the cause of infiltration problems. When sodium exists in the soil in an exchangeable form, it replaces the calcium and magnesium absorbed on the soil clays and causes dispersion of soil particles. The SAR has a proper criterion for irrigation water suitability. If calcium and magnesium are the predominant cations absorbed on the soil exchange complex, the soil tends to be easily cultivated and has a permeable and granular structure (Asadollahfardi *et al*. 2010). The irrigation water quality rating for SAR is given in table 1 (IS 11624 – 1986).

Table 1: Water Quality Rating based on SAR

Serial Number	Class	SAR Range (millimole/litre)
i)	Low	Below 10
ii)	Medium	10 – 18
iii)	High	18 – 26
iv)	Very high	Above 26

ANNs are loosely based on the neural structure of the brain which provides the ability to learn from the input data and then apply this to unknown data, in effect they can generalize and associate unknown data (Samson *et al.* 2010).The ANNs are capable of imitating the basic characteristics of the human brain such as self-organization, self-adaptability, and error tolerant and have been widely adopted for model identification, analysis and forecast, system recognition and design optimization (Niu *et al.* 2006, Sharma 2016).ANNs, especially back propagation network are closely related to statistical methods and are most suitable for predicting ground water quality applications (Sarala *et al*. 2014).

Asadollahfardi *et al.* (2013) predicted SAR values in Chelghazy River in Kurdistan, Iran using ANN. Input parameters were pH, discharge, sulfate, sodium, calcium, chloride, magnesium and carbonate. Ahanger *et al.* (2013) used Levenberg-Marquardt (trainlm) algorithm to train ANN for prediction of Manganese concentration in Chahnimeh1 reservoir, Iran. They selected the number of neurons in the hidden layer through trial-and-error method. Sarala *et al.* (2014) predicted pH, TDS and SAR of the irrigation water in Batlagundu town, Tamil Nadu using the ANN. They tested the performance of the ANN model by using correlation coefficient and Mean Square Error (MSE). Heydari *et al.*(2013) used ANN to derive and develop monthly models for prediction of monthly values of DO and SC for the Delaware River in Pennsylvania, USA. They evaluated the performance of the model by statistical criteria which are correlation coefficient (r), Root Mean Square Error (RMSE) and Mean Absolute Error (MAE).Mageshkumar *et al.*(2012) attempted to model the TDS concentration of a 40 kilometer stretch of Cauvery River.

Sarda and Sadgir (2015) developed two different ANN networks to estimate COD using various combinations of monthly input parameters. Shinde *et al.*(2015) concluded that the distribution of concentrations of various parameters due to seepage from a surface river water in basaltic shallow and deep aquifer may be evaluated by ANN.

The Study focuses on finding the ANN Model for the calculation of SAR and comparing it with the conventional method.

II. Study Area

In Maharashtra, Aurangabad is one of the famous developing city and which is well known for its Industrial auto cluster. It is situated in the central part of Maharashtra. The summer temperature is max 43º C and Min. 28ºC and winter temperature Max 32ºC to 5ºC. The sources of irrigation are streams, percolation tanks and wells in study area. Ground water plays a major role for irrigation as well as domestic uses. The

International Journal of Engineering Research
Volume No.5, Issue No.2, pp : 149- 152
ISSN:2319-6890)(online),2347-5013(print)
1 Feb. 2016

Study area covers the Aurangabad taluka and Gangapur taluka which lies between latitude 19º 53’ north and longitude 75º 20’ east along Kham River. The most important economic activity in the rural area is agriculture, with main crops being jawar, wheat, and maize, fodder crops for dairy animal and vegetable crops like onion, cauliflower, chili, tomato, and cucumber. Kham River, which is one of the major tributaries of the Godavari River, receives all domestic and industrial waste water from the Aurangabad city and MIDC waluj, which includes six and three nallahs respectively. Six Streams from Aurangabad city, which are Barudnagar Stream, Khadkeshwar Stream, Aushadhi Bhavan Stream, Stream behind Aurangabad Municipal Corporation, Gandhi Nagar Stream and Stream from Satara nagar parishad from urban area and three Stream from MIDC-Waluj, which are Tisgaon Stream, Stream near Oasis chowk, Stream beside Bajajnagar from waluj and ranjanagon industrial area as shown in location map (Figure 1). This river ultimately confluences with the Godavari upstream of Jaikwadi dam.

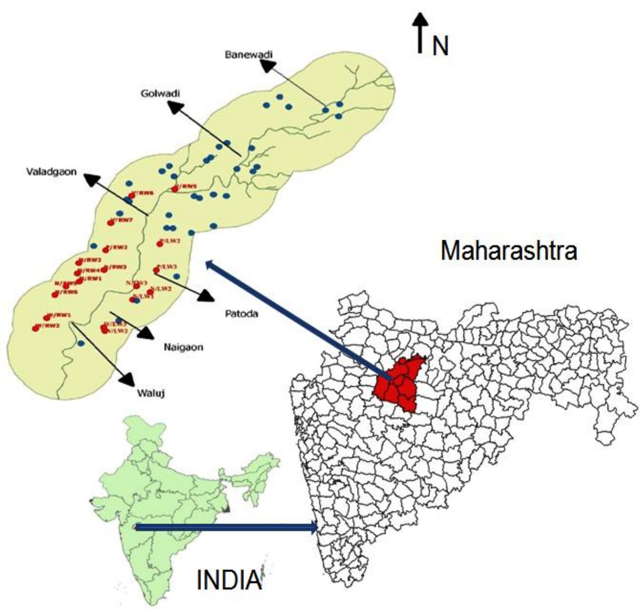

Figure 1: Detailed Map of the Study Area

Data for the study was obtained from the 54 wells along the Kham River – 27 wells along the left bank and 27 wells along the right bank of the river’s flow.

Correlation Analysis:

Correlation analysis is performed for the selection of the most suitable input parameters. It checks the correlation between any two parameters. The formula for correlation analysis is given as:

$$r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2 \sum_{i=1}^{n}(Y_i-\bar{Y})^2}} \quad .... \ (1)$$

Where,

Xi = observed value, Yi = predicted value

$\bar{X}$= mean of observed value, $\bar{Y}$ = mean of predicted value

An ‘r’ value near +1 indicates strong positive correlation, ‘0’ value indicates no relation and -1 indicates strong negative correlation.

Table 2: Pearson’s correlation coefficient:

	EC	Ca	Mg	Na	SAR
EC	1				
Ca	0.49614	1			
Mg	-0.1012	0.36267	1		
Na	0.33437	0.21451	-0.0288	1	
SAR	0.59360	0.18185	-0.2289	0.84119	1

From the above table, it is observed that Sodium values have a considerable effect on SAR values (0.8412). EC has a moderate correlation with SAR (0.5936), Calcium is considerably less correlated with SAR while Magnesium has a negative correlation.

Artificial Neural Network:-

The ANN is a data processing system, based on an idea similar to the processing of the human brain that treats data as a steady network parallel to each other in order to solve a problem. With the networks, the structure of data is designed to help programming knowledge in which the behavior is the same as natural neural and its component (Asadollahfardi *et al*. 2013).Neural networks, as used in artificial intelligence, have traditionally been viewed as simplified models of neural processing in the human brain. It is accepted by most scientists that the human brain is a type of computer. The origins of neural networks are based on efforts to model information processing in biological systems, which may rely largely on parallel processing as well as implicit instructions based on recognition of patterns of “sensory” input from external sources.

An artificial neural network is developed in this study using the experimental data. Finally, the neural network model is applied to collection of available data about the various water quality parameters of the well sources along the Kham River in Aurangabad. All the computations were performed using the Microsoft EXCEL 2013 and MATLAB (MATLAB® software R2011a).

Back Propagation Network:

The back propagation algorithm has made it possible to design multi-layer neural networks for numerous applications, such as adaptive control, classification of sonar targets, stock market prediction and speech recognition. Also, BPNN has the advantage of fast response and high learning accuracy. Hence an ANN with back propagation algorithm (BP) has been adopted here to model the potability behavior of well water along the River. One of the advantages of using the neural network approach is that a model can be constructed very easily based on the given input and output and trained to accurately predict process dynamics. This technique is especially valuable in processes where a complete understanding of the physical mechanisms is very difficult, or even impossible to acquire. Neural network is a logical structure with multi-processing elements, which are connected through interconnection weights.

The knowledge is presented by the interconnection weights, which are adjusted during the learning phase. There are several algorithms available among which the Levenberg-Marquardt algorithm (trainlm) has the fastest convergence. The trainlm is able to obtain lower mean square errors than any of the other algorithms. This BP network is a multi-layer of the

 ISSN:2319-6890)(online),2347-5013(print)
1 Feb. 2016

network architecture including the input layer, hidden layer(s) and output layer. Layers include several processing units known as neurons. They are connected with each other by variable weights to be determined. In the network, the input layer receives information from external source and passes this information to the network for processing. The hidden layer receives from the input layer, and does all information processing. The output layer receives processed information from the network, and sends the results to an external receptor (Rao 2013). The algorithm for the back propagation program is described below with the help of flow diagram as shown in Figure 2.

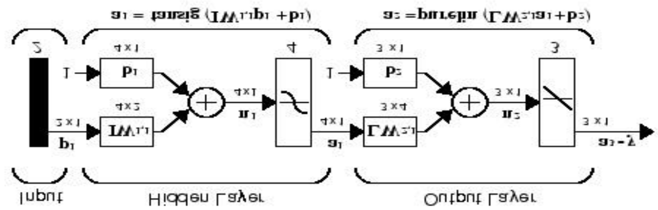

Figure2: Model for Feed Forward BP Network (Source:- MATLAB user guide)

Model Performance:

In this study, 12 BP algorithms were compared to select the best fitting one. For all algorithms, a two-layer network with a tan-sigmoid transfer function within the hidden layer and a linear transfer function within the output layer was used. In the selection of BP algorithm, the number of neurons was kept constant at 6.The performance of the BP algorithms was evaluated with the root-mean square error (MSE) and determination coefficient (R) between the measured data set and the modeled output. The best BP algorithm with minimum training error and maximum R was the Levenberg–Marquardt (trainlm) algorithm for SAR values.

Table 3: Comparison of back-propagation algorithms for predicting SAR in monitoring wells

Back-propagation algorithms	**R Value**	**MSE**	**Iteration Number**
Trainlm	0.9713	1.263	13
Traincgp	0.8926	2.2039	13
Traingd	0.6054	10.626	8
Traingda	0.6805	1.8345	47
Traingdx	0.6608	8.7161	9
Trainrp	0.9436	2.6671	28
Trainscg	0.9239	2.5353	16
Trainoss	0.9026	3.0084	34
Traincgf	0.9234	2.1983	11
Trainbfg	0.00715	69.6745	1
Traingdm	0.5238	47.8401	10
Traincgb	0.9012	2.6951	36

After selecting the best BP algorithm, Levenberg–Marquardt (trainlm) algorithm, the number of neurons was optimized keeping all other parameters constant (Table 4).

Table 4: Variation in R value with change in number of neurons

ANN	Architecture	Training	Validation	Test	All
Structure					
Output_input		R	R	R	R
S_ENMCa	4-2-1	0.8621	0.8904	0.8484	0.863
S_ENMCa	4-4-1	0.9728	0.9752	0.8979	0.9624
S_ENMCa	4-6-1	0.9367	0.922	0.8937	0.9424
S_ENMCa	4-8-1	0.9642	0.8715	0.8808	0.9352
S_ENMCa	4-10-1	0.912	0.8569	0.938	0.9069
S_ENMCa	4-12-1	0.9742	0.9522	0.8955	0.9589
S_ENMCa	4-14-1	0.9948	0.943	0.896	0.9689
S_ENMCa	4-16-1	0.3034	0.4332	0.4816	0.3504
S_ENMCa	4-18-1	0.9466	0.9484	0.9158	0.9422
S_ENMCa	4-20-1	0.9721	0.8306	0.7708	0.924

Note:Where, S = SAR, E = EC, N = Sodium, M = Magnesium and Ca = Calcium

It is observed from the above table that, an architecture using 10neurons in the hidden layer (4-10-1) gives the optimum Test R value (0.938) for prediction of SAR.So, the modeling was carried out using Levenberg–Marquardt (trainlm) algorithm with 10 neurons. Figure 3 shows the schematic diagram of the neural network structure of the model.

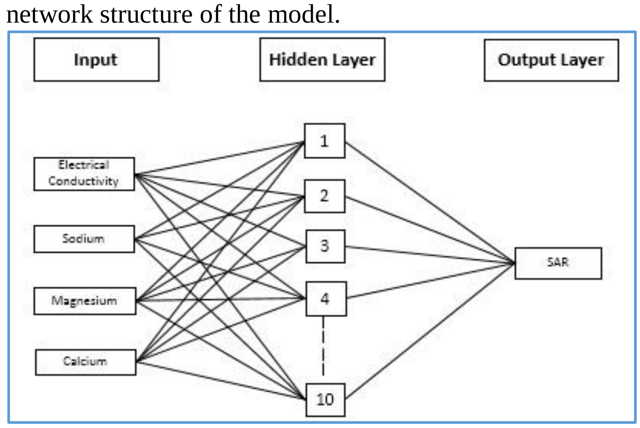

Figure 3: Neural network structure for the prediction of SAR

For calculation of the amount of error in predicting the desired parameter and performance evaluation of models, correlation coefficient R^2, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used, as shown in equations below-

$$\boldsymbol{R^2} = \mathbf{1} - \frac{\sum_{i=1}^{n}(X_i - Y_i)^2}{\sum_{i=1}^{n}(X_i - \overline{X_i})^2} \quad \text{....... (2)}$$

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - Y_i)^2} \quad \text{........ (3)}$$

$$\mathbf{MAE} = \frac{1}{n}\sum_{i=1}^{n}|X_i - Y_i| \quad \text{....... (4)}$$

Where, X_i and Y_i are the measured and predicted values respectively, $\overline{X}$ is the average of the measured data and n is the total number of values in the dataset.

III. Results and Discussion

It was crucial to determine the number of neurons in the hidden layer. Neurons play an important role that have an effect on the general characteristics of network and training time

ISSN:2319-6890)(online),2347-5013(print)
1 Feb. 2016

(Ahanger *et al.* 2013). The optimum number of neurons in a hidden layer was found by trial and error.

The algorithm used to train ANN in this study was selected as Levenberg–Marquardt back propagation (trainlm). The trainlm is an approximation to the Newton's method(Hagan *et al.*1994). This is very well suited to the training of the neural network. Figure 4 shows the sequential plot and scatter plots of the ANN model predicted versus actual values.

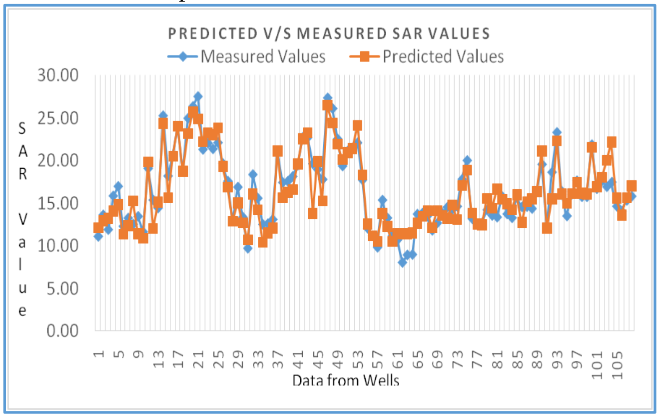

Fig 4 a) Sequential Plot for the Predicted v/s Measured SAR values

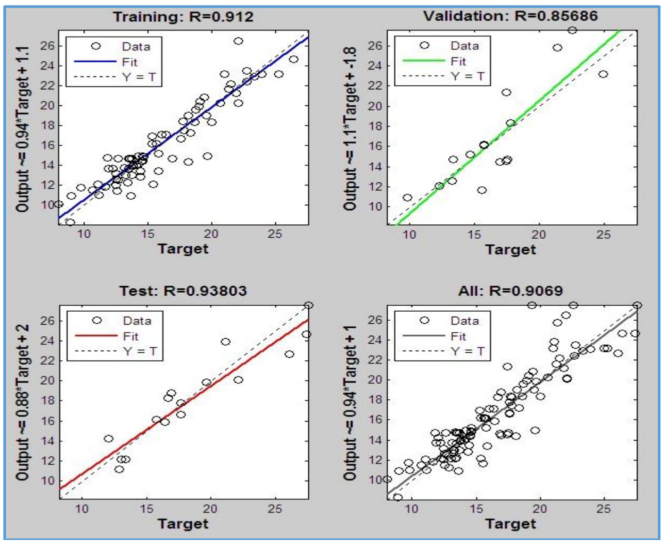

Fig 4 b) Scatter Plot for Predicted v/s Measured SAR values

The coefficient of correlation (R), Root Mean Square Value (RMSE) and MAE values for the model are given in the table below:

Table 5: Analytical results based on observed and predicted values

ANN Architecture	Performance of Model			
	RMSE	**MAE**	$\mathbf{R^2_{calculated}}$	$\mathbf{R_{calculated}}$
4-10-1	1.89	1.354	0.839	0.916

The value of Test R in the model and R value found using mathematical formula (Equation 2.2) are similar. The RMSE and MAE values were 1.89 and 1.354 (from equations 3.3 and 3.4) respectively.

IV. Conclusion

In this paper, ANN was used to predict the SAR values for the wells along the Kham River, Aurangabad, India. The identified model was applied to the SAR values measured from 2013–2015. By correlation analysis it was observed that Sodium has a strong relationship with SAR. Levenberg-Marquardt (trainlm) algorithm was selected as the best among all the BP algorithms. A network architecture consisting of 4 neurons in the input, 10 neurons in the hidden layer and 1 output neuron was found to be the most optimum neuron structure. Correlation coefficient value of 0.93803 proves the accuracy in prediction of the model. RMSE and MAE values were 1.89 and 1.54 respectively which indicate minimum error between predicted and measured dataset. The results of this study show that ANN can be used to model well data in an easy way.

References

i. *Ahangar A.G., Soltani J. and Abdolmaleki A.S. (2013), "Predicting Mn concentration in water reservoir using Artificial neural network (Chahnimeh1 reservoir, Iran)", International Journal of Agriculture and Crop Sciences, Vol. 6 (20), pp 1413-1420.*

ii. *Asadollahfardi G., Hemati A., Moradinejad S. and Asadollahfardi R. (2013), "Sodium Adsorption Ratio (SAR) Prediction of the Chalghazi River Using Artificial Neural Network (ANN) Iran", Current World Environment, Vol. 8 (2), pp 169-178.*

iii. *Hagan M. T. and Menhaj M. B. (1994), "Training Feed-forward Networks with the Marquardt Algorithm", IEEE Transactions, Vol. 5 (6), pp 989-993.*

iv. *Bureau of Indian Standards (1986), "IS: 11624-1986, Quality of Irrigation Water", Drinking water sectional committee UDC 631:671.03: 626.810 (026), India, pp.*

v. *Mageshkumar. P., Pradeep. T., StalinJohn M.R., AmalRaj. S. and Anandakumar. S. (2012), "Neural Network Modelling of TDS Concentrations in Cauvery River Water, Tamilnadu, India", International Journal of Applied Sciences and Engineering Research, Vol. 1 (6) pp.739-746.*

vi. *Heydari M., Olyaie E., Hamid Md. and Kisi O. (2013), "Development of a Neural Network Technique for Prediction of Water Quality Parameters in the Delaware River, Pennsylvania", Middle-East Journal of Scientific Research, Vol. 13 (10), pp 1367-1376.*

vii. *Niu Z., Zhang H. and Liu H. (2006), "Application of neural network to prediction of coastal water quality", Journal of Tianjin Polytechnic University, Vol.25, pp 89-92.*

viii. *Samson M., Swaminathan G. and Venkat K.N. (2010), "Assessing Groundwater Quality For Potability using A Fuzzy Logic And GIS – A Case Study of Tiruchirappalli City – India", Computer Modelling and New Technologies, Vol. 14 (2), pp 58–68.*

ix. *Sarala T.D. and Mageswari U.T.S.R (2014), "Modeling of Irrigation Water Quality using MultilayerPerceptron Back Propagation Neural Network (MLBP-NN)", International Journal of Chemical Technology Research, CODEN (USA),Vol. 6 (5), pp 3053-3061.*

x. *Sarda P.R. and Sadgir P.A. (2015), "Reservoir Water Quality Modeling for COD using Artificial Neural Network", International Journal of Engineering and Management Research, Vol. 5 (4).*

xi. *Sharma S.N. and Regulwar D.G. (2016), "Prediction of Evapotranspiration by Artificial Neural Network and Conventional Methods", International Journal of Engineering Research IJER, Vol. 5(1), pp.184-187.*

xii. *Shinde S.D., Patil K.A. and Sadgir P.A. (2015), "River and Ground Water Quality Assessment using GIS", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 4 (11).*