Amutha, R.; Sundarambal, P.; Porchelvan, P.

Ann Modelling of Groundwater Quality Index

Proceedings of ninth International Conference on Hydro-Science and Engineering (ICHE 2010), IIT Madras, Chennai, India.
2 - 5 August 2010

ANN MODELLING OF GROUNDWATER QUALITY INDEX

R. Amutha[1], P. Sundarambal[2] and P. Porchelvan[3]

***Abstract**: Artificial neural network (ANN) is machine-based computational technique. The technique is particularly suited to problems that involve the manipulation of multiple parameters and nonlinear interpolation, and consequently are not easily amenable to conventional theoretical and mathematical approaches. This paper presents the use of ANN as a viable means of modeling groundwater quality index. The chemical analyses of most common major constituents, Gibbs plot, Piper's diagram and Water Quality Index (WQI) method, is used to assess the general hydro-geochemical characteristics of groundwater and to evaluate its suitability for drinking purposes. ANN simplifies and speeds up the computation of the WQI, as compared to the currently used existing methods. By optimizing the calculation, a significant saving of money and time can be achieved. ANN models with different learning approaches, such as back propagation neural networks (multi-layered perceptron, ward networks) and polynomial nets (Group Method of Data Handling (GMDH)), are considered and adopted to model the WQI. ANN predicted WQI was compared to that conventionally computed and they were in close agreement (E >0.95).*

***Keywords**: ANN; Groundwater; Water Quality Index; Watershed; Modeling; Assessment.*

INTRODUCTION

Ground water used for domestic, industrial and for irrigation all over the world. In the last few decades, there has been a tremendous increase in the fresh water demand due to rapid growth of population and industrialization. Rapid urbanization, especially in developing countries like India, has affected the availability and quality of groundwater due to its overexploitation and improper waste disposal, especially in urban areas. According to WHO, about 80% of the diseases to human beings are caused by water. Once the groundwater is contaminated, its quality can not be restored by stopping the pollutants from the source. It is therefore imperative to monitor the quality of groundwater and device ways and means to protect it. After the samples of water have been collected analysis, a need arises to translate it to a form which is easily understandable. The formulation and use of indices have been strongly advocated by agencies responsible for water supply and control of water pollution. The water quality index (WQI) system is a well-known method of expressing water quality that offers a simple, stable and

1 Ph.D Scholar, School of Mechanical and Building Sciences, VIT University, Vellore, Tamilnadu - 632 014, India, Email: amuthanithy@gmail.com, Tel: 9444492468.
2 Research Fellow, Tropical Marine Science Institute, National University of Singapore, Singapore 119 223, Email: tmssp@nus.edu.sg.
3 Professor, School of Mechanical and Building Sciences, VIT University, Vellore, Tamilnadu, India.

reproducible unit of measure, which responds to changes in the principal characteristics of water (Brown et al., 1972). WQI is one of the most effective tools to communicate information on the quality of water to concerned citizens and policy makers (BIS 10500, 1991; WHO, 2008). Once the WQI are developed and applied, they serve as convenient tools to examine trends, to highlight specific environmental conditions, and to help governmental decision-makers in evaluating the effectiveness of regulatory programme (Ott, 1978). The following four steps are most often associated with the development of any WQI; depending on the sophistication is being aimed at, additional steps may also be taken (Prati et al., 1971; Ott, 1978; Schaeffer and Janardan, 1977; Amutha et al., 2008):

- Parameter selection,
- Transformation of the parameters of different units and dimensions to a common scale,
- Assignment of weightages to all the parameters and
- Aggregation of sub-indices to produce a final index score.

The WQI uses a scale from 0 to 100 to rate the quality of the water, with 100 being the highest possible score. Once the overall WQI score is determined, it will be compared against the following scale to determine how healthy the water is on a given day. WQI Water Rating Scale is 91-100: Excellent water quality, 71-90: Good water quality, 51-70: Medium or average water quality, 26-50: Fair water quality and 0-25: Poor water quality.

Applications of ANN in the areas of water engineering, ecological and environmental sciences are reported since the beginning of the 1990s. In recent years, ANNs have been used intensively for prediction and forecasting in a number of areas such as water resources (Liong et al., 1999; Maier and Dandy, 2000; Sundarambal et al., 2008). The present study is water quality index modeling using Artificial Neural network (ANN) techniques using the groundwater quality monitoring data for a Malattar sub-watershed, Gudiyattam Block, Vellore District, Tamil Nadu, India. This modeling helps in quick assessment of the WQI for the groundwater. The WQI computed using conventional method by Amutha et al., 2008, was used to compare with ANN model prediction in this study.

MATERIALS AND METHODS

Study area

The study area is located at Malattar sub-watershed, which is a major tributary of Palar river and covers a geographical area about 163 km^2 (Fig. 1). Malattar river originates in the hilly regions of Venkatagrikotta in Andra Pradesh and flows Niakeneri forest of Palamanar Thaluk, India. This river confluences Palar river 5 km east of Ambur near Sathampakkam village on the left side and flows through Pernampet block of Vellore District. The main tributaries of Malattar Rivers are Duggammaeru, Dandapaner venke, Gittargunta venka, Batavenka, Gooddar venka, Garisala venka and Kattar river. The watershed experiences tropical monsoon climate with normal temperature, humidity and evaporation throughout the year. The monsoon season in the watershed is from June to December. The occurrence of heavy rainfall during October and November contributes a significant amount of runoff in the watershed. In general, the annual rainfall is about 517.44 mm measured at the rainfall station, Modikuppam near Gudiyattam. Geology of the study area predominantly constitutes the fissile hornblende biotite gneiss, alluvium (recent) and dykes. Major lineaments are in the central part of the area and flowing through NW to SE and W to E. Groundwater pollution is widely perceived in the district.

The major sources are the effluents from the industrial units especially tanneries, which are engaged in the processing of raw leather. Disposal of urban sewage indiscriminately also can cause groundwater pollution. The non-point sources of pollution from fertilizers and pesticides applied in the agricultural fields. The dyeing units and other water-based industries located in Vellore District also contribute to Groundwater pollution.

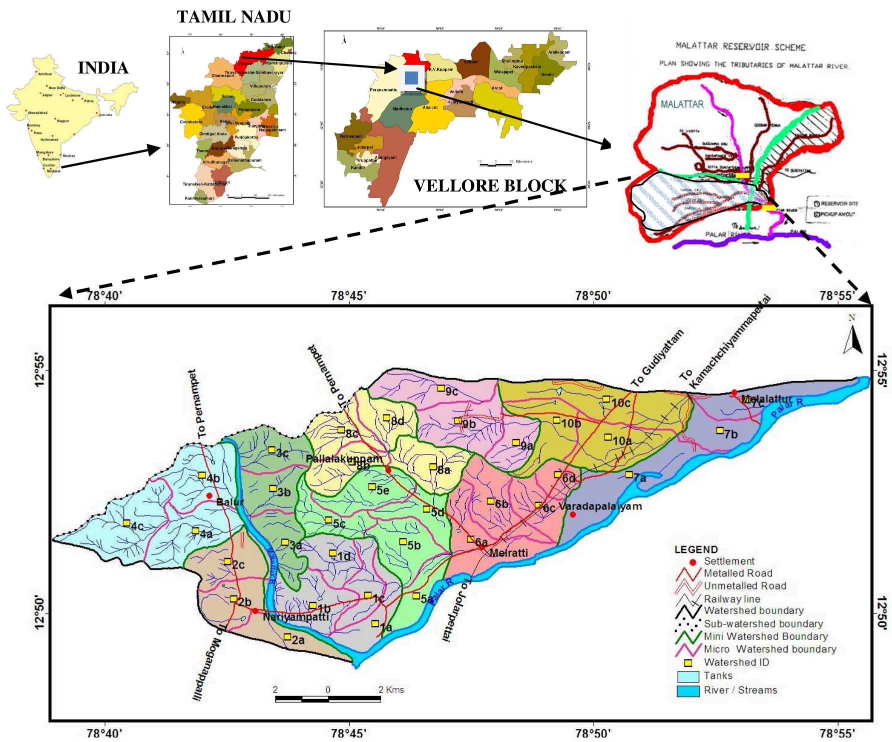


Fig. 1. The study area of the Malattar Sub-watershed.

Data

Thirty four groundwater samples were collected during May2006-April2007 covering the pre-monsoon and post monsoon period from both the bore wells and open wells in the study area at Malattar Sub-Watershed (Gudiyattam Block, Vellore) in Tamil Nadu, India (Fig. 1). The present study area constitutes different land use/land cover of about 50% of the area occupied by agricultural land, 60% area covers forest land, 17% area of cropland, 14% area of fallow land and remaining 21% of the area occupied by others such as water body, hills, settlement, upland with scrub and tanks. For calculation of WQI, selection of parameters has great importance since selection of too many parameters might widen the quality index and importance of various parameters depends on the intended use of water. The drinking water standards for these parameters are as recommended by the BIS. The WQI is calculated from physico-chemical parameters chosen for the present study namely pH, Total hardness (TH), Calcium (Ca^{2+}), Magnesium (Mg^{2+}), Chloride (Cl^-), Nitrate (NO_3^-), Sulfate (SO_4^{2-}), TDS, Fluoride (F) and Iron (Fe^{2+}), calculate. The collected samples are analyzed for the above said physicochemical parameters and their concentration ranges are shown in Table 1. With these chemical parameters, Gibbs and Piper plot diagram used for identifying the chemistry of ground water and concentrations of major cations and anions respectively by using Rock Works software (Amutha et al., 2008). The physical parameter pH was determined in the field at the time of sample collection. Chemical analyses is carried out for the major ion concentrations of the water samples collected from different locations using the standard procedures recommended by APHA (1994).

Table 1. Basic statistics of the measured physico-chemical parameters at Malattar Sub-Watershed.

Parameter (P_i)	Minimum	Maximum	Mean	Standard deviation	Median	Geomean
pH	5.50	7.50	6.29	0.43	6.50	6.28
TH	116.6	231.3	178.0	31.3	175.5	175.2
Ca^{2+}	13.0	38.0	25.9	6.8	25.5	25.0
Mg^{2+}	12.0	35.0	27.6	5.9	30.0	26.8
Cl^-	145.0	289.0	241.6	34.6	245.0	238.9
NO_3^-	9.0	19.0	14.6	2.2	14.5	14.4
SO_4^{2-}	123.0	287.0	231.3	44.2	245.0	226.3
TDS	423.0	674.0	577.9	62.0	590.0	574.4
F	0.01	0.40	0.12	0.13	0.06	0.07
Fe^{2+}	0.10	0.50	0.34	0.19	0.50	0.27

Method of WQI Determination

Conventional method

Water Quality Index (WQI) developed by the National Sanitation Foundation (NSF) in 1970 (Brown and others, 1970) is given in this section. As the magnitude of drinking water standard is smaller for the more harmful water pollutant and vice versa, the unit weight W_i for the ith parameter P_i is assumed to be inversely proportional to its recommended standard S_i ($i = 1, 2, \ldots, n$) where n is number of parameters considered (n=10 in the present case). Thus,

$$\mathbf{W_i = K/S_i} \qquad (1)$$

where the constant of proportionality K has been assumed to be equal to unity for the sake of simplicity. The unit weights W_i for the ten water quality parameters used here are shown in Amutha et al., 2008, where pH has been assigned the same weight as chloride. The quality rate q_i for the ith parameter P, all parameters except pH, is given by the relation

$$\mathbf{q_i = 100(V_i/S_i)} \qquad (2)$$

where V_i is the observed value of the ith parameter and S is its recommended standard for drinking water. For pH, the quality rating qpH can be calculated from the relation

$$\mathbf{qpH = 100[VpH \sim 7.0/1.5]} \qquad (3)$$

Where VpH is the observed value of pH and the symbol "~" means simply the algebraic difference between VpH and 7.0. Finally, the WQI can be calculated by taking the weighted arithmetic mean of the quality rating q_i, thus,

$$\mathbf{WQI} = [\sum(\mathbf{q_i W_i}) / \sum \mathbf{W_i}] \quad (4)$$

where both the summations are taken from i =1 to n.

ANN method

An artificial neural network is a computational tool that is able to acquire, represent and compute a mapping from one multivariate space of information to another, given a set of data representing that mapping. The advantage of ANN over traditional equations is that the exact function between a set of variables need not be known, this being a particular advantage to engineers where the underlying science of problems is not as yet determined or complex computation required and where data may be incomplete or noisy. A neural network 'learns', from a set of training data, a method of manipulating the input data in order to achieve the given result. Once a neural network has been trained in this way an additional set of data previously unseen can be presented to the network and the performance of the trained network assessed. Back-propagation neural networks have become a popular tool for modeling environmental systems (Maier and Dandy 1998).

In general, ANN models are specified by network topology, node characteristics and training or learning rules. It is an interconnected set of weights that contains the knowledge generated by the model. An ANN is composed of a large number of simple processing units, each interacting with others via excitatory or inhibitory connections (Fig. 2). Distributed representation over a large number of units together with interconnectedness among processing units, provide a fault tolerance. Three different layers can be distinguished:

(i) An input layer - connecting the input information to the network. In this study ten input nodes were applied, which are pH, TH, Ca^{2+}, Mg^{2+}, Cl^-, NO_3^-, SO_4^{2-}, TDS, F and Fe^{2+}.
(ii) Hidden layer (one or more hidden layer) - acting as the intermediate computational layer. Multi-layer feed forward networks formed by only one hidden layer.
(iii) Output layer - producing the desired output which in this case is WQI.

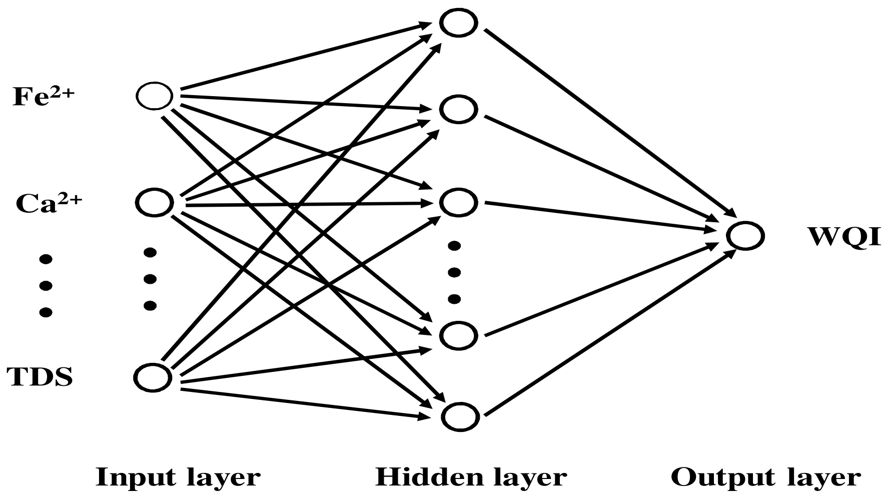


Fig. 2. Structure of a neural network for WQI prediction.

The circles depict nodes/neurons where information processing takes place and the lines depict the connection paths between each of the neurons. In order to learn the mathematical relationships represented in a process, ANNs are presented with a large number of training data sets. Once trained, the input neurons of the ANN are presented with the input values from unseen data sets for validation. The predicted output is compared to the actual value in the data set to enable the accuracy of the ANN to be determined. As ANN's are ideal for modeling non-linear relationships and they don't require any a priori knowledge on relationships of the process variables in question. An optimal ANN architecture may be considered as the one yielding the best performance in terms of error minimization, while retaining a simple and compact structure. There are two important issues concerning the implementation of artificial neural networks, that is, specifying the network architecture (the number of nodes and layers in the network) and finding the optimal values for the connection weights (selection of a training algorithm). In the process of specifying the network size, an insufficient number of hidden nodes cause difficulties in learning data whereas an excessive number of hidden nodes might lead to unnecessary training time with marginal improvement in training outcome as well as make the estimation for a suitable set of interconnection weights more difficult (Zealand et al., 1999). The performances of the models are evaluated using Nash-Sutcliffe coefficient of efficiency (E) (Nash and Sutcliffe, 1970), the square root of the mean square error (RMSE), mean square error (MSE), mean absolute deviation (MAD) and correlation coefficient (r). The ANN modeling steps include selection of performance criteria, data pre-processing (if necessary) and data division (a training set containing all possible extreme cases, over fitting test set and an independent validation set), selection of model inputs, outputs and network architecture, optimization of connection weights, training, testing and validation (Sundarambal et al., 2008).

In the present study, a commercial neural net software package NeuroShell 2™ Release 4.0 (Neuroshell, 2000) is used to develop the ANN model. To use the program, a set of inputs and outputs must be defined, and a suitable training set must be developed. ANN models with different learning approaches, such as back propagation neural networks (multi-layered perceptron (MLP), ward networks) and polynomial nets (Group Method of Data Handling (GMDH)), are considered and adopted to model the WQI in this study. The groundwater quality data (a total of 34 patterns) from 34 sampling wells in the study area (Fig. 1) were divided into three sets; training set, testing set and independent validation set contain 20 patterns, 8 patterns and 6 patterns respectively. The general approach of selecting a good training set from an available data series is to include all extreme events so that all possible minimum and maximum values present in the data set. The representative data, which was never seen before by trained ANN, were used as the validation set.

RESULTS AND DISCUSSIONS

ANN model was developed using the physico-chemical parameters as input (Table 1) and WQI as output. The results obtained from these modeling are discussed in this section. Relative

importance and contribution of the input variable iron to the output WQI of ANN prediction models using MLP and ward networks (Fig. 3). GMDH network also identified iron as the most important parameter in WQI prediction. Best WQI formula obtained from ANN prediction model using GMDH network was WQI = $-0.34+0.9X+0.32X^2$ where $X=2(Fe^{2+}-0.1)/0.4-1$.
Values of the WQI can be used not only to indicate the spatial variation of different bodies of water but also as a good indicator of behavior of water along environmental gradients. Each pattern in the Fig. 4a represents the bore wells and open wells in the study area. Changes of the WQI computed from both conventional method and ANN method for the groundwater samples collected from open and bore wells in the study area illustrated in Fig. 4a. ANN predicted WQI was compared that the conventionally computed (Fig. 4b) and they were in close agreement (E >0.98) (Table 2) for three selected ANN networks. WQI for the study area results that the area has poor and treatment necessarily requires before drinking water supply and overall water quality index observed that good groundwater supply.

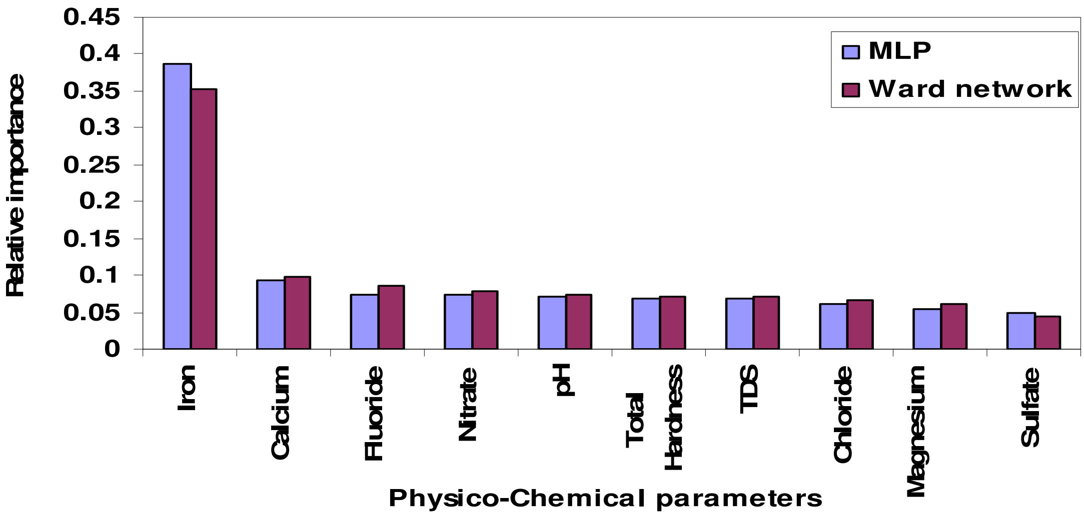
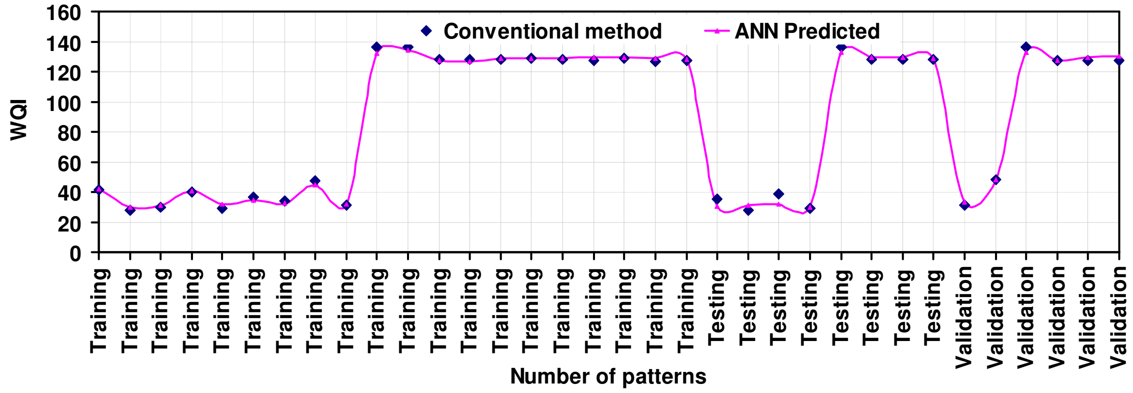
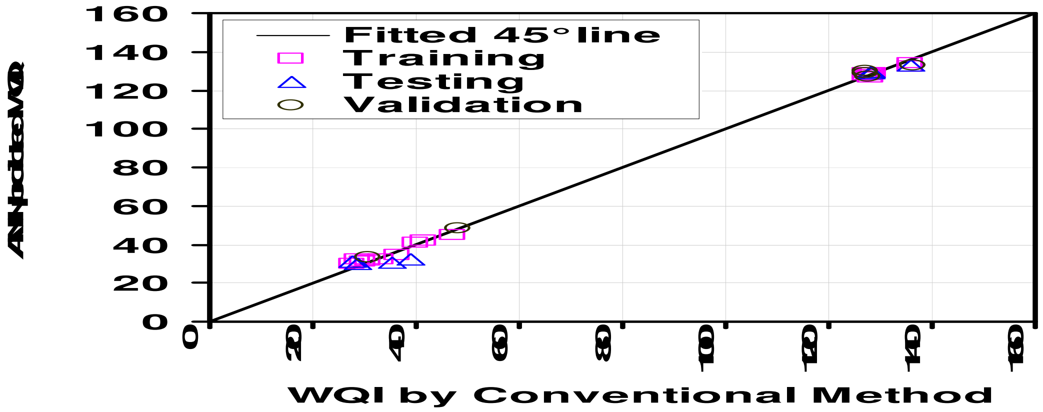


Fig. 3. Relative contribution of the input variables to the output of ANN prediction model.

Data points in Fig. 4b. were tightly clustered with WQI between 30 to 50 and 120 to 140, and displayed nearly equal dispersion above and below the 45° fitted line. ANNs are able to simulate the WQI with RMSE <5.5 and performance indicators shows that ward network are performing better than other two networks for WQI prediction. This appeared to confirm a strong correlation between the selected input parameters and WQI. The overall results of all three selected networks are comparable to each other. The performance indicators for WQI prediction by ANN models were found adequate even in the case of high WQI during monsoon changes and unknown local events (Table 2 and Fig. 4). Once validated WQI models are ready for prediction, the input parameters are analyzed from the collected samples and WQI can be predicted without tedious calculation using conventional method. The proposed ANN models show their efficiency in predicting the WQI profiles in Malattar Sub-Watershed, and it is in accordance with the results from conventional method. ANN will be seen as a powerful predictive alternative to traditional modeling techniques. The groundwater quality prediction can be easily affected with high uncertainty and specific circumstances, such as climatological, eco-regional and local human activities. Proposed models could show certain deviations. Thus, it is necessary to update the proposed models from time to time with actual measured values.



a.



b.

Fig. 4. (a) WQI for groundwater by Conventional method and ANN prediction; (b) Scatter plots for training, testing and validation of ANN model.

Table 2. Average WQI rating of groundwater sources in different season.

ANN architecture		Training	Testing	Validation
MLP	E	1.00	0.99	0.98
	MSE	10.7	16.2	29.0
	RMSE	3.27	4.03	5.38
	r	1.00	1.00	1.00
Ward Net	E	1.00	1.00	1.00
	MSE	5.4	11.9	5.0
	RMSE	2.32	3.45	2.22
	r	1.00	1.00	1.00
GMDH	E	1.00	0.99	0.99
	MSE	12.3	17.7	14.2
	RMSE	3.51	4.21	3.76
	r	1.00	0.84	1.00

CONCLUSIONS

In this paper, ANN models ere developed to predict WQI in the Malattar Sub-Watershed. ANN structure has been designed and trained using the Neuroshell software. The performance of theANN networks were tested using performance indicators. Results of simulation, presented in this paper, show that the application of the neural network to prediction of WQI gives satisfactory results. Proposed approach can be a very efficient tool and useful alternative for the computation of water quality index. ANN simplifies and speeds up the computation of the WQI, as compared to the currently existing method. By optimizing the calculation, a significant saving in terms of money and time can be achieved. ANN's "learn" by example as long as the input dataset contains a wide range of the types of patterns that the ANN will be asked to predict, and the model uses them successfully to predict the output using those patterns. Thus, if extreme data range events occur, the models have to be recalibrated and revalidated.

ACKNOWLEDGEMENTS

This research is a part of the main author's PhD research. We are grateful to Dr. Jegathambal Palanichamy for technical guidance and friends who helped in field monitoring and laboratory analysis.

REFERENCES

Amutha, R., Porchelvan, P. and Poovalinga. G. B., 2008. Remote sensing and GIS techniques for evaluation of groundwater quality index in Malattar sub- watershed, Gudiyattam block, Vellore District, Tamil Nadu. *Eco-Chronicle*, 283-290.

APHA (American Public Health Association), 1994. *Standard method for examination of water and wastewater*, NW, DC 20036.

BIS (Bureau of Indian Standards) 10500, 1991. *Indian standard drinking water- specification*. First revision, 1-8.

Brown, R.M., McClelland, N. I., Deininger, R. A and Tozer, R., 1970. A Water Quality Index – Do We Dare?, *Water Sewage Works*, 11, 339-343.

Brown, R. M., McLelland, N. I., Deininger, R. A. and O'Connor, M. F., 1972. A water quality index - crashing the psychological barrier. In: *Indicators of Environmental Quality*, Ed. W.A. Thomas, Plenum Press, N.Y.

Liong, S.Y., Lim, W.H., Audyal, G., 1999. Real Time River Stage Forecasting for Flood Stricken Bangladesh: Neural Network Approach. *Journal of Computing in Civil Engineering (ASCE)*, 14(1), 1–8.

Maier, H.R. and Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and Software*, 15, 101–124.

Nash, J.E. and Sutcliffe, J.V., 1970, River flow forecasting through conceptual models; part I - a discussion of principles. *Journal of Hydrology*, 10, 282–290.

Neuroshell 2^{TM}, 2000. Neuroshell Tutorial, Ward Systems Group, Inc., Frederick, MD, USA.

Ott, W.R., 1978. *Environmental Indices: Theory and Practice*. Ann Abbor. Science, Ann Arbor, Michigan, 371 pp.

Ott, W.R., 1978. *Water Quality Indices: A Survey of Indices used in the United States*. EPA –

600\4-78-005, Environmental Protection Agency, Washington, D.C.

Prati, L., Pavanello, R. and Pesarin, F., 1971. Assessment of Surface Water Quality by a Single Index of Pollution. *Water Research*, 5, 741-751.

Schaeffer, D. J. and Janardan, K. G., 1977. Communicating Environmental Information to the Public: A Water Quality Index. *J. Environ. Educ.*, 8(4), 16-26.

Sundarambal, P., Liong, S. Y., Tkalich, P., 2008. ANN application for water quality forecasting. *Marine Pollution Bulletin*, 56(9), 1586–1597.

WHO (World Health Organization), 2008. *Guidelines for drinking water quality*. 3rd Ed., 1, 515 pp.

Zealand, C. M., Burn, D. H. and Simonovic, S. P., 1999. Short term stream flow forecasting using artificial neural networks, *Journal of Hydrology*, 214, 32–48.