

DEVELOPING A pH MODEL USING ARTIFICIAL NEURAL NETWORK AND VISUAL MODFLOW TO EVALUATE GROUNDWATER QUALITY

JESSY MOL, I.^{1*} – BASKARAN, T.² – JUSTIN JOSE, D.³

¹*Civil Department, St. Xavier's Catholic College of Engineering, Nagercoil 629003, Tamil Nadu, India
(e-mail: jessymol@gmail.com)*

²*Civil Department, Thiagarajar College of Engineering, Madurai 625015, Tamil Nadu, India
(e-mail: tbaskaran70@gmail.com)*

³*Computer Science Department, Marthandam College of Engineering, Kuttakuzhi 629177, Tamil Nadu, India
(e-mail: mail2djose@gmail.com)*

**Corresponding author
e-mail: jessymol@gmail.com*

(Received 15th Apr 2020; accepted 13th Aug 2020)

Abstract. Groundwater pollution cause excessive levels of water quality parameters such as pH, Total Dissolved Solids (TDS), hardness etc, therefore, it is a severe problem for people. The present study describes the prediction of pH in groundwater using Artificial Neural Network (ANN) model and visual MODFLOW. It is observed that the pH level of study area is lower than the standard limit so that the water turns to acidic. The water is not suitable for drinking purposes thus a prediction model is needed to rectify this issue. A smart prediction model for pH was developed using ANN. Selected input variables were Hardness, Calcium (Ca), Magnesium (Mg), Sodium (Na), Potassium (K), Nitrite (NO₃), Chloride (Cl) and Sulphate (SO₄) and the best performance value was 0.025637. Also to simulate the groundwater of the study area, visual MODFLOW model was created. Using this model, pH is simulated for 365 days. From the calibration plot, it is known that the correlation coefficient of the observed and the simulated value is 0.92. From the simulation it is concluded that the pH level was almost the same for 365 days.

Keywords: Mondaikadu, water quality, solute transport, simulation, prediction

Introduction

Groundwater has been essential to sustain India's economy and to fulfill its domestic needs, and it also has agricultural and industrial use. Around 33% of the total population relies upon the utilization of groundwater for drinking (Nickson et al., 2005). Groundwater accounts for about 98% of all of the planet's available fresh water, and it is about 60 times more concentrated than the fresh water present in lakes and streams. India is the world's biggest groundwater user, with a yearly use of roughly 230 cubic kilometers for each annum (World Bank, 2012). The nation as a whole has an annual groundwater supply of 398 billion cubic meters (Water and Related Statistics, 2015). More than 60% of farming and 85% of drinking water supplies depending on it, groundwater is a basic asset for Indian rural zones.

Nowadays, groundwater pollution is considered as a world level issue. Many organizations are taking serious actions against this issue and creating regular monitoring programmes. A research studies on various features of groundwater, such as, storage potential, hydrogeology, water quality, vulnerability, and sustainability and

so on was conducted by many researchers (Chapagai et al., 2010; Pandey et al., 2011, 2012). Particularly the quality of water is considered a key contributor to various diseases. In general, groundwater is less vulnerable to pollution as compared to surface water. Contamination of groundwater is due to the presence of certain pollutants in groundwater which exceed limits prescribed for drinking water. Arsenic, nitrate and carbon, which are of a geogenic origin, are the widely found pollutants. Many pollutants include bacteria, phosphates, and heavy metals that arise from human activities like domestic waste, agricultural practices, and industrial effluents. Groundwater quality may change from place to place as well as season to season. Weak environmental management schemes in factories contribute to water discharges of hazardous and agricultural wastes. This resulted in the contamination of those surface and groundwaters from which water is drawn for irrigation and household use (Pai and Mulye, 2016). Landuse pattern is one of the reasons for groundwater contaminations.

Many intelligent tools, such as linear and non-linear tools are successfully used to predict the water quality. Nowadays artificial neural network is used to find a solution for many cases of water quality problems. Wagh et al. (2016) used the ANN model to predict sodium adsorption ratio (SAR) values, residual sodium carbonate, magnesium adsorption ratio, Kelly's ratio and sodium percentage in Nanded tehsil ground water. It is found from the results that there is a strong agreement between the actual data and the ANN outputs for indices of irrigation suitability for training and testing datasets. Mohammadi et al. (2016) utilized ANN model to simulate concentrations of fluoride in groundwater resources in Khaf and surrounding villages based on the water's physical and chemical properties. The best results were observed by the MLP1 model with eight parameter inputs such as root mean square error (RMSE) and correlation coefficient of real and expected outputs. The simulation results from the MLP1 test stage as well as the high correlation between experimental and projected data suggested that this model can be used to predict fluoride concentrations in groundwater supplies with its high confidence coefficient. Kheradpisheh et al. (2015) developed the artificial neural network for Cl^- , EC, SO_4^{2-} and NO_3^- in order to assess the influence of the key input parameters, the test results were estimated and compared with the measured values. Palani et al. (2008) predicted the quantitative characteristics of the water bodies using ANN. The ANN model was developed at any location in the domain of interest for fast assessment and selected water quality variables were forecasted. The input parameters measured at other locations serve as respective variables. Input variables include salinity, temperature, dissolved oxygen, and chlorophyll-alpha. The findings demonstrate the great potential of the ANN for simulating variables in water quality.

During the summer, winter and monsoon seasons, the level of pH in the study area is much less (5.4 to 6.2) when compared to the World Health Organization (pH 6.5-8.5) (WHO, 2008) and Indian Standards (pH 6.5-8.5) (BIS, 2009), therefore the water becomes acidic in nature and the quality of the water is degraded. The reason for low pH is due to the coconut retting zone of the study area (Jessy Mol and Baskaran, 2017). Since the level of pH is decreasing, modelling of pH is needed to find out the pH level in the future. The novel aspect of this study is the Visual MODFLOW combined with MT3DMS, it is used for the groundwater quality model, where the coconut retting zone of the study area is taken as the recharge zone for boundary conditions in MODFLOW. Therefore, the influence of the retting zone on the pH level can be predicted very easily.

Artificial neural network (ANN)

ANN is a tool for mathematical modelling. It works based on neural processing in the human brain. In the human brain, each neuron is connected as a network. In the same manner, in the neural network, the layout has a set of neurons and they are linked and organized. The information passes through a neural network, from the input to the output. The Network consists of interconnected node layers. These nodes are called neurons which function as elementary units of processing. From the various inputs each neuron receives the information and produces an output by the value. Its activation function happens when the weighted whole of its information is the statement (Kilicaslan et al., 2014). Every layer is related by its respective weights to the next layer (Nasr et al., 2012). Eq.1 is used to calculate the weighted sum of the layer.

$$y_j = \sum_{i=1}^n W_{ij}x_i + \theta \quad (\text{Eq.1})$$

where y_j is the weighted sum of the j th neuron for the input data obtained from the n -neuron proceeding layer, W_{ij} is the weight between the i th neuron and the j th neuron in the proceeding layer, x_i is the contribution of the i th neuron in the proceeding layer and θ may be biased by the j th neuron. The input layer of the network is used to feed the data to the network; the hidden layer of the network is used to act as a collection of features detector, and the output layer is used to get out the result (Seyamand Mogheir, 2011).

In the forward direction, the input signal passes through the network from one to another layer. This process is named as Multi Layer Perception (MLP). Back Propagation Algorithm is used in MLP to solve the problems. This algorithm consists of Feed Forward and Feed Backward Network (Subbarayan, 2013). Feed Forward Back Propagation Network is mostly used in many water resources problems. The optimal weight is calculated by minimizing the Mean Square Error (MSE) of the training stage (Nasr et al, 2013). Mean Square Error is calculated by Eq.2.

$$MSE = \frac{1}{n} \sum_{i=1}^n (T_i - Out_i)^2 \quad (\text{Eq.2})$$

where T_i and the out_i are the desired neural network target and output for the i th neuron respectively, n is the number of samples in the network. Neural networks have three stages; these are training stage, validation stage and testing stage. In the training stage, the network's weight and bias are computed. Learning functions are used at this stage to update weight and bias of the layer. Here, the trained network is validated. After that, the data set is used to examine the generalization of the network. This stage is the testing stage.

The network is classified as two types namely, supervised and unsupervised classification. Supervised classification means that the input and output data are given and the error is calculated for the predicted output and the output given to the network. Mostly used training function is `trainlm` function, Levenberg – Marquardt back propagation (`trainlm`) function locates the minimum of a multivariate equation it can be represented as the sum of the non-linear real-evaluated functions squares. It's an iterative technique that operates in such a way that every iteration of the algorithm is always reduced by output level. This component makes `trainlm` the quickest training algorithm for reasonably sized networks (Sharma and Venugopalan, 2014). This

training function is often capable of getting lower mean square errors than the other training function (Kumar, 2010).

Visual MODFLOW model

Visual Modflow is used to simulate the Groundwater flow because of its most comprehensive and user-friendly modeling environment. MT3DMS is one of Visual Modflow's problem solving tool which is used to evaluate the simulation of contaminant transport. This model is completely incorporated, combining the strong analytical tools with a clear menu structure. In model creation or analysis of results, the model info parameters and results can be envisioned in 2D and 3D at any minute. The mathematical equation for groundwater model is given by Eq.3.

$$\frac{\partial}{\partial x} \left\{ Kx \frac{\partial h}{\partial x} \right\} + \frac{\partial}{\partial y} \left\{ Ky \frac{\partial h}{\partial y} \right\} + \frac{\partial}{\partial z} \left\{ Kz \frac{\partial h}{\partial z} \right\} = Ss \frac{\partial h}{\partial t} \quad (\text{Eq.3})$$

where, Kx, Ky are the Hydraulic conductivity of the x and y directions respectively, Kz is the Hydraulic conductivity of the z direction, x, y, z are the Cartesian coordination, h is the Hydraulic head, Ss is the Specific storage of the porous material and t is the Time.

Transport equation

The equation represents motion of the solute mass transport through a volume of control. It indicates that the amount of masses absorbing or producing solute with the volume of control must be equal to a shift of solvent concentration with the volume of control. Eq.4 gives the solute transport equation that is implanted in MT3D used for modeling.

$$\frac{\partial C}{\partial t} = \left[\frac{\partial}{\partial x_i} \left(D_{ij} \frac{\partial C}{\partial x_j} \right) \right] - \left[\frac{\partial}{\partial x_i} (V_i C) \right] \quad (\text{Eq.4})$$

where,

V_i = Seepage velocities in x, y, z directions, m/s (LT^{-1}),

D_{ij} = Dispersion coefficients, m^2/sec ($L^2 T^{-1}$),

C = Solute concentration, mg/m^3 (ML^3),

t = Time.

Input data

The detailed hydrogeological description of the study area was collected to allocate the required modeling inputs. The data required to input the MODFLOW for groundwater modelling are elevation, quality data, observation well details, subsurface thickness, hydraulic head, contaminant concentration, conductivity, storage properties, dispersion coefficient, evapo-transpiration, recharge and Boundary conditions.

Simulation

A stress cycle is characterized as a time period during which all system stresses are constant. Design Visual Modflow was utilizing the data from the first time period of each given boundary condition for a steady state simulation. For transient simulations,

all of the data described in different time period for each boundary condition and well was automatically merged into the stress period format provided by Modflow.

Results and Discussion

ANN model for pH

A back-propagation Feed Forward Neural Network (FFNN) algorithm is used to compare the relationship between input and output. This model is performed using MatlabTool box for ANN (MATLABand StatisticsToolbox, 2012). Based on prior research by Sathyamurthy (2013), the ANN inputs for pH are identified. The selected output is pH and the input variables are Hardness, Ca^{2+} , Mg^{2+} , Na^+ , K^+ , NO_3^- , Cl^- and SO_4^{2-} . Several trials were conducted for each group to attain the suitable network structure until the appropriate training rate, the number of hidden layers and the number of neurons per hidden layer was reached. The best structure is the one that produces the minimum MSE in both the training and testing data (Nasr and Zahran,2014). The algorithm for back propagation minimizes the Mean Square Error in the output layer between the observed and the predicted output. The structure resulting in a minimal number of errors is the one selected as shown in figure. The network properties are as follows:

- Data Used: 160 for training 20 for testing.
- Network input: Hardness, Ca^{2+} , Mg^{2+} , Na^+ , K^+ , NO_3^- , Cl^- and SO_4^{2-} .
- Network output: pH.
- Network type: feed-forward back –propagation.
- Training function: Levenberg–Marquardt algorithm (TRAINLM).
- Adaptation learning function: LEARNGDM.
- Performance function: MSE.
- Epochs: 1000.
- Number of neurons: 25, Fig. 2 shows the neural network diagram of the model.

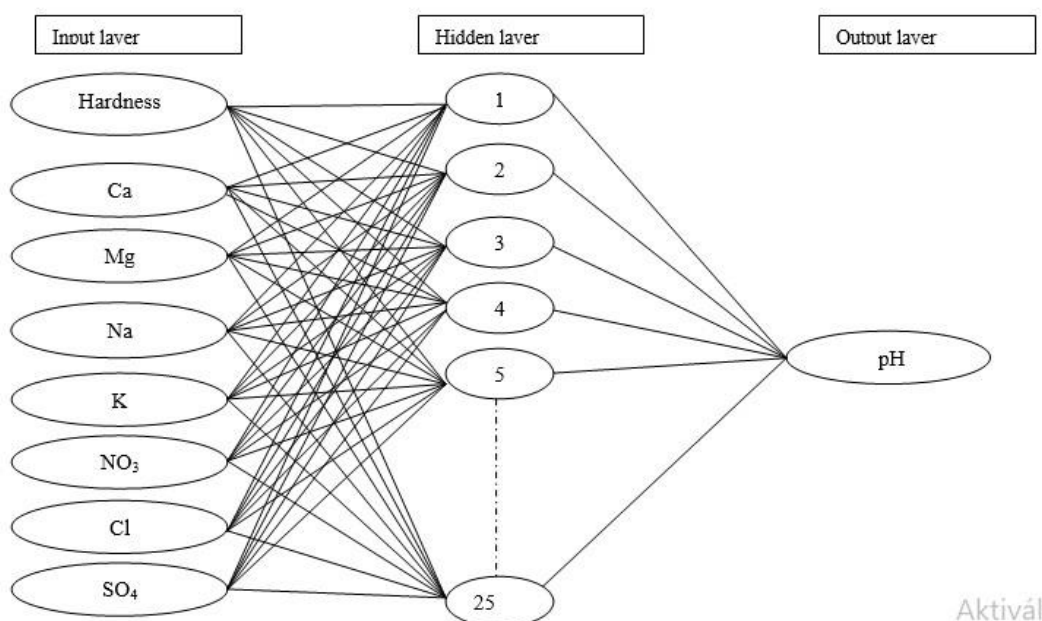


Figure 2. Neural Network Diagram of pH

To terminate the network training, the gradient's magnitude and the iterations of validation checks is used. The number of validity checks is 9, which is the right value to quit training. The output plot indicates the function's value as opposed to the iteration number in terms of preparation, evaluation, and test behaviors. Based on the MSE the best validation performance is 0.025637 at epoch 3 as shown in *Fig. 3*. Since the validation and test graphs are close, the training did not pose any major problems or over fit. During preparation, each layer neuron changes its weight vector toward the input vector group nearest to it.

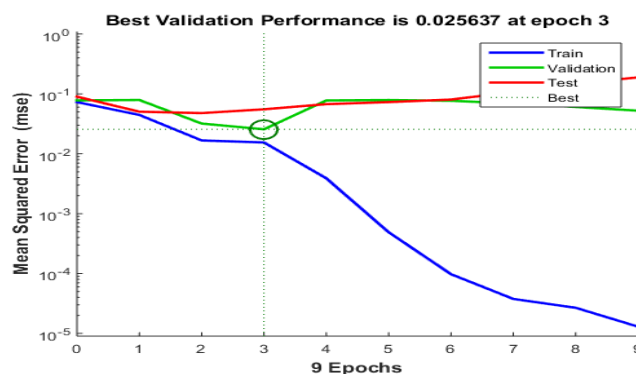


Figure 3. Training performance graph

As shown in *Fig. 4*, for training, validation, and testing, a linear regression analysis is performed to find out the correlation between the network outputs and goals. The dotted line represents the optimal outcome in each map, i.e. outputs = goals, while the dark line indicates the best linear regression that suits. The effects of regression (R-value), respectively, are 0.96, 0.99 and 0.98 for preparation, validation and research. Those results corresponded to a cumulative 0.96.

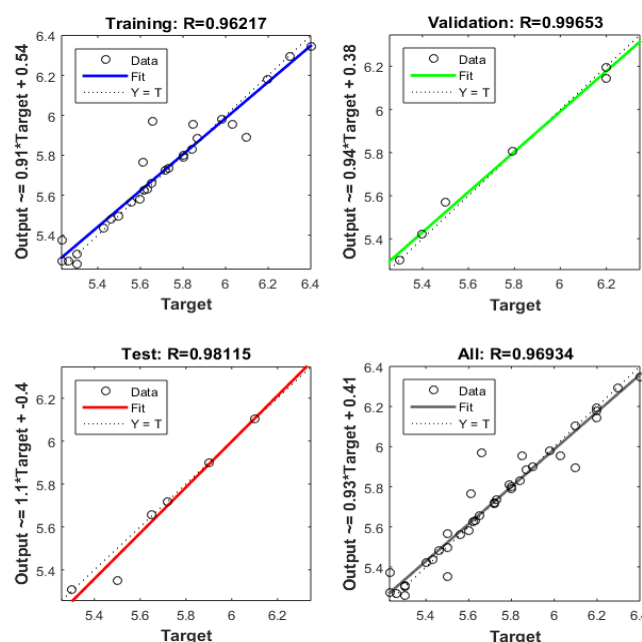


Figure 4. Regression plot for training, validation and testing

After training, validation and testing of the network, the created network can be used to predict the pH via the new input data. From *Fig. 5* it was found that the experimental pH analysis is similar to the expected data determined from the network and this confirms this model's validity. *Fig. 6* shows the linear relationship between the observed and predicted pH and it found R^2 0.927. From this, it is inferred that the model established is apt to predict groundwater pH.

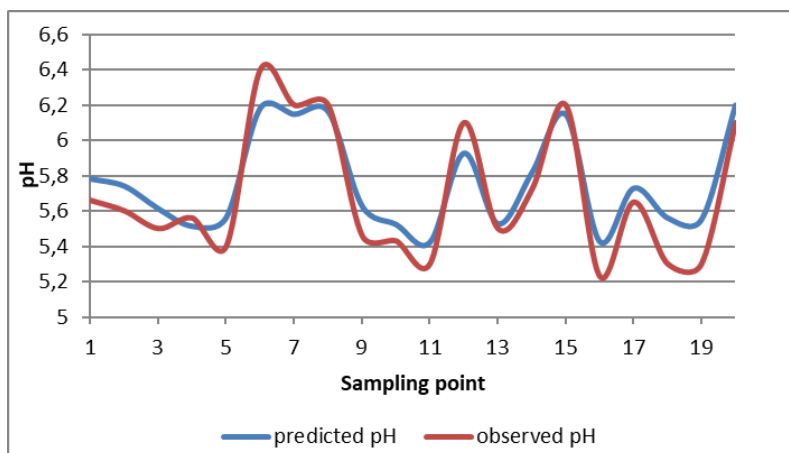


Figure 5. Comparison of observed and predicted pH

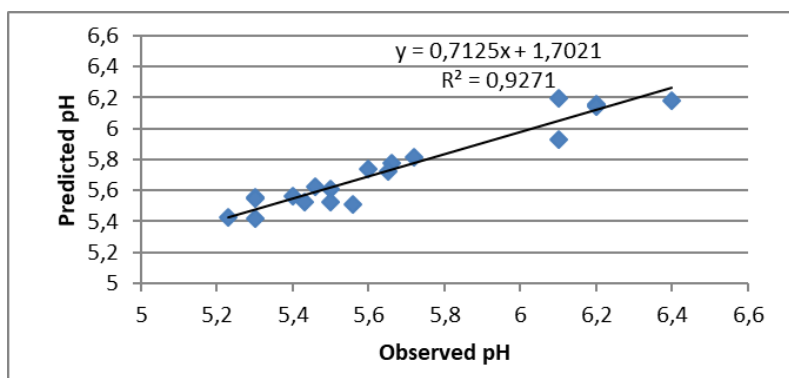


Figure 6. Linear relation between observed and predicted EC

MODFLOW model

Modeling of groundwater flow and transport is performed using Visual MODFLOW and MT3DMS tools (Harbaugh, 2005). The three Visual MODFLOW sections are Input, Run, and Output. The groundwater parameters and aquifer properties are given in the input section as input to the model. Based on the specifications, the correct engine and solver parameters are selected in the run portion, and the model is simulated.

Input of transport modeling

Hydro-geological characterization of the study area is performed to allocate the required modeling data. *Fig. 7* shows output obtained from the model, such as location of pumping wells and concentration wells and boundary conditions. The assigned

boundary conditions of the study area are concentration and recharge boundary conditions. Since the pH level of the study area is less than the drinking water standards (WHO) the level of pH is given as the input to the concentration wells.

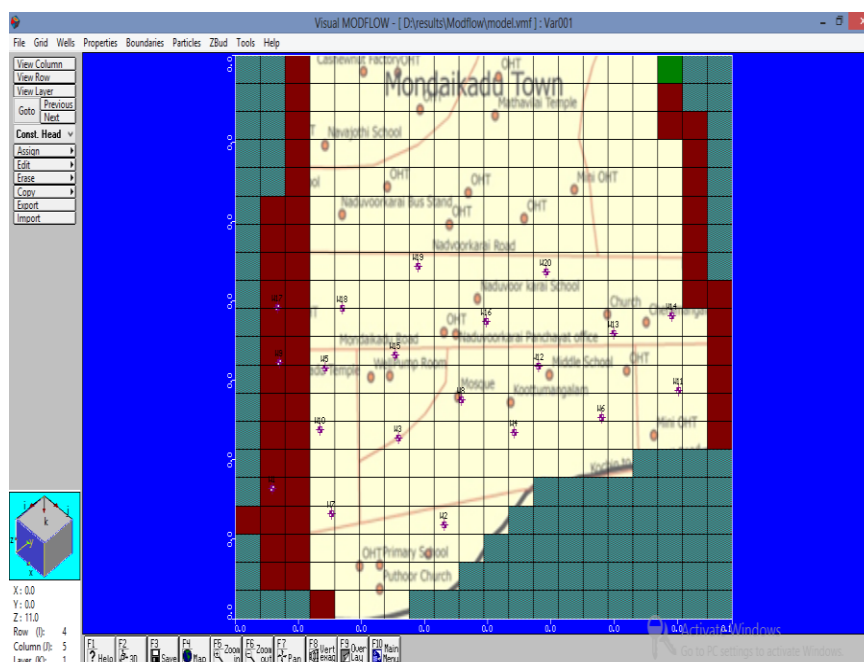


Figure 7. Boundary conditions of study area

Model simulation

The results of the transport contamination were simulated using the following assumptions.

Run Type	-	Transient flow
Time Steps	-	10 stress periods
Simulation Period	-	365 days
Solver	-	WHS solver
Layer	-	Single layer of unconfined aquifer

Model output

In *Fig. 8* the flow direction is initially from the south side towards the other three sides of the study area, because the south side is given as the recharge zone of the study area. The vectors of velocities represent the speed and direction of the particle of water as it travels through the field of flow.

Calibration of model

The pH is simulated and calibrated for 365 days; the calibration plot between the observed and the predicted pH is shown in *Fig. 9*. It shows the observed pH is 5.68 and the calculated pH is 5.35 same as the observed value of pH for another sample is 5.15 and observed value is 5.20. The correlation coefficient between the observed and the calculated value is 0.92; this value is nearer to 1, and therefore the model is calibrated in a correct manner.

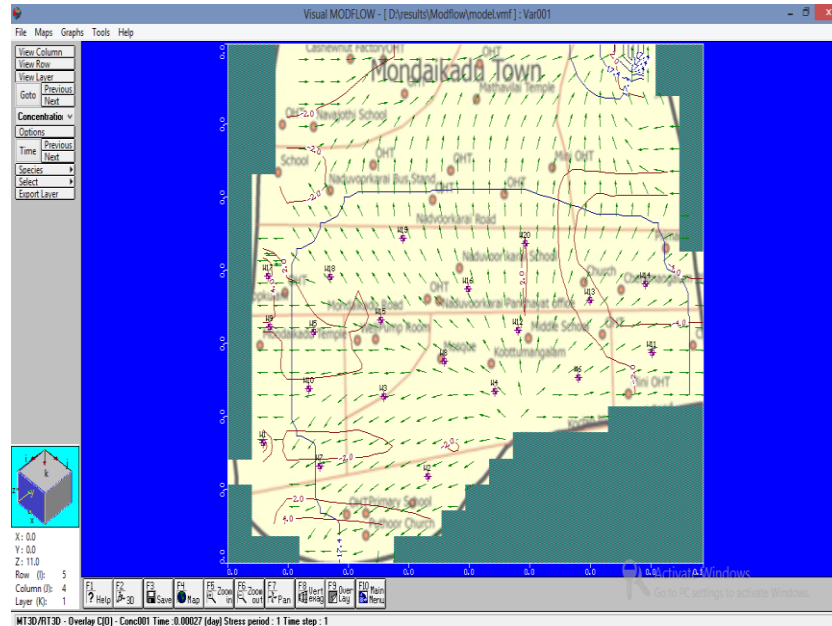


Figure 8. Flow directions and contour lines

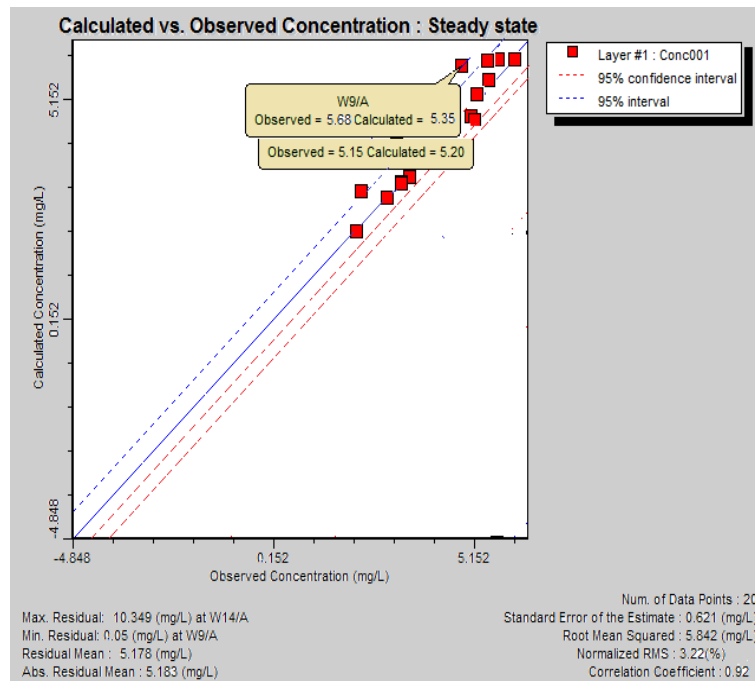


Figure 9. Model Calibration Chart

Conclusions

The pH was predicted by ANN model, using Hardness, Ca^{2+} , Mg^{2+} , Na^+ , K^+ , NO_3^- , Cl^- and SO_4^{2-} as inputs. The best validation performance was 0.025637. Since the ANN model is based on the output data, the approach can be used for the other basins of study area, depending on the availability of data. To simulate the pH, Visual MODFLOW

model was used. The flow concentration was simulated for 365 days, and the correlation coefficient of the observed and the simulated value was 0.92. The decision makers and the groundwater managers associated with Mondaikadu will benefit if they use the findings of this paper as a decision support tool. It is recommended that the research can be done in the future by implementing various artificial recharge structures in the study area to enhance the groundwater quality.

Acknowledgements. We would like to acknowledge our relatives, friends and colleagues for their valuable suggestions and encouragement for preparing this manuscript. We would like to appreciate the reviewers and editors for their suggestions and comments.

REFERENCES

- [1] APHA. (1998): Standard Methods for the Examination of Water and Wastewater. – 20th ed., American Public Health Association, Washington DC: 2005-2605.
- [2] BIS. (2009): Indian standards specifications for drinking water. – BIS:10500- Bureau of Indian Standards. New Delhi.
- [3] Chapagai, S., Pandey, V.P., Shrestha, S., Nakamura, T., Kazama, F.(2010): Assessment of deep Groundwater quality in Kathmandu Valley using multivariate statistical techniques. – Water, Air and Soil Pollution 210(1-4):277-288.
- [4] ENVIS centre Tamil Nadu: Kanyakumari district. – <http://tnenvis.nic.in/files/KANYAKUMARI%20%20.pdf>.
- [5] Harbaugh, A.W.(2005): MODFLOW, the U.S. Geological Survey modular ground-water model. – the Ground-Water Flow Process: U.S. Geological Survey Techniques and Methods 6-A16.
- [6] Jessy Mol, I., Baskaran, T. (2017): Evaluation of groundwater quality at coconut husk retting area. – Applied Ecology and Environmental Research 15: 1745-1756.
- [7] Kheradpisheh, Z., Talebi, A., Rafati, L., Ghaneeian, M.T., Ehrampoush, M.H. (2015): Groundwater quality assessment using artificial neural network: A case study of Bahabad plain, Yazd, Iran. – Desert 20(1):65-71.
- [8] Kilicaslan, Y., Tuna, G., Gezer, G., Gulez, K., Arkoc, O., Potirakis, S. M. (2014): ANN-Based Estimation of Groundwater Quality Using a Wireless Water Quality Network. – International Journal of Distributed Sensor Networks 10(7):1-8.
- [9] Kumar, N. (2010): Analysis of Groundwater for Potability from Tiruchirappalli City Using Backpropagation ANN Model and GIS. – Journal of Environmental Protection 1(2): 136-142.
- [10] MATLAB and Statistics Toolbox Release (2012): The MathWorks. – Natick, Massachusetts, United States.
- [11] Mohammadi, A.A., Ghaderpoori, M., Yousefi, M., Rahmatipoor, M., Javan, S. (2016): Prediction and modeling of fluoride concentrations in groundwater resources using an artificial neural network: a case study in Khaf. – Environmental Health Engineering and Management Journal 3(4): 217-224.
- [12] Nasr, M., Tawfik, A., Ookawara, S., Suzuki, M.(2013): Prediction of hydrogen production using artificial neural network. – Seventeenth International Water Technology Conference, IWTC17, Istanbul.
- [13] Nasr, M., Zahran, H. F. (2014): Using of pH as a tool to predict salinity of groundwater for irrigation purpose using artificial neural network. – Egyptian Journal of Aquatic Research 40(2):111-115.
- [14] Nickson, R.T., McArthur, J.M., Shrestha, B., Kyaw-Myint, T.O., Lowry, D.(2005): Arsenic and other drinking water quality issues, Muzaffargarh District, Pakistan. – Applied Geochemistry 20:55-68.

- [15] Pai, S., Mulye, P. (2016): Groundwater Levels Plunge, Contamination Rises, Crisis Grows. – India Spend, Journalism India.
- [16] Palani, S., Liong, S.-Y., Tkalich, P. (2008): An ANN Application for Water Quality Forecasting. – Marine pollution bulletin 56: 1586-97.
- [17] Pandey, V.P., Kazama, F. (2011): Hydro geological characteristics of groundwater aquifers in Kathmandu Valley Nepal. – Environmental Earth Science 62(8): 1723-1732.
- [18] Pandey, V.P., Kazama, F. (2012): Groundwater storage potential in the Kathmandu Valley's shallow and deep aquifers. – Kathmandu Valley Groundwater Outlook, pp. 31-38.
- [19] Seyam, M., Mogheir, Y. (2011): Application of Artificial Neural Networks Model as Analytical Tool for Groundwater Salinity. – Journal of Environmental Protection 2: 56-71.
- [20] Sharma, B., Venugopalan, K. (2014): Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images. – IOSR Journal of Computer Engineering 16(1): 31-35.
- [21] Subbarayan, S. (2013): Modelling of Irrigation Water Quality of Coastal Area Using Back Propagation-Multi Layer Perceptron Artificial Neural Network. – Universal Journal of Environmental Research and Technology 3(2): 273-280.
- [22] Wagh, V.M., Panaskar, D.B., Muley, A.A. (2016): Prediction of groundwater suitability for irrigation using artificial neural network model: a case study of Nanded tehsil, Maharashtra, India. – Modelling of Earth System and Environment 2: 1-10.
- [23] Water and Related Statistics (2015): Central Water Commission.
- [24] World Bank report (2012): India Groundwater: a Valuable but Diminishing Resource.
- [25] World Health Organization (2008): Guidelines for drinking water quality. – Recommendations Geneva 1: 1308.