

Evaluation of ground water quality contaminants using linear regression and artificial neural network models

G. Charulatha¹ · S. Srinivasalu¹ · O. Uma Maheswari² · T. Venugopal³ · L. Giridharan⁴

Received: 6 January 2016 / Accepted: 19 January 2017 / Published online: 20 March 2017
© Saudi Society for Geosciences 2017

Abstract Contamination of groundwater is one of the major health concerns in the rapidly urbanizing and industrializing world. Since groundwater is one of the most important resources for the domestic, industrial, and agricultural purposes, the quality and quantity is of prime importance. Nitrite which is a reduced form of nitrate ion is one of the potential contaminants in the groundwater. The detection of nitrite ion is one of the laborious works and also it gets easily oxidized to nitrate ion and hence modeling approaches for the nitrite concentration will be one of the resilient quantification techniques. In the present study, the effective performance of the linear and non-linear models such as multiple linear regression (MLR), principal component regression (PCR), artificial neural network (ANN), and the integrated technique of principal components and artificial neural network (PC-ANN) is evaluated in the prediction of the nitrite concentration. The MLR and PCR showed better results either in generation step or in the validation step but not both. ANN shows better results in both generation and validation steps but the results in the validation steps, though good but accuracy is comparatively lower than the generation step. In the case of PC-ANN, the prediction of the model is found to be good both in the generation and in the

validation steps. The Nash–Sutcliffe efficiency test clearly illustrates better performance of PC-ANN in comparison with other models in the present study for the quantification of nitrite concentration in groundwater.

Keywords Groundwater · Nitrite · Multiple linear regression · Principal component regression · Artificial neural network · Principal components-artificial neural network

Introduction

Regression analysis is a statistical approach of relating a variable of interest, namely the dependent variable to a set of independent variables. The dependent and independent variables are referred as response and explanatory variables, respectively. The regression model adequately describes, predicts, and controls the dependent variable on the basis of the independent variables. In simple linear regression, scores on one variable are predicted from the scores of another variable. The variable being predicted is generally referred as criterion variable and the variable which forms the basis for the predictions is referred as predictor variable. When there is only one predictor variable, the prediction method is referred as simple regression. Multiple linear regression attempts to model the relationship between two or more explanatory variables and a response variable by fitting a linear equation to observed data. In the present work, multiple variables are used as predictor variables to fix the criterion variable in the linear regression model.

In order to study the modeling of groundwater contaminant using non-linear model, artificial neural network (ANN) and principal component-artificial neural network (PC-ANN) techniques are used (Sarala Thambavani and Uma Mageswari 2015). ANN is a massively parallel-distributed

✉ G. Charulatha
charunprakash@gmail.com

¹ Institute for Ocean Management, Anna University, Chennai, Tamil Nadu, India

² Department of Electronics and Communication Engineering, Anna University, Chennai, Tamil Nadu, India

³ Department of Chemistry, Government college of Engineering, Salem, Tamil Nadu, India

⁴ Department of Geology and Mining, Government of Tamil Nadu, Chennai, India

information processing system that has certain performance characteristics resembling biological neural networks of the human brain (Haykin 1999). Artificial neural networks have the ability to discover knowledge automatically using the functions and means of learning procedures analogous to the human brain and neural biology (Fausett 1994). ANN is a highly interconnected network of many simple processing units called neurons, which are analogous to the biological neurons in the human brain. The hierarchy structure of the network is constructed with the neurons having similar characteristics arranged in groups called layers.

The exacting arrangement of the processing elements and links produce a particular ANN model, suitable for the application of certain tasks. A multilayer perceptron (MLP) is a kind of feed-forward ANN model, consisting of three adjacent layers, viz., input, hidden, and output layers. The hidden layers are introduced to the network to increase its ability to model complex functions (Chauhan et al. 2010). The performance of ANN very much depends on its generalization capability, which in turn depends upon the data matrix. The key factor which plays a major role in the characteristic of data is non-correlation among the various variables.

In statistical means, the set of data presented to the multilayer perceptron may consist of non-correlated information which introduces confusion to the MLP during the learning process and hence, generalization capability becomes low to resolve the unseen data (Hornik 1991). This intricacy can be resolved by applying the principal component analysis (PCA) technique (Jolliffe 2002) onto input data sets prior to the multilayer perceptron training process as well as interpretation stage. PC-ANN is a chemometrics approach based on the combined use of the PCA and ANN. ANN models have been extensively functional to the water quality issues (Hornik 1991; Lee et al. 1996; Capolo et al. 1999; Chen and Huang 2001).

Various research studies have been carried out using the PCA as a preprocessor to the usual multilayer perceptron ANN in different fields. O'Farrella et al. (2005) applied ANN for classification and PCA in the preprocessing stage to classify the quality of food products by a feed-forward ANN with one hidden layer and the back propagation algorithm. Bucinski et al. (2005) combined PCA and ANN in medical field applying back propagation ANN to classify patients. Marengo et al. (2006) predicted polluting emissions from a cement production plant using principal component regression (PCR) and ANN. Sousa et al. (2007) use PCA analysis with varimax rotation to extract factors using ozone concentration data.

Liu and Yi (2007) applied a hybrid approach by integrating the ANN with the adaptive principal components extraction (APEX) algorithm. Ravi and Pramodh (2008) applied principal component neural network (PCNN) architecture to solve bankruptcy prediction problem in commercial banks. This

approach is becoming an effective and popular alternative for conventional methods (Yazdani et al. 2009; Cho et al. 2011; Ghasemloo et al. 2011). In the present work, in order to compare the performance of linear and non-linear models and to characterize their merits and demerits in the training and validation processes, one of the groundwater contaminants, viz., nitrite ion, is chosen and studied in detail.

Study area

The study area (south of Chennai) is located on the east coast of India and lies between $12^{\circ} 47''$ N $80^{\circ} 15''$ E and $13^{\circ} 00''$ N $80^{\circ} 05''$ E with an aerial extent of about 60 km^2 (Fig. 1). Chennai is one of the main metropolitan cities in South India and the capital of Tamil Nadu State. The population density in the city was 24,682 per km^2 in 2001, which makes it one of the most densely populated cities in the world. Chennai experiences a tropical wet and dry climate (Jayaprakash et al. 2012). In the last few decades, compared to central and northern part of Chennai, the study area visualizes fast urbanization and many industrial hubs especially IT industries were sprung up. Settlements from other districts and also from central and northern part of Chennai City are being attracted to this region. Hence, in the study area, both urbanization and industrialization are found to be in the rapid pace. South Chennai has emerged as an important center for economical, historical, cultural, and trade development in the state over the last few decades. The study area gets most of its seasonal rainfall from the northeast monsoon winds, during the period from October to December and the average annual rainfall is about 1300 mm (Arunprakash et al. 2014). The area comprises of coastal plains that prevails backwater zone in the north and categorized by sand dunes underlain by crystalline rocks of Archaean age.

In this region, groundwater movement is influenced by the easterly hydraulic gradient of approximately 0.004 with confined aquifer. The alluvium in the area has an average thickness of 15–20 m, underlain by crystalline rocks along the coast, shales, and clays of Gondwana to the west. The contact surface of the crystalline rocks and alluvium slopes in either direction of the coast forming a ridge-like structure (Ballukraya and Ravi 1998). The thickness of unconfined aquifer varies from few meters to as much as 40–50 m, with an average thickness of about 20 m. Variations in the thickness of this aquifer from east to west is not very significant since the levels of bedrock and Gondwana formations are generally at the same elevation. From the groundwater point of view, sandy formations underlying the dune/sand ridges along the coast are more prolific. The alluvium over the crystalline ridge is sandier than the deposits overlying the Gondwanas on the east.

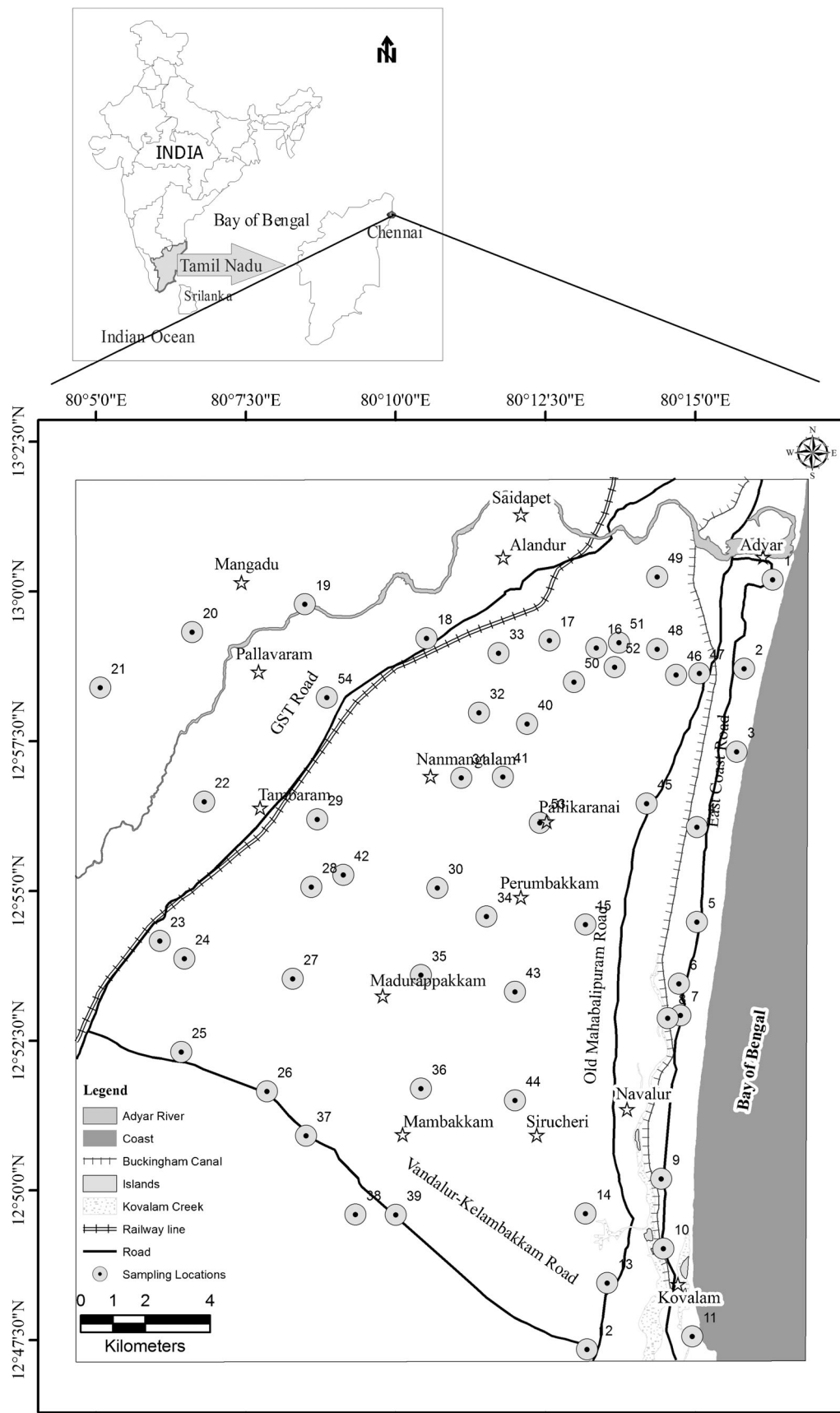


Fig. 1 Location map of the study area

Analytical methodology

Groundwater samples were collected during July 2013 (pre-monsoon) and January 2014 (post-monsoon). Samples were collected in new 1-L HDPE bottles pre-washed with distilled water, dried and rinsed three to four times with the water sample, and then labeled accordingly. Prior to analysis in the laboratory, these samples were stored at a temperature below 4 °C. A total of 54 samples for 9 water quality parameters were used for analysis. The parameters are temperature, electrical conductivity (EC), pH, total dissolved solids (TDS), nitrate, sodium, chloride, silicate, and fluoride. The standard methods of APHA 1995 were followed for collection, preservation, and analysis of the samples. Some of the parameters such as temperature, electrical conductivity, and pH of water samples were measured in the field immediately after the collection of the samples using pH and electrical conductivity meters. The pH meter was calibrated with reference buffer solutions of pH = 4 and 7 before each measurement. Evaporation and calculation methods of Hem 1991 were applied for the measurement of total dissolved solids (TDS). Ion chromatography (Metrohm883, Basic Plus) is used for the determination of all major cations and anions.

Statistical methodology

Multiple linear regression

Multiple linear regression (MLR) is a **multivariate statistical technique** for examining the **linear correlations** between two or more **independent variables** and a single **dependent variable**. The variable being predicted is generally referred as criterion variable and the variable which forms the basis for the predictions is referred as predictor variable. The dependent variable is referred as predictand, and the independent variables the predictors. MLR is based on least squares: the model is fit if the sum-of-squares of differences of observed and predicted values are minimized. In the present study, multiple variables of groundwater such as pH, TDS, nitrate, sodium, chloride, silicate, and fluoride are taken as predictors and the nitrite ion is taken as the criterion variable or the predictand. The following equation of MLR is used initially to find the best fit.

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$$

Since some of the chosen variables show deviation from linearity, best fit result could not be arrived and hence logarithmic transformation of the data is carried out and the results are found to be satisfactory. The MLR model used for the present study is

Fig. 2 a–h Diagrams depicting the comparison results between the observed and predicted nitrite concentrations of groundwater during pre-monsoon (a–d) and post-monsoon (e–h)

$$\log(y) = \beta_0 + \sum_{i=1}^n \beta_i \log x_i$$

where y is the dependent variable, x_i is the explanatory variable, β_i is the regression coefficient of explanatory variables, and β_0 is the value of the intercept in the log-linear fitting.

Principal component regression

Principal component regression (PCR) is a statistical method for analyzing multiple regression data that suffer from multicollinearity. Least squares estimates will be unbiased when multicollinearity occurs, but their variances will be large so they may be far from the true value. The standard errors in the principal component regression can be reduced by adding a degree of bias to the regression estimates. In the present case, multiple variables of groundwater such as pH, TDS, nitrate, sodium, chloride, silicate, and fluoride are taken as independent variables and the nitrite ion is taken as the dependent variable. PCR coalesce principal component analysis (PCA) decomposition with multiple linear regression (Jolliffe 2002). This generates a new set of variables otherwise called as principal components (PCs) and the orthogonal transformation of the original data generates principal component scores which are used as explanatory variables in the regression.

Artificial neural network

ANN is a computational model composed of a large number of highly interconnected neurons called the processing elements (PEs) with links between them. A configured arrangement of the PEs and links produces an ANN model, suitable for specific tasks. MLP is a kind of feed-forward ANN model consisting of three adjacent layers, viz., the input, hidden, and output layers. Multilayer perceptrons learn from input-output samples and they are capable of giving outputs based on inputs. MLP develops a mapping function between the inputs and outputs during the learning process which employs a learning algorithm. During this learning process, the input processing elements receive data from the external environment and pass them to the hidden PEs, which are responsible for mathematical computations involving the weights and the input values. The results from the hidden PEs are mapped onto appropriate threshold function of each PE and the final outputs are produced. The output result thus produced has then become the input in the adjacent layer and the computation process gets repeated and cease when an acceptably small error is achieved. In order to arrive the final acceptable small error value, the outputs are continuously computed. The

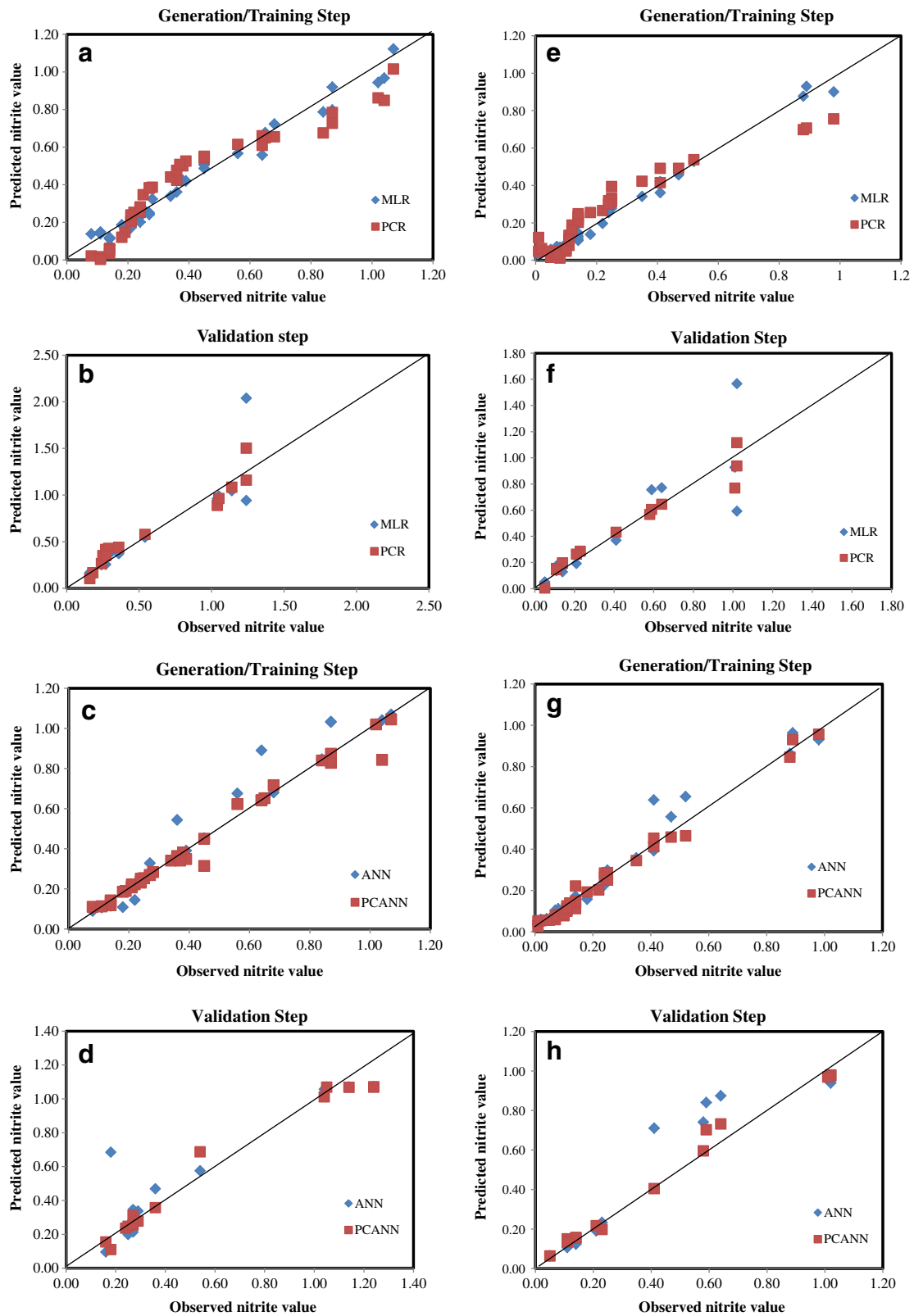


Table 1 Generation, training, and validation steps for pre-monsoon and post-monsoon samples

	Pre-monsoon					
	Generation/training steps			Validation steps		
	MAE ^a	NSE ^b	<i>R</i> (Maedeh et al. 2013)	MAE	NSE	<i>R</i> (Maedeh et al. 2013)
MLR	0.035554	0.976293	0.976	0.038948	0.699271	0.796
PCR	0.07652	0.893592	0.893	0.032886	0.929615	0.934
ANN	0.02841	0.943039	0.962	0.034808	0.860919	0.864
PC-ANN	0.017659	0.977401	0.979	0.019069	0.962834	0.971
Post-monsoon						
MLR	0.023114	0.984315	0.984	0.039525	0.688043	0.772
PCR	0.063935	0.877899	0.888	0.020512	0.946454	0.950
ANN	0.029522	0.952499	0.962	0.030631	0.85885	0.888
PC-ANN	0.018399	0.987961	0.988	0.012422	0.98245	0.984

^a Mean absolute error^b Nash–Sutcliffe model Efficiency coefficient

difference between the MLPs and the actual outputs and the entire training process is iterative in nature.

Principal component-artificial neural network

PC-ANN has some additional features and advantages over ANN in certain research fields. In the statistical analysis of the data, in certain research areas, the variables in the data may not have correlation and in certain cases, overlapping of the data is also observed. In such cases, the data presented to the multilayer perceptron may introduce misperception during the learning process which can be resolved by the application of PCA technique onto the data sets prior to the MLR training process. PC-ANN amalgamates PCA decomposition with ANN. The raw input variables from the analytical results of the groundwater are applied and the principal component scores thus generated from the orthogonal linear transformation are used as the input variables in the ANN. The output thus generated is referred as PC-ANN output.

Results and discussions

MLR for predicting the nitrite ion concentration in the ground water

MLR results for prediction of the nitrite concentration in the ground water for both generation and validation step for pre-monsoon and for post-monsoon is shown in Fig. 2a–h. The performance measures such as the generated regression coefficient b_i , standard error SE_{b_i} , and variance inflation factor (VIF) for both pre-monsoon and post-monsoon for the MLR model is presented in Table 1. When $b_i > 2SE_{b_i}$, then the variable is considered as

significant (Rawlings et al. 1998). In this work, during pre-monsoon, sodium and chloride ion is considered as significant and in post-monsoon nitrate, sodium and chloride is considered as significant. The modeling accuracy of various models is compared using Nash–Sutcliffe efficiency (NSE) coefficient test which is calculated using the predicted value and the observed value.

In general, NSE value can be divided into two categories, viz., NSE value < 0.5 of the predicted value is considered as inaccurate and the model will be rejected, while NSE value between 0.5 and 0.7 gives an acceptable predicted value, but when the NSE value > 0.7 indicates that the predicted value is in good agreement with the observed value. When the NSE values are nearer to 1, then the model is considered as accurate. The results suggest that overall MLR shows good coefficient of determination for the generation step and the model shows a good fit which is again proved by the NSE value which is closer to 1 (Table 1). The trend is found to be the same for both pre-monsoon and post-monsoon data. But in the validation, MLR shows poor results with *R* (Maedeh et al. 2013) value for both pre-monsoon and post-monsoon with values less than 0.8 and NSE value less than 0.7. Hence, in the validation step, the developed MLR model tends to underestimate the nitrite concentration. The collinearity statistics of MLR model is studied by following variance inflation factor (VIF). The VIF value for both the pre-monsoon and post-monsoon was found to be very high with an average greater than 30 which indicates that the input variables strongly rely on the degree of its correlation with other variables (Bowerman and O'Connell 1990; Myers 1990).

PCR for predicting nitrite ion concentration in the ground water

Seven PCs are extracted from the input data, but only three PCs are taken for linear regression, for both pre-monsoon and

Table 2 Factor component matrix for pre-monsoon data

Parameters	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6
TDS	0.995	-0.025	-0.01	-0.022	0.056	0.064
Nitrate	0.99	-0.074	0.019	-0.006	-0.098	-0.059
Na	0.988	-0.066	0.022	-0.007	-0.12	0.03
Cl	0.984	-0.034	-0.036	-0.01	0.162	-0.035
pH	0.155	0.865	-0.286	0.382	-0.006	0
Silicate	-0.052	-0.916	0.037	0.395	0.007	0.002
F	0.054	0.298	0.948	0.1	0.01	0
Eigenvalue	3.944	1.689	0.983	0.313	0.053	0.01
% variance	56.343	24.126	14.049	4.465	0.763	0.138
Cumulative %	56.343	80.469	94.519	98.983	99.747	99.885

post-monsoon. Eigenvalue is considered for choosing PCs and PCs with Eigenvalues greater than 1 is considered for the analysis (Tables 2 and 3). The three PCs for both the monsoon data has a cumulative percentage of more than 93 implying that these three PCs almost explain and represent the whole data matrix. The first principal component of both pre-monsoon and post-monsoon contributes more than 50% of the total variance and also the component values greater than 0.9 for the major ions TDS, Nitrate, Na, and Cl. In the pre-monsoon, silicate and pH have high loadings in PC-2 and fluoride have high loadings in PC-3.

In the post-monsoon, contribution of pH towards PC-2 is reversed with a high negative loading and the rest of the parameter contribution other than pH is same as that of pre-monsoon. Figure 2a, b illustrates the comparison of observed and predicted nitrite concentration in the groundwater of the study area during pre-monsoon and post-monsoon for both generation and validation steps. Table 1 shows the performance measure values such as regression coefficient (*R*) (Maedeh et al. 2013), mean absolute error (MAE), and NSE for both the monsoons. The results illustrate that the PCR model show high value of *R* (Maedeh et al. 2013) and NSE

for the validation step, but in the case of the generation step, the values are deprived indicating that the PCR model fits well for the validation step but does not work well in the case of generation step. If a mathematical model is superior, it should fit both generation and validation steps. It is interesting to note the change in the collinearity data variance inflation factor (VIF). The average VIF is nearly 1 in both the monsoon data indicating that there is a tremendous change from the MLR which showed a value greater than 30. The principal component dependence on one another is considerably less than the component in MLR.

ANN model for predicting nitrite ion concentration in the ground water

Pattern search algorithm is used for determining the number of hidden layers, number of neurons in the hidden layer, optimum momentum rate, and the learning rate for the ANN modeling. Normally, the number of neurons in the hidden layer for the ANN model is fixed 2 to 3 times the number of inputs in the modeling (Brion and Lingireddy 1999). When the number of hidden layer neurons is too small, then the

Table 3 Factor component matrix for post-monsoon data

Parameters	Component 1	Component 2	Component 3	Component 4	Component 5	Component 6	Component 7
TDS	0.981	0.133	0.028	0.001	-0.011	-0.108	0.09
Nitrate	0.978	0.125	-0.036	0.007	-0.085	0.128	0.044
Na	0.974	0.167	0.029	0.005	-0.096	-0.046	-0.105
Cl	0.971	0.134	-0.02	0.003	0.193	0.027	-0.028
pH	0.304	-0.854	0.061	0.418	-0.001	-0.002	-0.002
Silicate	-0.339	0.84	0.007	0.424	0.002	0.001	0.003
F	-0.017	0.045	0.998	-0.028	0.004	0.01	0.002
Eigenvalue	4.018	1.516	1.004	0.355	0.054	0.031	0.022
% variance	57.406	21.656	14.341	5.071	0.77	0.444	0.312
Cumulative %	57.406	79.062	93.403	98.474	99.244	99.688	100

prediction from the ANN would be inaccurate and when the number of hidden layer neurons is too high, more computational time would be required to finish the modeling.

Hence, for both ANN and PC-ANN, the number of neurons in the hidden layer is fixed at 15. Levenberg-Marquardt feed-forward back propagation training algorithm is used in this study. The transfer function “tan-sigmoid” is used in neurons of the hidden and output layers. The ANN were trained, tested, and validated. Figure 2a–h illustrates the comparison of observed and predicted nitrite concentration in the ground water during pre-monsoon and post-monsoon periods for both generation and validation steps. Table 1 represents the R (Maedeh et al. 2013) and the NSE value. Results show that the NSE value for the generation step is nearly 0.95 for both the pre-monsoon and post-monsoon with R (Maedeh et al. 2013) values greater than 0.96 which indicates that the ANN model has accurate prediction value in the generation step (Table 1). But the NSE and R (Maedeh et al. 2013) value for both the monsoon is found to be less than 0.9 for validation step. This implies that ANN is able to predict accurately during the generation step but it is less accurate for the validation step.

PC-ANN model for predicting nitrite ion concentration in the ground water

Figure 2a–h represents the comparison of observed and predicted nitrite concentration by the PC-ANN model on the ground water for pre-monsoon and post-monsoon for both generation and validation steps. Table 1 signifies the R (Maedeh et al. 2013) and the NSE value. The values of NSE and R (Maedeh et al. 2013) for both the monsoons and for both the training and validation steps show more than 0.96, which implies that PC-ANN model fits for both the generation and validation steps.

Comparison between MLR, PCR, ANN, and PC-ANN

The results clearly illustrate that the R (Maedeh et al. 2013) and NSE values for both validation and generation steps for both the monsoon data is observed to be high for PC-ANN compared to other models. The values are found to be greater than any other models studied in this work. The results also imply that the non-linear model is more suitable for the determination of nitrite concentration in groundwater rather than linear model such as MLR or PCR. From the results, it can be inferred that the MLR gives better value for the training/generation step while in validation step, it does not give an accurate result.

On the other hand, PCR does not yield a good result in the generation step but provides a better result in the validation phase. ANN performs very well as far as accuracy is

concerned during validation step. But when ANN run with the three PCs obtained from the PCA, the results were found to be more accurate with the observed value. The results clearly demonstrate that the highest accuracy in prediction of nitrite concentration in the groundwater of the study area is given by the model PC-ANN. The best performance of PC-ANN model is clearly illustrated from the correlation coefficient values 0.979 (training) and 0.971 (validation) for pre-monsoon and 0.988 (training) and 0.984 (validation) for post-monsoon. The PC-ANN non-linear mathematical model shows more effective results than MLR or PCR.

Conclusions

The present study on the groundwater contaminant clearly authenticates the ability of the linear and non-linear models in the training and validation steps. The MLR model shows good determination of correlation coefficient 0.976 (pre-monsoon) and 0.984 (post-monsoon) for the generation step and the model shows a good fit which is again proved by the NSE value which is closer to 1. In the validation phase, MLR shows poor results with R (Maedeh et al. 2013) values 0.796 and 0.772 displaying values less than 0.8 for both pre monsoon and post monsoon respectively, and NSE value less than 0.7 for both monsoons. The PCR model illustrates that the model shows high value for both R (Maedeh et al. 2013) and NSE for the validation step, but in the case of the generation step, the values are found to be poor indicating that the PCR model fits well for the validation step but does not work well in the case of generation step. The results of ANN model show that the NSE value for the generation step is nearly 0.95 for both pre-monsoon and post-monsoon with R (Maedeh et al. 2013) values greater than 0.96 which indicates that the ANN model has accurate prediction in the generation step. But the NSE and R (Maedeh et al. 2013) value for both the monsoon is found to be less than 0.9 for validation step. This implies that ANN is able to predict more accurately during the generation step but in the validation step, the accuracy is found to be less. The PC-ANN model reveals that the values of NSE and R (Maedeh et al. 2013) for both the monsoons and for both the training and validation steps are greater than 0.96, which implies that PC-ANN model fits for both the generation and validation steps. Finally, the results clearly exemplify that the R (Maedeh et al. 2013) and NSE values for both validation and generation steps for both the monsoon data are observed to be high for PC-ANN compared to other models. The PC-ANN non-linear model exhibited better prediction results for predicting nitrite concentration in groundwater rather than the linear models such as MLR or PCR.

References

- Arunprakash M, Giridharan L, Krishnamurthy RR, Jayaprakash M (2014) Impact of urbanization in groundwater of south Chennai City, Tamil Nadu, India. *Environmental Earth Sciences* 71(2):947–950
- Ballukraya PN, Ravi R (1998) Natural fresh water ridge as barrier against sea water intrusion in Chennai City. *J Geol Soc India* 52(3):279–286
- Bowerman BL, O'Connell RT (1990) *Linear statistical models: an applied approach*, Second edn. Duxbury Press, Belmont, CA
- Brion GM, Lingireddy S (1999) A neural network approach to identifying non-point sources of microbial contamination. *Water Res* 33(14):3099–3106
- Bucinski A, Baczek T, Wasniewski T, Stefanowicz M (2005) Clinical data analysis with the use of artificial neural networks (ANN) and principal component analysis (PCA) of patients with endometrial carcinoma. *Reports on Practical Oncology and Radiotherapy* 10:239–248
- Capolo M, Andreussi P, Soldati A (1999) River filled forecasting with a neural network model. *Water Resource Res* 35:1191–1197
- Chauhan S, Sharma M, Arora MK, Gupta NK (2010) Landslide susceptibility zonation through ratings derived from artificial neural network. *Int J Appl Earth Obs Geoinf* 12:340–350
- Chen RM, Huang YM (2001) Competitive neural network to solve scheduling problems. *Neurocomputing* 37:177–196
- Cho KH, Sthiannopkao S, Pachepsky YA, Kim KW, Kim JH (2011) Prediction of contamination potential of groundwater arsenic in Cambodia, Laos, and Thailand using artificial neural network. *Water Res* 45(17):5535–5544
- Fausett L (1994) *Fundamentals of neural network architecture, algorithms, and applications*. Prentice-Hall, UK
- Ghasemloo N, Mobasheri MR, Rezaei Y (2011) Vegetation species determination using spectral characteristics and artificial neural network (SCANN). *J Agr Sci Tech* 13(Suppl):1223–1232
- Haykin S (1999) *Neural networks: a comprehensive foundation*. Macmillan College, London
- Hornik K (1991) Approximation capability of multi-layer feedforward networks. *Neural Netw* 4:241–257
- Jayaprakash M, Nagarajan R, Velmurugan PM, Sathiyamoorthy J, Krishnamurthy RR, Urban B (2012) Assessment of trace metal contamination in a historical freshwater canal (Buckingham Canal), Chennai, India. *Environ Monit Assess* 184(12):7407–7424
- Jolliffe IT (2002) *Principal components analysis*. Springer-Verlag, New York, p 488p
- Lee I M, Lee J H (2006) Prediction of pile bearing capacity using artificial neural networks. *Comput Geotech* 18(3):189–200
- Liu G, Yi Z, Yang S (2007) A hierarchical intrusion detection model based on the PCA neural networks. *Neurocomputing* 70:1561–1568
- Maedeh A, Mehrdadi N, Nabi Bidhendi GR, Zare Abyaneh H (2013) Application of artificial neural network to predict total dissolved solids variations in groundwater of Tehran plain. *Iran Int J Environment and Sustainability* 2(1):10–20
- Marengo E, Bobba M, Robotti E, Liparota MC (2006) Modeling of the polluting emissions from a cement production plant by partial least-squares. *Principal Component Regression and Artificial Neural Networks Environ Sci Technol* 2006(40):272–280
- Myers R (1990) *Classical and modern regression with applications*, Second edn. Duxbury Press, Boston, MA
- O'Farrella M, Lewisa E, Flanagan C, Lyonsa WB, Jackman N (2005) Combining principal component analysis with an artificial neural network to perform online quality assessment of food as it cooks in a large-scale industrial oven. *Sensors Actuators B* 107:104–112
- Ravi V, Pramodh C (2008) Threshold accepting trained principal component neural network and feature subset selection: application to bankruptcy prediction in banks. *Appl Soft Comput* 8:1539–1548
- Rawlings JO, Pantula SG, Dickey DA (1998) *Applied regression analysis: a research tool*. Springer, New York
- Sarala Thambavani D, Uma Mageswari TSR (2015) Comparative application of ANN and PCA in modeling of ground water. *J Adv Chem Sci* 1(1):22–26
- Sousa SIV, Martins FG, Alvim-Ferraz MCM, Pereira MC (2007) Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ Model Softw* 22:97–103
- Yazdani MR, Saghafian B, Mahdian MH, Soltani S (2009) Monthly runoff estimation using artificial neural networks. *J Agr SciTech* 11(3):355–362