Groundwater quality prediction using Bayesian automatic relevance determination modeling

Anasuya Das*, and Saumen Maiti,
Department of Applied Geophysics, Indian Institute of Technology (Indian School of Mines), Dhanbad-826004, India
Email: saumen_maiti2002@yahoo.co.in
Fax: +91-326-2296618

Summary

In this study, an application of automatic relevance determination-based Bayesian neural network (ARD-BNN) approach is explored to quantify the influence of water quality variables for groundwater quality modelling of coastal Maharashtra, India. The ARD-BNN model is trained using training samples generated from the bounds of water quality parameters. Upon successful training, the trained model is tested in validation and test data sets. Since the value of hyper-parameters in ARD-BNN modelling is important for categorizing the water quality variables, the stability of the hyper-parameters is assessed by optimization test history. The groundwater quality map will be useful for management of the groundwater system of coastal Maharashtra. The novel application of this type could be potential alternative scheme for analysing complex/fuzzy, interlinked water quality variables collected around the world.

Keywords: Bayesian neural networks, Automatic relevance determination, Water quality parameter, Groundwater quality status, Konkan.

Introduction

Groundwater is the most important natural resource used for drinking, irrigation and variety of other domestic purposes. Groundwater quality has been under risk due to growing pollution of freshwater lenses by various pollution sources (e.g. agriculture, industry, human societies, sea water intrusion) (Brown et al. 1974; APHA 1985; Das et al. 2016). There are various factors which usually control groundwater quality. The common factors are evolved as: (i) local geology of the study area,(ii) degree of chemical weathering of the rocks, (iii) rock-water interaction process, (iv) evaporation, (v) industrial return,(vi) irrigation return flow, (vii) hydrological changes during a season (Todd 1959; Hem 1970; Fetter 1988; Maiti et al. 2013). Moreover, the organic matters derived from various fertilizer, pesticides and other additive applied in agricultural lands are also known to be potential sources in degrading groundwater quality (Brown *et al.* 1974; APHA 1985). Further, over-exploitation of groundwater due to rapid growth of industrialization, urbanization and human population lead to the deterioration of groundwater quality (Brown *et al.* 1974; APHA 1985). The optimal use of groundwater resource and its sustainability is guided by the accurate assessment of trends in groundwater quality. Many authors have attempted to model the groundwater quality (Brown 1998;Maiti et al. 2013; Stedmon et al. 2011; Najah et al. 2013; Maiti et al. 2012; Sadat-Noori et al. 2014). Evaluating groundwater quality of an area is not a new problem, but it is still unresolved problem to quantify individual's impact of water quality parameter on groundwater quality. Since many of the water quality parameters are interlinked, comprehensive efforts are required to better understand the individual's impact for groundwater management in any region.

Here we explore the specific use of ARD-BNN to model relative contribution for estimation of groundwater quality. The use of ARD-BNN for groundwater quality estimation is the first study in its type in India. The objective of this study were to (i) find the relative weight of the water quality parameter by Bayesian automatic relevance determination method(ii) Predict the water quality status.

Materials & Method

The present study area in the Sindhudurg district, western Maharashtra, India is demarcated by Latitude $15.7^0 - 15.9^0$ N and Longitude $73.6^0 - 73.8^0$ E (Figure 1)(CGWB 2009). Geo-electrical survey along with hydro-geochemical sample collection was carried out in the study area. Water samples were collected from the study area in pre-monsoon season. In total 36 sampling sites were chosen, which were evenly distributed according to the complaints made by the local villagers (one sample each which provided a total of 36 sampling sites altogether). Water samples were collected in 1.0 litre plastic bottles. Water sample analysis was performed using a widely used method (Brown et al. 1974; APHA 1985).

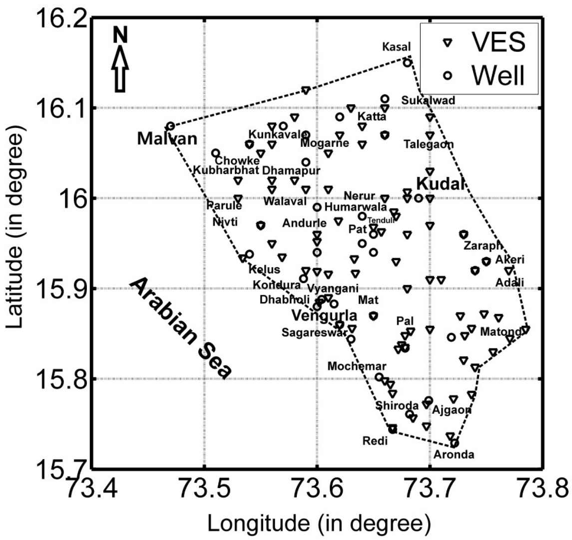

Figure 1: The location map of the study area. The location of vertical electrical sounding (VES) and groundwater sampling from respective well (Well) are given in the map

A joint information system including geo-electrical/VES and hydro-geochemical is combined to evaluate groundwater quality. As per report (WHO 1984), given in Table 1, groundwater quality of coastal Maharashtra could be classified into three classes (1) very good (assigned real code is 1.00), (2) good (assigned real code is 0.50), and (3) unsuitable (assigned real code is 0.00) (Maiti et al. 2013). We use ARD-BNN to estimate the individual water quality parameter's contribution in evaluating groundwater quality condition.

Multi-Layer Perceptron (MLP) is a special configuration of an artificial neural networks (ANN) (Bishop 1995;Karmakar et al. 2017) that has been widely used to solve non-linear mapping problem using examples.

Table 1: The geochemical and geophysical parameters used for ARD-BNN forward modelling.

Water quality parameters	Very Good (1.00)	Good (0.50)	Unsuitable (0.00)	Permissible limit (WHO 1984)
pH(-log10H^+)	7.51-8.5	7.1-7.5	0.01-7.0 and 8.51-14.0	7.5-8.5
EC(µS/cm)	0.01-750	751-1500	1501-3000	750
TDS(mg/l)	0.01-500	501-1500	1501-3000	1000
$HCO3^-$ (mg/l)	0.01-200	201-300	301-500	300
Cl^-(mg/l)	0.01-100	101-200	201-1000	200
$SO4^{2-}$ (mg/l)	0.01-100	101-200	201-500	200
$NO3^-$ (mg/l)	0.01-25	25-45	45-100	45
Ca^{2+}(mg/l)	0.01-50	51-75	76-500	75
Mg^{2+}(mg/l)	0.01-20	21-30	31-200	30
Na^+(mg/l)	0.01-50	51-200	201-500	200
K^+(mg/l)	0.01-10	11-100	101-200	100
ρ (Ohm-m)	11-1000	1001-40000	0.01-10	___

The underlying process sets up objective function from a finite data set $z = \{x_k, y_k; k = 1,...N\}$ in the form

$$E_Z = \frac{1}{2}\sum_{k}^{N}\left\{y_k - o_k(x_k; w_k)\right\}^2 \qquad (1)$$

where, y_k and o_k are respectively the target/desired and the actual network output at each node in the output layer .

Often, regularization is incorporated to smooth the error function

$$E(w) = \mu E_Z + \lambda E_R \qquad (2)$$

where λ and μ which control other parameters (synaptic weight and biases) known as hyper-parameters. For example, $E_R = \frac{1}{2}\sum_{i=1}^{R} w_i^2$ is known as weight-decay regularization function and R is the total number of weights and biases in the network.

ARD-BNN which is a special configuration of a Bayesian neural network approach automatically restricts the influence of irrelevant inputs for making improved prediction/classification (Calster et al. 2008). Moreover, ARD-BNN allows accessing the relative impact of all input variables in output prediction/classification quantitatively. In the plain BNN approach (in eq. 2), it is assumed that the general state of prior information about weight of input and hidden layers of the network (Maiti and Tiwari 2009). This is attributed to application of same restriction to all weights (input and hidden layers) which is unfavourable to scaling properties of network (Bishop 1995).

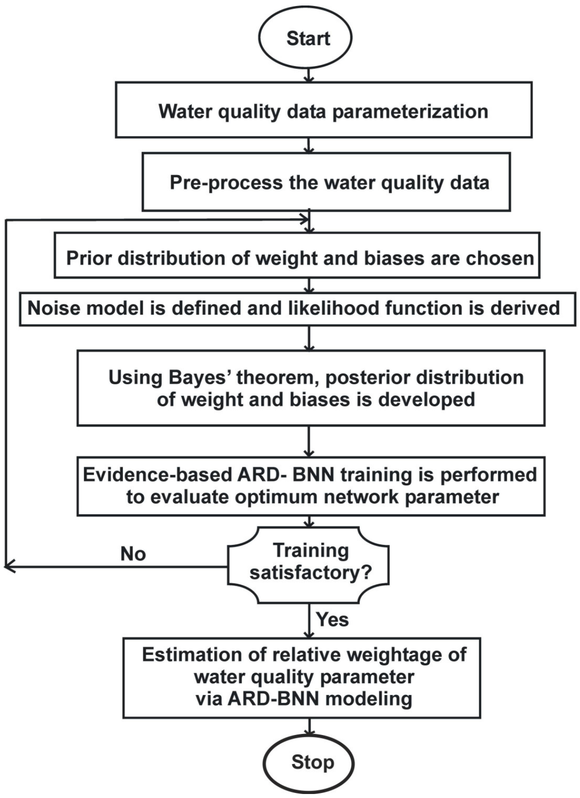

Figure 2: Description of flowchart.

Therefore, input layer weight (w) and hidden layer weight (v) could be controlled by different λ values, which lead to the following expression (Nabney 2004; Maiti and Tiwari 2010)

$$E(w) = \mu E_Z + \lambda E_R$$

$$= \frac{\mu}{2}\sum_{k=1}^{N}\{O(x_k;w) - y_k\}^2 \qquad (3)$$

$$+ \frac{\lambda_1}{2}\sum_{i=1}^{w} w_i^{\,2} + \frac{\lambda_2}{2}\sum_{i=1}^{v} v_i^{\,2}$$

Likewise the λ value can further be partitionioned conceptualizing that each input weight is controlled by their individual λ. The magnitude of λ would control the value of the corresponding weight. If the value of λ is large, optimization process would control the weight to be very small and the particular input would therefore have low relevance to the output prediction (Nabney, 2004). Total methodology of the present work is presented in the flowchart (Fig. 2).

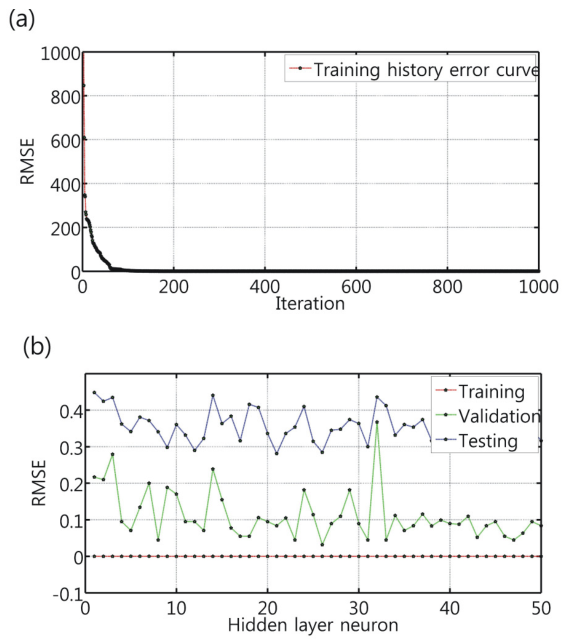

Figure 3: (a) Training history of automatic relevance determination-based Bayesian neural network (b)Root-Mean Squared Error (RMSE) history in training, validation and testing interval.

The total data set (1500) is divided into three parts. Training data is composed of 50% of the total data (i.e. 750). Out of the rest 50% of the total data, 25% (i.e. 375) data is kept for validation of the network and rest 25 % (i.e. 375) is reserved for testing the network.

The performance of ARD-BNN model was evaluated by value of Pearson correlation coefficient (r), root mean squared error (RMSE), reduction of error (RE), and index of agreement (IA) (Maiti et al. 2013).The successful calibration of the model is assessed by the model performance in two independent data sets, called validation and test data sets. Table 2 shows that the trained network gives satisfactory prediction results in validation and test data sets.

Table 2: Performance of ARD-BNN for groundwater quality classification

Data set	r	RMSE	RE	IA
Validation data	0.91	1.2	0.98	0.95
Testing data	0.90	1.4	0.97	0.93

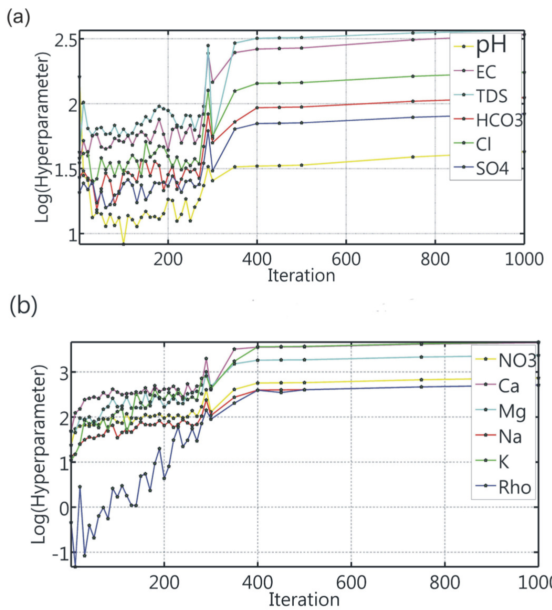

Figure 4(a-b): The stability of water quality parameter for modeling and quantifying automatic relevance to the groundwater quality prediction.

It can be learned from figure 4a-b that consistency of logarithmic values of hyper-parameters is likely to occur when the training iteration is beyond 500. The output of ARD-BNN is depicted in Figure 5. The output value close to 1.0 may be interpreted as "very good" class whereas ARD-BNN output value close to 0.0 may be declared un-fit for drinking. It is found that most of the places in coastal Maharashtra is likely to have 'good' to 'excellent' quality of groundwater except Kelus and Shiroda (Figure 5), which is in agreement with HMC-BNN based results (Maiti et al. 2013; their Fig.7). The deplorable condition of groundwater quality at Kelus and Shiroda is possibly due to the effect of saline water intrusion from the Arabian Sea. It is recognized that pollution of coastal groundwater system is mainly attributed to saltwater intrusion.

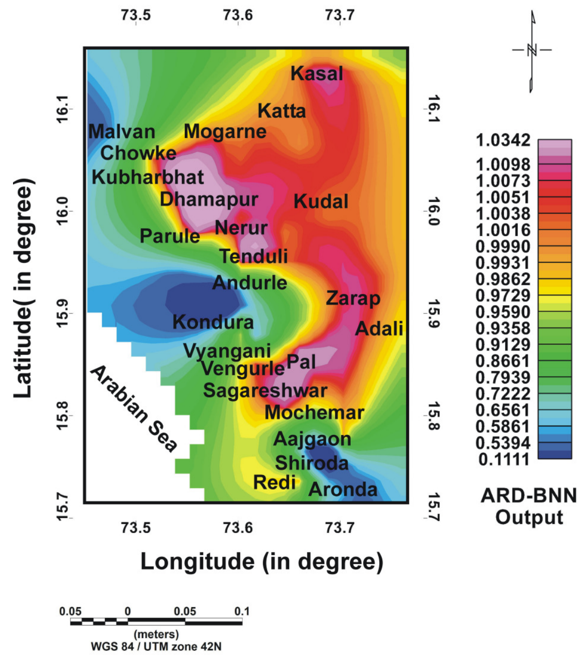

Figure 5: Automatic relevance determination-based Bayesian neural network prediction of groundwater quality status of the study area. The output value close to 1.0 is considered as excellent and the value close to 0.0 is considered as unfit for drinking.

Conclusions

The automatic relevance determination based Bayesian neural network (ARD-BNN) approach is able to predict the groundwater quality of coastal Maharashtra, India. Once the network is properly trained with global training data set, it is able to predict the groundwater quality instantaneously from new data sets. The present results also indicate that most of the places except Kelus and Shiroda are characterized to have good quality groundwater.

References

American Public Health Association APHA. 1985 *Standard methods for the examination of water and waste.* 16th edition. Washington, DC, USA: Am. Public Health Assoc., p. 100

Bishop, C. M. 1995 Neural networks for pattern recognition. Oxford University Press.

Brown, E. Skougstad, M. W. & Fishmen, M. J. 1974 *Methods for collection and analysis of water samples for dissolved minerals and gases*. Govt. Printing Office. Washington.

Calster, B.V. Timmerman, D. Nabney, I.T. Valentin, L. Testa, A.C. Holsbeke, C.V. Vergote, I. Huffel, S.V. 2008 Using Bayesian neural networks with ARD input selectionto detect malignant ovarian masses prior to surgery. Neural Comput. & Application, 17,489-500.

Central Ground Water Board (CGWB). 2009 Groundwater information Sindhudurg district Maharashtra, p.16.

Das, A. Maiti, S. Naidu, S. & Gupta, G. 2016 Estimation of spatial variability of aquifer parameters from geophysical methods: A case study of Sindhudurg district, Maharashtra, India, Stochastic Environmental Research and Risk Assessment, http://dx.doi.org/10.1007/s00477-016-1317-4

Fetter, C.W. 1988 Applied hydrogeology. Charles E. Merrill,Columbus.

Hem, J.D. 1970 Study and interpretation of the chemical characteristicsof natural water. Water Supply Paper-1473, US Geological Survey, pp 363

Karmakar, M., Maiti, S., Singh, A., Ojha, M., Maity, B., 2017 Mapping of rock types using a joint approach by combining the multivariate statistics, self-organizing map and Bayesian neural networks: an example from IODP 323 site, Marine Geophysical Research, http://dx.doi.org/10.1007/s11001-017-9327-2

Maiti, S., Erram, V.C., Gupta, G., & Tiwari, R.K., 2012 ANN based inversion of DC resistivity data for groundwater exploration in hard rock terrain of western Maharashtra (India). Journal of Hydrology, 464-465,281-293, http://dx.doi.org/10.1016/j.jhydrol.2012.07.020.

Maiti, S., Erram, V.C., Gupta, G., Tiwari, R.K., Kulkarni, U.D., & Sangpal, R.R., 2013 Assessment of groundwater quality: A fusion of geochemical and geophysical information via Bayesian Neural Networks, Environmental Monitoring and Assessment, 185, 3445-3465, http://dx.doi.org/10.1007/s10661-012-2802-y

Maiti, S. & Tiwari, R.K. 2009 A hybrid Monte Carlo method based artificial neural networks approach for rock boundaries identification: A case study from the KTB bore hole. Pure and Applied Geophysics, 166, 2059-2090, doi 10.1007/s00024-009-0533-y.

Maiti, S. & Tiwari, R.K. 2010 Neural network modeling and an uncertainty analysis in Bayesian framework: A case study from the KTB borehole site. Jour. Geophys. Res., 115, B10208, doi:10.1029/2010JB000864.'

Najah, A. El-Shafie, A. Karim, O.A. El-Shafie, A.H. 2013 Application of artificial neural networks for water quality prediction. Neural Compu & Applic, 22, S187-S201,doi 10.1007/s00521-012-0940-3.

Nabney, I. T, 2004 Netlab Algorithms for pattern recognition. Springer, New York.

Stedmon, C. A. et al. 2011 A potential approach for monitoring drinking water quality from groundwater systems using organic matter fluorescence as an early warning for contamination events. Water Res. 45, 6030-6038

Todd, D.K, 1959 *Groundwater hydrology*. New York: Wiley, p. 277-294.

World Health Organization (WHO) 1984 *Guideline of drinking quality*, Washington, DC, USA: World Health Organization, p. 333-335.

Acknowledgments

Authors are thankful to Director, IIT (ISM), Dhanbad for his kind permission to publish the work. AD is thankful to IIT (ISM) SRF fellowship and Gautam Gupta of IIG, New Panvel for many fruitful discussions. SM is thankful to the Ministry of Earth Sciences, Govt. of India, New Delhi, India, and acknowledge the grant (Grant No: MoES/P.O. (Geosci)/44/2015).