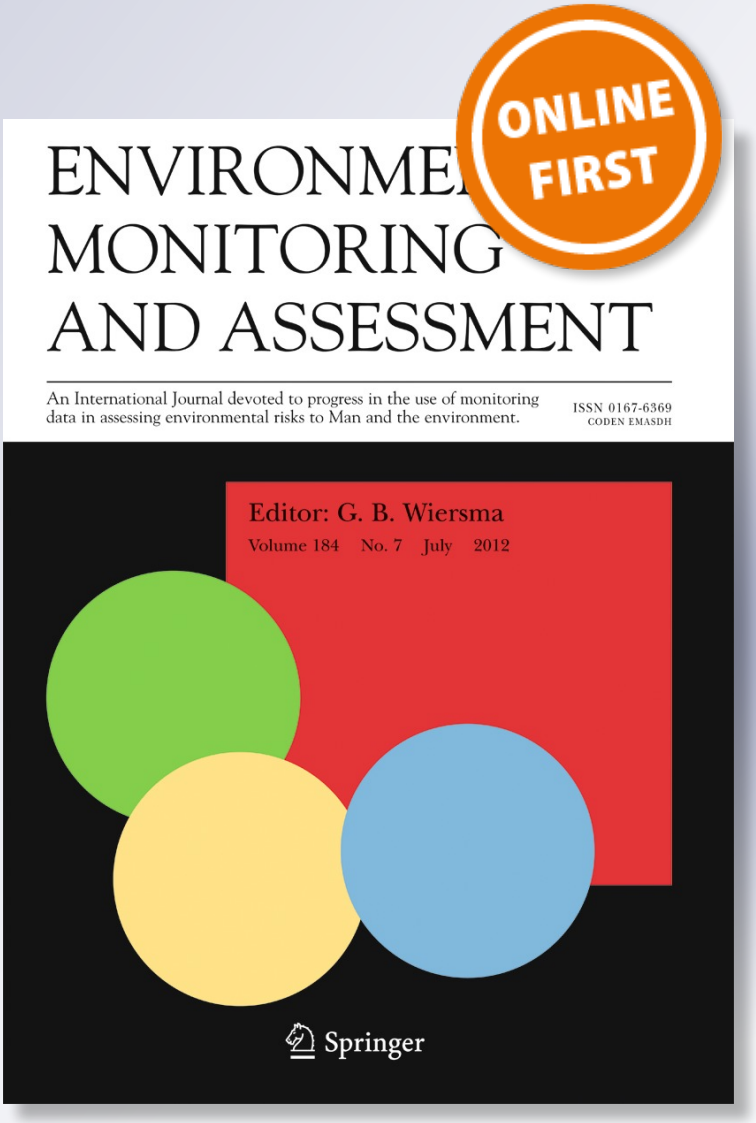
Assessment of groundwater quality: a fusion of geochemical and geophysical information via Bayesian neural networks

Saumen Maiti, V. C. Erram, Gautam Gupta, Ram Krishna Tiwari, U. D. Kulkarni & R. R. Sangpal

Environmental Monitoring and Assessment
An International Journal Devoted to Progress in the Use of Monitoring Data in Assessing Environmental Risks to Man and the Environment

ISSN 0167-6369

Environ Monit Assess
DOI 10.1007/s10661-012-2802-y



Springer

Environ Monit Assess
DOI 10.1007/s10661-012-2802-y

Assessment of groundwater quality: a fusion of geochemical and geophysical information via Bayesian neural networks

Saumen Maiti · V. C. Erram · Gautam Gupta · Ram Krishna Tiwari · U. D. Kulkarni · R. R. Sangpal

Received: 17 February 2012 / Accepted: 23 July 2012

Abstract Deplorable quality of groundwater arising from saltwater intrusion, natural leaching and anthropogenic activities is one of the major concerns for the society. Assessment of groundwater quality is, therefore, a primary objective of scientific research. Here, we propose an artificial neural network-based method set in a Bayesian neural network (BNN) framework and employ it to assess groundwater quality. The approach is based on analyzing 36 water samples and inverting up to 85 Schlumberger vertical electrical sounding data. We constructed a priori model by suitably parameterizing geochemical and geophysical data collected from the western part of India. The posterior model (post-inversion) was estimated using the BNN learning procedure and global hybrid Monte Carlo/Markov Chain Monte Carlo optimization scheme. By suitable parameterization of geochemical and geophysical parameters, we simulated 1,500 training samples, out of which 50 % samples were used for training and remaining 50 % were used for validation and testing. We show that the trained model is able to classify validation and test samples with 85 % and 80 % accuracy respectively. Based on cross-correlation analysis and Gibb's diagram of geochemical attributes, the groundwater qualities of the study area were classified into following three categories: "Very good", "Good", and "Unsuitable". The BNN model-based results suggest that groundwater quality falls mostly in the range of "Good" to "Very good" except for some places near the Arabian Sea. The new modeling results powered by uncertainty and statistical analyses would provide useful constrain, which could be utilized in monitoring and assessment of the groundwater quality.

Keywords Bayesian neural networks · Hydrogeochemistry · DC resistivity · Groundwater quality · Uncertainty analysis · Coastal Maharashtra

S. Maiti (✉) · V. C. Erram · G. Gupta
Indian Institute of Geomagnetism (DST),
Navi Mumbai 410218, India
e-mail: saumen_maiti2002@yahoo.co.in

R. K. Tiwari
CSIR—National Geophysical Research Institute,
Hyderabad 500606, India

U. D. Kulkarni · R. R. Sangpal
Nowrosjee Wadia College of Science,
Pune 411001, India

Introduction

Groundwater is regarded as one of the major sources of drinking water for the society. Groundwater is also used for agricultural, industrial, and various other domestic purposes. However, intrusion of saline water, particularly in the coastal areas, creates a major concern for its safe uses (Park et al. 2005; Mondal et al. 2011). In addition to this, various harmful contaminants, such as agricultural fertilizer, pesticides, waste

Published online: 17 August 2012

deposits, industrial effluents, etc. carried by surface water, further modify its chemical composition and concentration (Kannan and Joseph 2010; Prasanna et al. 2011). The chemical composition of groundwater is controlled by multiple processes, viz. (1) atmospheric inputs, (2) dynamic interaction of water with soil and rock, and (3) anthropogenic activities (Kannan and Joseph 2010). Precipitation and climate change fall in the category of atmospheric inputs. Weathering and erosion of crustal materials occur due to the interaction of water with soil and crustal rock. The anthropogenic activities by industrial and agricultural pollution, increasing consumption, and urbanization degrade the quality of groundwater and impair its safe uses for agricultural, industrial and domestic purposes (Kannan and Joseph 2010). The evaluation of groundwater quality could, therefore, provide a scientific basis for taking rational decisions and appropriate utilization and protection, which in turn would play a key role in environmental management, monitoring and decision making (Zou and Wang 2010).

During the past decades, various geochemical methods have been successfully used to assess the groundwater quality (Park et al. 2005; Subba Rao 2006; Song et al. 2007; Naik et al. 2009; Kannan and Joseph 2010; Mondal et al. 2011). However, it is now being realized that fusion of geochemical and geophysical information could provide somewhat better footprints of partial or complete mixing of contaminant with groundwater that control the groundwater chemistry in coastal regions. The appropriate quantitative analysis would alleviate some of the above problems (Prasanna et al. 2011). The conventional methods such as principal component analysis, discriminant function analysis, cluster analysis, etc. have generally been used in previous investigations for classification of water quality (Brown 1998; Panda et al. 2006; Long and Valder 2011). These techniques are linear and semi-automated and therefore not suitable to take into accounts the inherent nature of nonlinearity and complexity of the groundwater records.

Artificial Neural Network (ANN) has now become an alternative approach to resolve the above problem since the method has a natural ability to map between two domains with arbitrary precision even if there is a lack of deterministic or linear relationship. The conventional ANN methods have been widely used in almost all the branches of geophysics (e.g., Van der Baan and Jutten 2000; Poulton 2001; Sun et al. 2004; Zou and Wang 2007). There are, however, several cited limitations in conventional ANN approaches (Bishop 1995; Nabney 2004; Maiti et al. 2011; Tiwari and Maiti 2011). For instance, a complex model can fit training data well but it is not necessarily guaranteed that it will offer smaller errors in analyzing unseen data/new data (Bishop 1995; Maiti and Tiwari 2010). This is due to the fact that ensuing approaches do not take into account the uncertainty between the input and the output (Bishop 1995; Nabney 2004). So there is a pressing need to develop a quantitative method which can produce reliable results in the face of uncertainty.

The main objectives of the present study are to (1) develop a model based on ANN in Bayesian framework (BNN); (2) evaluate the applicability of the BNN to assess and classify the groundwater quality of western Maharashtra, India; and (3) to quantify uncertainty in the present model.

Materials and methods

Geology and hydrogeology of the study area

The study area lies between latitude 15.7°–16.2°N and longitude 73.4°–73.8°E in the Sindhudurg district, western Maharashtra, India (Fig. 1). Kudal, Malvan, and Vengurla are some of the important townships in the district. Geologically, the study area exposes rocks ranging in age from Archaean to Quaternary period. The Archaeans are represented by granite gneiss and is seen in the southern part of the Sindhudurg district near Vengurla and Sawantwadi. The palaeo- to Meso-Proterozoic represented by Dharwar supergroup overlie the Archaeans and occupy a major part of the area comprising psammatic metasediments consisting of metagabbro, quartz chromites, amphibolites schist, and ferruginous phyllite. The general geologic succession encountered in the Sindhudurg district is given in Table 1. The Archaean granite and gneisses are medium to coarse grained and consist of quartz, microcline, orthoclase biotite and hornblende (CGWB 2009; Maiti et al. 2012).

Dharwarian metasediments (Archaean), Kaladgi formation (Precambrian), Deccan Trap lava flows (Upper Cretaceous to lower Eocene), laterites (Pleistocene), and Alluvial deposits (Recent to sub-Recent) are the water-bearing formation observed in

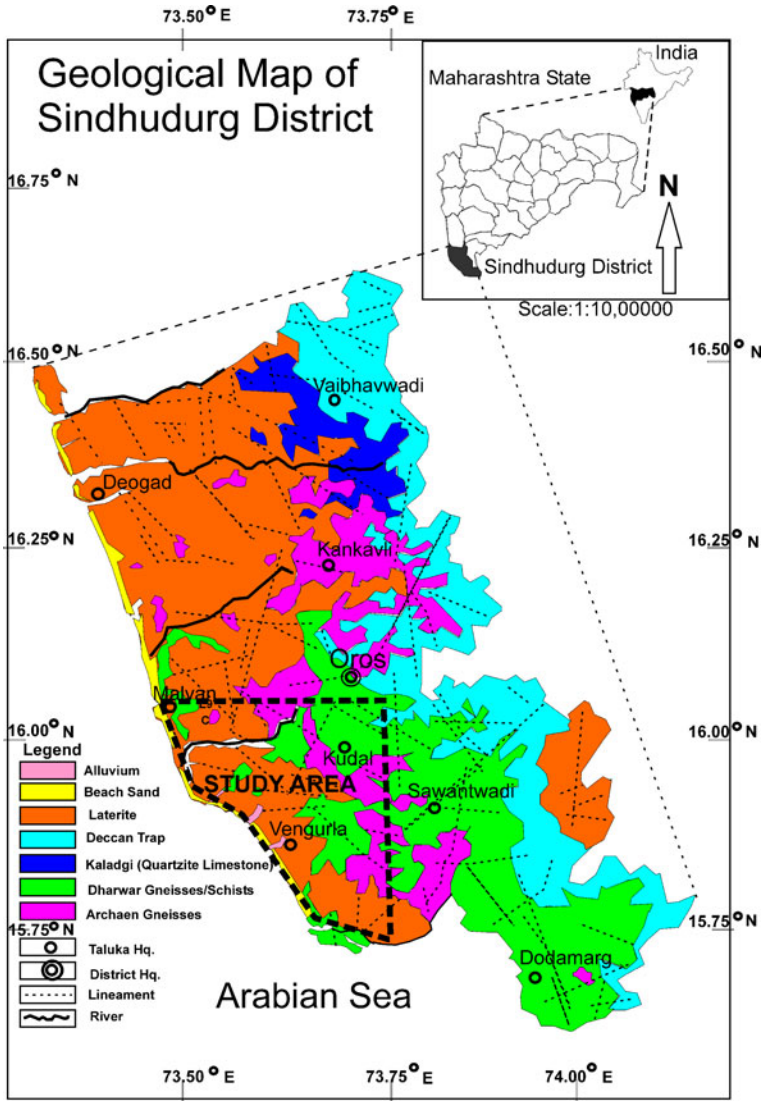


Fig. 1 Geological map of the study area

Sindhudurg district. However, the Kaladgi formation occurs only in very limited patches and does not form potential aquifer in the district. The alluviums also have limited areal extent found mainly along the coast (CGWB 2009; Maiti et al. 2012). Dharwarian gneiss/schists are devoid of primary porosity and permeability. Kaladgi rocks are mainly represented by orthoquartzite, limestone, sandstone, and shales. They are jointed in diverse directions and weathered portion actually controls their water-bearing properties (CGWB 2009; Maiti et al. 2012). Since primary porosity is negligent in the Deccan trap basalts, secondary porosity due to jointing and fracturing plays an important role in groundwater circulation. Laterite has more porosity than the Deccan trap basalt, thereby forming many potential aquifers in the area. Groundwater level in the study area varies from 2 m to 20 m (Maiti et al. 2012).

Table 1 General geological succession in Sindhudurg district, Western Maharashtra, India (source—CGWB report)

Geological time	Formation
Recent to sub-Recent	Alluvium beach sand
Pleistocene	Laterite and lateritic spread
Miocene	Shale with peat and pyrite nodules
Cretaceous to Eocene	Deccan trap basalt lava flows
Upper Pre-Cambrian	Kaladgi series, quartzite, sandstone shale, and associated limestone
Dharwar Super Group	Phyllite, conglomerate, quartzite

Fusion of geochemical and geophysical data and model development

Groundwater quality index in the present study is classified into three classes following the standard limit of World Health Organization (WHO 1984): (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) (Table 2). The model is derived by parameterizing the geochemical information such as (1) hydrogen ion concentration (pH), (2) electrical conductivity (EC), (3) total dissolved solids (TDS), (4) bicarbonate $\left(HCO_3^-\right)$, (5) chloride (Cl^-), (6) sulfate $\left(SO_4^{2-}\right)$, (7) nitrate $\left(NO_3^-\right)$, (8) calcium (Ca^{2+}), (9) magnesium (Mg^{2+}), (10) sodium (Na^+), (11) potassium (K^+), according to the standard of the World Health Organization (WHO 1984), and (12) true resistivity (ρ) obtained by Singular Value Decomposition (SVD)-based inversion of apparent resistivity data of VES Schlumberger survey. The statistics of the geochemical attributes/parameters are given in Table 3. Note that true resistivity (ρ) values from each of the sounding points were considered at a depth of 20 m for modeling since groundwater level varies from 2 m to 20 m in the study area. It is important to note that the classification of groundwater quality is being made based on how good the water is for safe drinking. For instance, pH of the water is expressed on a scale that runs from 0 to 14. Number 7 is neutral, meaning that there is a balance between acidic and basic (alkalinity). A pH value measured below 7 means that water is acidic while a pH value above 7 means that water is basic. Accordingly, a pH value in the range 7.5–8.5 is assigned for "very good", a pH value between 7.1 and 7.5 is assigned for "good", and a value between 0.01 and 7.0 and 8.51–14.00 is assigned for "unsuitable" for drinking. The pH value below 6.5 and above 8.5 is considered as aesthetic (WHO 1984). EC value below 1,500 μS/cm can be considered as safe for drinking. A greater value than 1,500 μS/cm of groundwater can be considered as saline (WHO 1984; Mondal et al. 2011). Groundwater with TDS value measured below 1,000 mg/l can be considered as fresh and recommended for safe drinking (WHO 1984). HCO_3^- value up to 300 mg/l in fresh groundwater is considered as safe for consumption. Cl^- is one of the major ions in saline water. The value below 200 mg/l can be suitable for drinking. The value of SO_4^{2-} below 200 mg/l can be suitable for drinking as per WHO recommendation (WHO 1984). Although the value of NO_3^- below 45 mg/l is permitted for drinking, the value less than 10 mg/l could be ideal for drinking (WHO 1984). Ca^{2+} and Mg^{2+} are considered as one of the major ions in fresh groundwater. The limits of Ca^{2+} and Mg^{2+} for safe drinking are up to 75 mg/l and 30 mg/l respectively, according to WHO guidelines (WHO 1984). Na^+ is considered as one of the major ions in saline water. Na^+ quantity in groundwater below 200 mg/l is fit for safe drinking (WHO 1984; Mondal et al. 2011). Groundwater can be suitable for drinking if the quantity of K^+ is below 100 mg/l (WHO 1984). The resistivity of geologic layer (ρ) varies on wide scale. The value of ρ relies

Table 2 The geochemical and geophysical parameters used for forward modeling

Geochemical/ geophysical attributes	Very good (1.00)	Good (0.50)	Unsuitable (0.00)	Permissible limit (WHO 1984)
pH ($-\log 10\ H^+$)	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	751–1,500	1,501–3,000	750
TDS (mg/l)	0.01–500	501–1,500	1,501–3,000	1,000
$HCO_3^-(mg/1)$	0.01–200	201–300	301–500	300
Cl^- (mg/l)	0.01–100	101–200	201–1,000	200
$SO_4^{2-}(mg/1)$	0.01–100	101–200	201–500	200
$NO_3^-(mg/1)$	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 (Ω-m)	11–1,000	1,001–40,000	0.01–10	–

Table 3 The statistics of geochemical and geophysical parameters in the study area

Geochemical/ geophysical attributes	Minimum	Maximum	Mean	STD	Skew	Kurtosis
pH ($-\log 10\ H^+$)	5.81	6.98	6.98	0.48	−0.65	0.20
EC (μS/cm)	70	4,450	421.47	725.5	5.19	29.03
TDS (mg/l)	45.5	2,845	271.03	463.33	5.20	29.1
HCO_3^- (mg/1)	24.4	280.6	89.27	60.40	1.53	2.02
Cl^- (mg/l)	13.19	832	54.50	139.26	5.34	29.82
SO_4^{2-} (mg/1)	76.84	298.2	159.08	58.72	0.57	−0.20
NO_3^- (mg/1)	0.08	10.19	1.69	2.03	2.55	8.11
PO_4^- (mg/1)	0.00	1.84	0.35	0.48	1.82	2.42
Ca^{2+} (mg/l)	28.05	180.2	57.28	32.01	2.07	5.13
Mg^{2+} (mg/l)	2.43	112.3	22.90	19.67	3.1	12.23
Na^+ (mg/l)	3.78	276.1	33.90	48.48	3.99	18.58
K^+ (mg/l)	0.39	8.6	2.92	2.2	1.19	0.83
Th (mg/l)	120	912	236.38	148.46	2.97	11.73
SAR	0.15	3.98	0.85	0.81	2.16	5.91
ρ (Ω-m)	1	38,217	3,362	9,485	3.18	9.08

on many factors including chemical composition of the water saturating the geologic layer and the composition of the geologic layer and their internal packing pattern. It is known that the resistivity of geologic layer can be below 10Ω-m if the layer is saturated with saline water (resistivity of the saline water is 0.2 Ω-m). Accordingly, we parameterized the whole range of geochemical and geophysical attributes for groundwater quality assessment study. The range of each parameter and corresponding groundwater quality index is given in Table 2. The relationship between groundwater quality (GWQ) and input variables can be expressed as

$$\mathrm{GWQ} = f(\mathrm{pH}, \mathrm{EC}, \mathrm{TDS}, \mathrm{HCO_3}, \mathrm{Cl}, \mathrm{SO_4}, \mathrm{NO_3}, \mathrm{Ca}, \mathrm{Mg}, \mathrm{Na}, K, \rho) \quad (1)$$

where pH, EC, TDS, HCO_3, Cl, SO_4, NO_3, Ca, Mg, Na, K, and ρ denote respectively hydrogen ion concentration, electrical conductivity, total dissolved solids, bicarbonate, chloride sulfate, nitrate, calcium, magnesium, sodium, potassium of water samples, and true resistivity of the geologic layer.

Sources of geochemical information

Geochemical information were obtained by analyzing 36 water samples collected during March 2011 from representative dug wells and bore wells distributed throughout the area. The location of the geochemical sample collection points has been marked in Fig. 2. Water samples were collected at 0.5 m below the groundwater table from each well. All the wells are functional, and duration of each pumping lasted at least 5 min. Samples were collected in a 1.0-l-capacity plastic bottle. The bottles were first washed with diluted HNO_3 and then with distilled water thoroughly in order to make it ready for sample collection. EC and pH were measured immediately, while all other parameters were determined in the laboratory. EC was measured in microsiemens per centimeter at 25 °C. Except pH and EC, all other parameters are expressed in milligrams per liter (mg/l). We note that methods of water sample collection and analysis were essentially followed by Browen et al. (1974) and APHA (1985).

During sampling, greater emphasis was given on the areas where pollution was suspected. Most of the bore well samples were collected at locations where the local people complained of brackish taste of water due to proximity of the area with Arabian Sea. Attempt was made to collect at least one water sample from every village in the study area. It is important to note that geophysical data could be complimented to geochemical information where no water samples would be collected due to the unavailability of well.

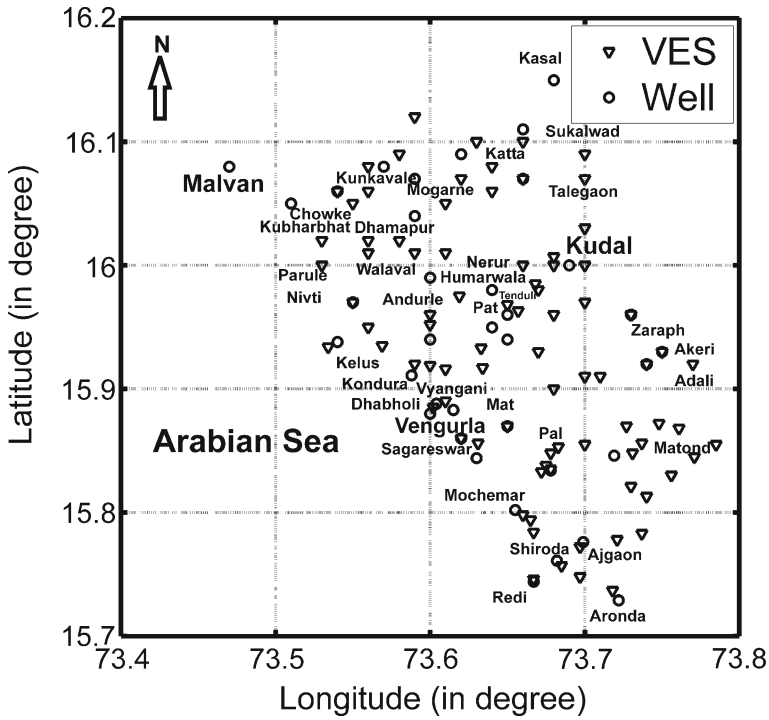


Fig. 2 Location map of the geochemical sampling point (well) and the geophysical vertical electrical sounding (VES) point

Sources of geophysical information

Geophysical information was obtained by inverting 85 of the VES data from the study area (latitude 15.7°–16.2°N and longitude 73.4°–73.8°E). The VES site was chosen according to the suitability of terrain condition to cover the area uniformly as much as possible. We note that the majority of the VES locations happen to be very close to the geochemical sampling location. A maximum half of current electrodes spacing (AB/2) of 100 m was used with half of potential electrodes spacing (MN/2) of 0.5 and 15 m in successive steps (Fig. 2). To improve the quality of data, the resistivity meter performed a minimum stack of three for each data point. The location of the sounding point has been marked in Fig. 2. We have inverted all 85 VES data points using very popular damped least-squares algorithm (Maiti et al. 2012). The damped least-squares-based inversion solution is used following (Menke 1984; Maiti et al. 2012)

$$\Delta m = \left(G^T G + \beta^2 I\right)^{-1} G^T \Delta d \tag{2}$$

where Δm is the parameter correction vector, Δd is the data difference vector, and G is the Jacobian matrix that contains partial derivatives of data with respect to the initial model parameters. I is the identity matrix, and the term β is called damping factor which is a scalar quantity that actually controls both speed of convergence and solution. The inversion solution (i.e., true resistivity (ρ)) and assumed homogeneous isotropic layer thickness (h) from recorded apparent resistivity data were estimated by SVD following (Maiti et al. 2011; Maiti et al. 2012). The program was developed and was implemented at the different stages of processing using commercial software package Matlab (Matlab 2010).

Bayesian neural network modeling

In the present study, synthetic model was optimized in Bayesian neural network (BNN) framework by fusing geophysical and geochemical information. Thus, the BNN model was developed to map between an input space (here geochemical and geophysical attributes) and an output space (groundwater quality, GWQ). The model was trained on a finite data set (input/target pairs) say $s = (\{d_k, x_k\})^{N_{k=1}}$ by minimizing misfit error

$$E_s = \frac{1}{2}\sum_{k}^{N}\{x_k - o_k(d_k; w_k)\}^2 \tag{3}$$

to adjust network parameters (weight and bias). The repeated evaluation of the gradient of error E_s was carried out using the "back-propagation" scheme (Bishop 1995). The forward function used in the study is non-linear function (tan sigmoid) which actually helps to solve the weakly non-trivial problem (Fig. 3a).

Bayesian philosophy strongly believes in practicing prior state of information (before inversion). In the Bayesian approach, a suitable prior probability distribution, say $P(w)$ of weights, is considered before observing the data instead of just a single set of weights as usually used in conventional ANN approach. Using Bayes' rule, a posteriori probability distribution for the weights, say $P(w|s)$, can be given as (Bishop 1995)

$$P(w|s) = \frac{P(s|w)P(w)}{P(s)} \tag{4}$$

where $P(w|s)$ is the data set likelihood function and the denominator, $P(s)$, is a normalization factor/scaling factor. The denominator $P(s)$ is intractable so direct estimation of posterior $P(w|s)$ is not possible. It is interesting to note from the above equation that maximization of $P(w|s)$ leads to the maximization of posterior $P(w|s)$ and the scaling factor is of no interest to us; therefore, the proportionality in Eq. (4). Using the rules of conditional probability, the distribution of outputs for a given input vector x can be written in the form (Bishop 1995)

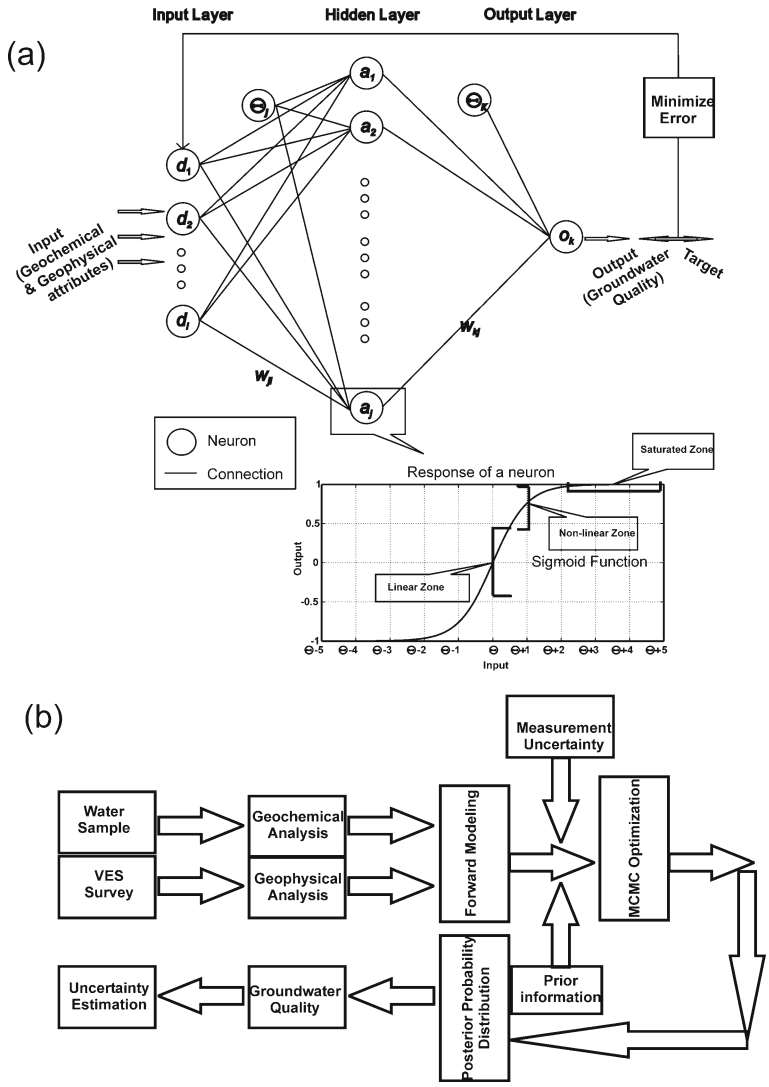


Fig. 3 **a** Layout of MLP with a three-layer neural network with "d" representing input, subscript "I" representing the number of nodes in the input layer, w_{ji} representing the connection weight between the ith node in the input layer and the jth node in the hidden layer, and w_{jk} representing the connection weight between the jth node in the hidden layer and the kth node in the output layer. Θ_j and Θ_k are bias vectors for the hidden and the output layer. When the sum of the argument of a neuron is comparable to the threshold value Θ_j, the sigmoid function squashes linearly; otherwise, it saturates with value +1; −1 gives non-linearity for non-linear mapping between an input and an output space. Sigmoid function plays a key role for mapping non-trivial groundwater quality problem. **b** Flow diagram of the methodology for groundwater quality assessment

$$P(x|d,s) = \int P(x|d,w)P(w|s)dw \quad (5)$$

The major problem in Bayesian computation is evaluating integrals for posterior *pdf* of weights (Eq. 4) and network output (Eq. 5). Monte Carlo/Markov Chain Monte Carlo (MCMC) sampling-based numerical method was used for evaluating posterior integrals. Using MCMC, Eq. (5) can be approximated as

$$P(x|d,s) = \frac{1}{N}\sum_{n=1}^{N} P(x|d,w_m) \quad (6)$$

where $\{w_m\}$ represents a MCMC sample of weight vectors obtained from the distribution $P(w|s)$ and N is the number of points w sampled from $P(w|s)$. The Hybrid Monte Carlo (HMC) algorithm (Duane et al. 1987; Maiti et al. 2011) is a global optimization scheme where each trajectory was updated by integrating the resulting Hamiltonian differential equations via a second-order (leapfrog) discretization scheme. Detailed description of the HMC/MCMC method can be found elsewhere (Bishop 1995; Tiwari and Maiti 2009).

General flow chart

Contrary to the joint inversion where each information (geochemical and geophysical) was integrated by individual inversion to interpret the results, we fuse the geochemical and geophysical information and then invert it to derive a model for final interpretation. The general flow chart of Bayesian neural network-based algorithm for groundwater quality assessment is described as follows (Fig. 3b).

Step 1: The geochemical and geophysical forward modeled data are preprocessed to scale the input/output series in between [−1 +1] by using a simple linear transformation algorithm; normalized input=2×(input−minimum input)/(maximum input−minimum input)−1 (Maiti et al. 2012), so that network can perform non-linear mapping within the bound.

Step 2: We assume 2 % observational error/measurement uncertainty due to the sampling of training data. We also assumed that measurement uncertainty to be Gaussian so that solution can be analytically tractable. While modeling a suitable prior state of information (e.g., location of well, topography, proximity to sea, or pollutants, etc.) is incorporated such that the value of resistivity of geological layer and geochemical attributes should not be negative. The impact of the exposure of salt pan very near to the sampling well would be more in general than that of the remote well. Likewise, if the topography undulation is rapid and sharp, then the residence time of the water with rock would be less, thereby decreasing the concentrations of dissolved solids.

Step 3: Expression of posterior probability distribution is estimated using Bayes' theorem and the network is optimized via Markov Chain Monte Carlo (HMMC) algorithm after setting up step size, iteration, simulation directions, etc.

Step 4: The new data is post-processed and applied to the trained network, and groundwater quality maps are generated (from network output) along with the associated uncertainty maps (from the posterior covariance matrix). The posterior distribution may be approximated to Gaussian distribution so that the desired calculation (statistical properties) can be done easily.

The mean $o(d; w_m)$ and the variance of the above distribution can be given by (Bishop 1995)

$$\delta_i^2 = \frac{1}{\mu} + g^T H^{-1} g \quad (7)$$

Here, physically, μ represents the inverse variance of the noise model, g denotes the gradient of network output $o(d; w_m)$ with respect to the weights w evaluated at w_m, and H is the Hessian matrix of the total error function (regularized) which we can simplify following Bishop (1995) and Maiti and Tiwari (2010). δ_i is the standard deviation (STD) of the predictive distribution for the target model.

Model initiation

The training samples were generated by suitable parameterization of the geochemical and geophysical variables according to the explicit limit of the standard prescribed by WHO (1984) given in Table 2. The networks

structure used in the study is 15–5–1, i.e., 15 input nodes, five hidden nodes, and one output node. Input nodes take the geochemical and geophysical attributes. Since three-layer models can produce sufficient accuracy for non-linear mapping, we used one hidden/intermediate layer. Five hidden nodes were found sufficient for optimization. One node at the output layer denotes the GWQ index. Although in Bayesian approach it is not required to keep some data separately for validation and testing the model because of its mathematical soundness to overcome the over-fitting, we do cross-validation (CV) technique for keeping uniformity with earlier development. The pH could take at least three (usually, minimum, mean, and maximum) different values from each category of "very good", "good", and "unsuitable". Likewise EC, TDS, HCO_3^-, Cl^-, SO_4^{2-}, NO_3^-, Ca^{2+}, Mg^{2+}, Na^+, K^+, and ρ could take at least three different values from each category of "very good", "good", and "unsuitable". In total, 1,500 representative samples (input and corresponding output) were generated by the data range given in Table 2. Out of 1,500 samples, 50 % (i.e., 750) were used for training, 25 % (i.e., 375) were used for validation, and 25 % (i.e., 375) were used for testing the model. We used a single prior hyperparameter for all the weights in a network that controls the extent to which the regularization term influences the form of the solution. In order to define an objective function in the Bayesian framework, an error model for the data likelihood is required (Maiti and Tiwari 2010). Having assumed that the target data is formed from a smooth function with additive zero mean Gaussian noise, we estimated the corresponding hyperparameters for both hidden and output layer weights that is essentially controlled by the noise present in the target data.

Model verification

The performance of BNN model were evaluated according to the statistical criteria such as Pearson correlation coefficient (r), coefficient of determination (r^2), root mean square error (RMSE), reduction of error (RE), and index of agreement (IA) (Willmott 1981).

$$r = \frac{\sum_i^N \left(O_i - \overline{O_i}\right)\left(P_i - \overline{P_i}\right)}{\sqrt{\sum_i^N \left(O_i - \overline{O_i}\right)^2 \left(P_i - \overline{P_i}\right)^2}} \tag{8}$$

$$r^2 = \frac{\left(\sum_i^N \left(O_i - \overline{O_i}\right)\left(P_i - \overline{P_i}\right)\right)^2}{\sum_i^N \left(O_i - \overline{O_i}\right)^2 \left(P_i - \overline{P_i}\right)^2} \tag{9}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^N (O_i - P_i)^2}{N}} \tag{10}$$

$$\mathrm{RE} = 1.0 - \frac{\sum_i^N (O_i - P_i)^2}{\sum_i^N O_i^2} \tag{11}$$

$$\mathrm{IA} = 1.0 - \frac{\sum_i^N (O_i - P_i)^2}{\sum_i^N \left[\left|P_i - \overline{O_i}\right| + \left|O_i - \overline{O_i}\right|\right]^2} \tag{12}$$

where O_i is the observed value, P_i is the predicted value, $\overline{O_i}$ is the average of the estimated value, $\overline{P_i}$ is the average of the predicted value, and N is the sample number.

Uncertainty analysis

The uncertainty of mapping between input and target can be due to the following two reasons: (1) intrinsic noise (correlated, uncorrelated, and observational error) embodied in hyperparameter μ (Maiti and Tiwari 2010) and (2) theoretical error (modeling uncertainty due to inaccurate theory and imperfections in the data learning) described by the posterior distribution of the weight vector w (Bishop 1995). The main diagonal elements of the output covariance matrix represents the "variance" with respect to the "mean", and off-diagonal elements of the output covariance matrix represents how the "variance" are correlated (Tarantola 1987; Maiti and Tiwari 2010). The standard deviation (STD) is calculated by taking the square root of the diagonal elements of output covariance matrix. The STD of the predictive distribution of the network output can be interpreted as an error bar on the mean value $o(d; w_m)$. Following

Nabney (2004), the 90 % confidence interval of the mean network $o(d; w_m)$ was calculated by adding and subtracting STD from $o(d; w_m)$.

Results and discussions

Spatial distribution of geochemical and geophysical parameters

Quality of groundwater determines the usefulness of water for the purposes of drinking, irrigation, and industrial uses. The chemical quality of groundwater is influenced to a considerable extent by the chemical composition of rocks and soil mass through which it moves under various physico-chemical conditions. As the water percolates downwards through the zone of aeration, not many changes in the composition of groundwater are expected as it moves faster in this zone (Naik et al. 2009). Water movement is usually very slow in the underground reservoirs and, therefore, it gets sufficient time for reaction with the rocks to take place (Fetter 1988). Groundwater also undergoes frequent changes in chemical composition by secondary phenomena such as base exchange, oxidation, reduction, etc., which may modify the composition of water (Schoeller 1959). Schoeller (1959, 1967), Hem (1970), and many others have discussed the nature of groundwater in different rock types. Hydrological changes during season also play an important role in variation of major-ion concentration of groundwater (Rajmohan and Elango 2006). Some excellent work were reported that the rising water table in post-monsoon period dissolves more saline matter from the soils and increases the salinity of water after the monsoon (Rajmohan and Elango 2006). Although apart from these process, evaporation and irrigation return flow influence the major-ion chemistry in groundwater (Hem 1970). Statistical parameters, including minimum, maximum, mean, and standard deviation, of different chemical constituents of the analyzed groundwater samples (N=36) from the study area were computed and shown in Table 3. Most parameters had wide ranges and high standard deviations, as shown in Table 3. In particular, the EC value had a wide range between 70 and 4,450 μS/cm (mean=421.47 μS/cm). The ranges of Na^+ and Cl^- ions were from 3.78 to 276.1 mg/l (mean=33.90 mg/l), and 13.19 to 832 mg/l (mean=54.50 mg/l), respectively. The concentration of $SO4^{2-}$ also widely varied from 76.84 to 298.2 mg/l (mean=159.08 mg/l). Such wide ranges of solute concentrations suggest that multiple sources and/or complex hydrochemical processes acted in agreement to generate the chemical composition of groundwater. In order to access the water quality, it is essential to understand the chemical properties of the water samples.

Hydrogen ion concentration (pH)

The pH value in the study area ranges from 5.81 to 6.98. The pH concentration map indicates that the groundwater is relatively acidic in the places of Kondura (pH value 5.81), Mochemar (pH value 5.92), Kelus (pH value 5.98), Humarmala (pH value 6.34), Nerur (pH value 6.36), and Kankavali (pH value 6.49) than the rest of the places (Appendix Fig. 9). A positive correlation is reported between pH of the groundwater and the fluoride content, meaning that alkalinity can influence the fluoride content in groundwater (Duraiswami and Patankar 2011). Low pH in the groundwater could be due to the presence of the laterite which is porous and permeable (Raghunath et al. 2001). The excessive use of fertilizers like ammonium sulfate and phosphate in agriculture also causes low pH in groundwater (Appelo and Postma 2005). In view of health hazard, a low pH (pH<6.5) actually can cause gastrointestinal disorders: hyperacidity, ulcers, stomach pain, and burning sensation. The acidic nature of groundwater in the study area can also accelerate the corrosion rate of metallic substances in water which in turn causes the rise of metallic substance in groundwater. The low pH value in the northern part of the study area is due to the interaction of water with rock like quartz, chlorite, amphibole, schist, and ferruginous phyllite.

Electrical conductivity (EC)

The EC is an important constituent in groundwater. The EC value varies from 70.00 μS/cm to 4,450.00 μS/cm. The EC recorded at Kelus (EC value 4,450.00 μS/cm) and Shiroda (EC value 1,240.00 μS/cm) falls beyond the acceptable level for drinking prescribed by WHO (1984) (Appendix Fig. 9). The high EC value at Shiroda and Kelus is due to the intrusion of saline aerosol from Arabian Sea. The SE and NW parts of the study area have less EC concentration, meaning that wells in this

part of the study area are not affected by seawater intrusion.

Total dissolved solids (TDS)

The TDS is regarded as one of the main factors which determine the use of groundwater for any purpose. TDS value found at Kelus (TDS value 2,845.00 mg/l) and Shiroda (TDS value 792.00 mg/l) exceeds the acceptable limit prescribed by WHO (1984) (Appendix Fig. 9). Nutrient enrichment due to fertilizers and saline water intrusion could enhance TDS and, in turn, increases the EC in the study area.

Bicarbonate (HCO_3^-)

The spatial distribution of the bicarbonate ions recorded at places Vyangani (HCO_3^- value 280.60 mg/l), Shiroda (HCO_3 value 219.70 mg/l), Nivti (HCO_3 value 207.4 mg/l), Matond (HCO_3^- value 195.3 mg/l), and Sagareshwar (HCO_3^- value 146.4 mg/l) is slightly higher (Table 2, Appendix Fig. 9). NW and the central part of the study area indicate low bicarbonate concentration value. However, NE and SE parts of the study area indicate relatively more bicarbonate concentration. The primary source of bicarbonate ions in groundwater is attributed to the dissolution of carbonate minerals (e.g., calcite). The secondary source of bicarbonate is due to the reaction of water with the carbon dioxide gas. The chemical reaction of groundwater with augite (by dissolution) also gives bicarbonate ions (Naik et al. 2009). It is important to note that H^+ is a small ion and can easily enter the crystal structure, thereby releasing other ions into the groundwater.

Chloride (Cl^-)

Cl^- occurs naturally in all types of water. Cl^- is the dominant ion of seawater (Song et al. 2007). Cl^- concentration recorded at Kelus (Cl^- value 832 mg/l) and Shiroda (Cl^- value 255.2 mg/l) falls beyond the permissible limit (200 mg/l) (Appendix Fig. 10). The high concentration of Cl^- at these places is primarily due to the saline water intrusion and secondarily due to the influence of discharged agricultural, industrial, and domestic waste waters. NW and SE parts of the areas show less concentration of chloride while NE part shows high chloride concentration (Appendix Fig. 10).

Sulfate (SO_4^{2-})

The SO_4^{2-} values recorded were highest at Malvan (SO_4^{2-} value 298.2 mg/l) followed by Matond (SO_4^{2-} value 293.4 mg/l) (Appendix Fig. 10). Rainwater and anthropogenic activities can be important sources for other constituents such as sulfate (Pawer et al. 1998). The primary mineral sources of sulfate ions include evaporate minerals (calcium and gypsum) and sulfates of magnesium and sodium (Pradhan and Pirasteh 2011). NE part of the study area, including Matond, shows higher sulfate concentration than that of Shiroda and Kelus.

Nitrate (NO_3^-)

The (NO_3^-) concentration recorded at all the places in the study area falls within the permissible limit (45 mg/l) (Appendix Fig. 10). (NO_3^-) concentration is found to be comparatively very low because the present samples were collected in summer. Since (NO_3^-) is loosely bound to soils, it is expected to be more driven by the runoff in the rainy season. This may cause its rise of concentration in groundwater particularly in the rainy season (Rao et al. 2004). NE and SE parts show higher nitrate concentration value while Matond shows the highest nitrate concentration. Note that in the context of health risk, greater concentration of (NO_3^-) may have a toxic effect on young infants.

Calcium (Ca^{2+})

Calcium is a major ion in fresh water (Song et al. 2007). The spatial distribution of Ca^{2+} in the study area shows that the recorded values at places Kelus (Ca^{2+} value180.20 mg/l), Matond (Ca^{2+} value 116.20 mg/l), Malvan (Ca^{2+} value 104.20 mg/l), Vyangani (Ca^{2+} value 100.20 mg/l), Mochemar (Ca^{2+} value 96.19 mg/l), Shiroda (Ca^{2+} value 104.2 mg/l), Aronda (Ca^{2+} value 80.4 mg/l), and Nivti (Ca^{2+} value 80.16 mg/l) falls beyond the permissible limit(75 mg/l) (Table 2, Appendix Fig. 10). The crystalline limestone and prolonged agricultural activities could influence directly or indirectly to augment the mineral dissolution in groundwater (Bohlke 2002). Ca^{2+} concentration is less in the N and NE parts of the study area while Kelus records the highest calcium concentration.

Magnesium (Mg^{2+})

The content of Mg^{2+} is comparatively less than that of Ca^{2+}. The spatial distribution of Mg^{+2} concentration shows highest value recorded at Kelus (Mg^{2+} value 112.3 mg/l) which falls beyond the permissible limit (Table 2, Appendix Fig. 11). Its solubility in water is around five times that of calcium (Pradhan and Pirasteh 2011). The concentration is high in the NE part of the study area.

Sodium (Na^{+})

Na^{+} is one of the important naturally occurring cations, and its concentration in fresh water is generally lower than that of Ca^{+} and Mg^{+}. The Na^{+} concentration recorded at Kelus (Na^{+} value 276.10 mg/l) falls beyond the permissible limit (50 mg/l) whereas values of Shiroda (Na^{+} value 136.8 mg/l) and Aronda (Na^{+} value 62.21 mg/l) are significant (Appendix Fig. 11). Sodium ions are leached in the area primarily due to the reaction of feldspars with the groundwater (hydrolysis). The higher value is due to the interaction of groundwater with rock type of plagioclase feldspar and granite gneiss. In coastal area, sodium could be added to the groundwater through saline aerosols from Arabian Sea whereas in lateritic terrain groundwater is contaminated due to leaching during the process of lateralization. The rise of Na^{+} level in the NE part of the area may be attributed to the use of Na^{+}-rich fertilizer. Health hazard may occur due to a higher sodium intake which usually causes hypertension, congenital heart diseases, and kidney problems.

Potassium (K^{+})

The K^{+} concentration recorded at all the places in the study area falls within the permissible limit (10 mg/l) (Appendix Fig. 11). This involves the dissolving of feldspar minerals in the granite by hydrogen. The feldspar reacts with hydrogen in water producing kaolin (china clay) in the processes of kaolinization—this occurs as water circulates through the granite. Granite weathered by hydrolysis becomes weakened as the quartz crystals remain as loose crystals, unaffected by the hydrolysis process. The concentration of K^{+} shows very low values in all the places. Particularly, the low concentration of K^{+} is owing to the fact that a part of the K^{+} goes into the clay structure. Note that clay is a very common lithology observed in the coastal part of the study area. Although Na^{+} and K^{+} are released in equal amounts during weathering, K^{+} concentration in groundwater is found comparatively less as compared to sodium (nearly one tenth or one-hundredth that of sodium) because the potassium minerals are resistant to decomposition by weathering. The K^{+} concentration is found high at the NE part of the study area.

True resistivity (ρ)

True resistivity and layer thickness of each geologic layer is deduced by the inversion of VES apparent resistivity data. Inversion results suggest that there are three to four geologic layers in the study area (Maiti et al. 2012). Note that true resistivity contour map reflects the distribution pattern of true resistivity at a depth of 20 m. The contour map of ρ at a depth of 20.00 m suggests that the true resistivity found at Kelus is the minimum ($\rho \sim 1\,\Omega$-m). The true resistivity found at Shiroda and Malvan is of the order of 0.95 Ω-m and 2 Ω-m, respectively (Appendix Fig. 11). The low resistivity of these places is attributed to the saline water intrusion from the Arabian Sea. The coastal parts are conductive possibly due to the effect of saline water intrusion whereas the NE and SE parts of the study area are highly resistive. The relatively low resistivity at Nerur may be attributed to leaching of normal salt. Note that the mismatches between EC and ρ maps are likely because the EC map is based on the direct groundwater sample analysis whereas the true resistivity map is derived from gross resistivity values of geologic layer saturated with or without water.

Gibb's diagram

Gibb's diagrams, the ratio of ions $Na^{+}/(Na^{+}+Ca^{2+})$ and $Cl^{-}/(Cl^{-}+HCO_3^{-})$ as a function of TDS are widely used to assess the functional relationship of sources of dissolved chemical constituents (Gibbs 1970). The ratio of ions $Na^{+}/(Na^{+}+Ca^{2+})$ and $Cl^{-}/(Cl^{-}+HCO_3^{-})$ of groundwater samples are plotted against the respective values of TDS. The plot indicates that more than 95 % of the groundwater samples fall in the rock dominant category and the rest into the evaporation domain (Fig. 4). Rock dominance

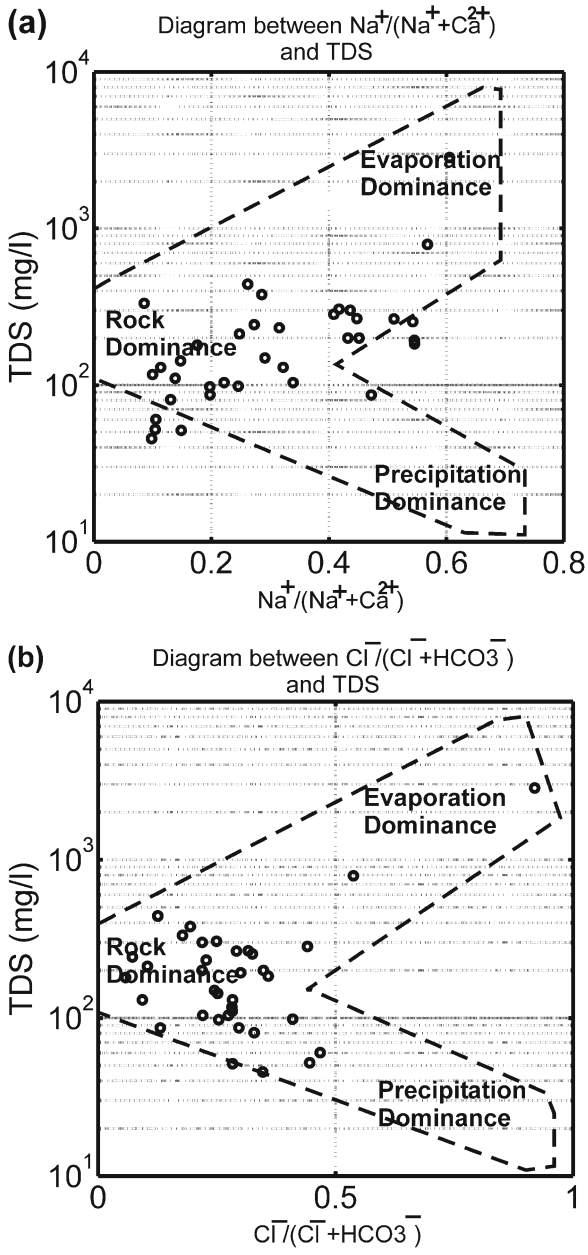


Fig. 4 **a**–**b** Gibb's diagram of collected geochemical samples from the study area. The plots are semi-log type

of the samples is caused by the interaction between the aquifer rocks and groundwater, and can be regarded as the good quality water/fresh water.

Cross-correlation analysis

Cross-correlation analysis among chemical attributes is presented in Table 4. It had shown that the highest correlation had been found between EC and TDS of the order of 0.99. The correlation had been found between Cl^- and TDS of the order of 0.97 and between Na^+ and TDS of the order of 0.94 (Table 4, Appendix Fig. 12). The coefficient of determination (r^2) between Na^+ and TDS is 0.88, which means that 88 % of the total variation in Na^+ can be explained by the linear relationship between Na^+ and TDS (Appendix Fig. 12). The correlation coefficients for other constituents, such as Mg^{2+}, Ca^{2+}, K^+, and SO_4^{-2}, with TDS declined progressively with values 0.82, 0.75, 0.43, and 0.24, respectively. This indicates proportionately diminishing contributions of these constituents to groundwater pollution. The coefficient of determination (r^2) between Mg^+ and TDS is 0.67, which means that 67 % of the total variation in Mg^+ can be explained by the linear relationship between Mg^+ and TDS (Appendix Fig. 12). Likewise, the coefficient of determination (r^2) between Cl^- and TDS is 0.94, which means that 94 % of the total variation in Cl^- can be explained by the linear relationship between Cl^- and TDS (Appendix Fig. 12). The coefficient of determination (r^2) between SO_4^{-2} and TDS is 0.05, which means that 5 % of the total variation in SO_4^{-2} can be explained by the linear relationship between SO_4^{-2} and TDS (Appendix Fig. 12). Table 4 shows the highest correlation between Na^+, Mg^{2+}, and Cl^- with TDS, which were estimated to be 0.94, 0.82, and 0.97, respectively (Table 4, Appendix Fig. 12). The correlation coefficient between TDS and SO_4^{-2} was found to be 0.24 (Table 4, Appendix Fig. 12). The correlation coefficients between Mg^{2+}, Na^+, and Ca^{2+} with Cl^- were found to be 0.91, 0.84, and 0.70, respectively, whereas for the Cl^- and K^+ it was 0.39 (Table 4, Appendix Fig. 12). This indicates that there is a strong positive correlation between Mg^{2+} and Cl^-, Na^+ and Cl^-, and relatively less positive correlation between Ca^{2+} and Cl^-, but much less positive correlation between Cl^- and K^+. The coefficient of determination (r^2) between Mg^+ and Cl^- is 0.65, which means that 65 % of the total variation in Mg^+ can be explained by the linear relationship between Mg^+ and Cl^- (Appendix Fig. 12). The coefficient of determination (r^2) between Na^+ and Cl^- is 0.88, which means that 88 % of the total variation in Na^+ can be explained by the linear relationship between Na^+ and Cl^- (Appendix Fig. 12). The coefficient of determination (r^2) between Ca^{2+} and Cl^- is 0.49, which means that 49 % of the total variation in Ca^{2+} can be explained by the linear relationship between Ca^{2+} and

Table 4 Correlation matrix among different chemical constituents of groundwater samples collected from the western India

	pH	EC	TDS	Ca^{2+}	Mg^{2+}	Na^{+}	K^{+}	HCO_3^-	Cl^-	SO_4^{2-}	NO_3^-
pH	1.00	–	–	–	–	–	–	–	–	–	–
EC	−0.37	1.00	–	–	–	–	–	–	–	–	–
TDS	−0.37	0.99	1.00	–	–	–	–	–	–	–	–
Ca^{2+}	−0.26	0.75	0.75	1.00	–	–	–	–	–	–	–
Mg^{2+}	−0.19	0.82	0.82	0.69	1.00	–	–	–	–	–	–
Na^{+}	−0.30	0.94	0.94	0.74	0.80	1.00	–	–	–	–	–
K^{+}	−0.30	0.43	0.43	0.50	0.48	0.49	1.00	–	–	–	–
HCO_3^-	0.05	0.13	0.12	0.59	0.28	0.23	0.33	1.00	–	–	–
Cl^-	−0.29	0.97	0.97	0.70	0.81	0.94	0.39	0.06	1.00	–	–
SO_4^{2-}	−0.22	0.24	0.24	0.57	0.49	0.33	0.53	0.50	0.16	1.00	–
NO_3^-	0.16	−0.05	−0.05	0.24	0.22	0.00	0.06	0.34	−0.05	0.56	1.00

Cl^- (Appendix Fig. 12). Likewise, the coefficient of determination (r^2) between K^+ and Cl^- is 0.15, which means that 15 % of the total variation in K^+ can be explained by the linear relationship between K^+ and Cl^- (Appendix Fig. 12).

The Ca^{2+}/Na^+ and HCO_3^-/Cl^- molar ratios offer an indicator for demarcating groundwater types. The ratio of chloride to carbonate+bicarbonate is called Simpson's ratio (Todd 1959). In order to evaluate the groundwater quality qualitatively, the TDS verses inter-ionic ratios (Ca^{2+}/Na^+ and HCO_3^-/Cl^- diagrams were examined. The TDS vs. HCO_3^-/Cl^- exhibits a strong negative correlation. For "very good" groundwater (TDS ranges, 0–500 mg/l), molar ratios of Ca^{2+}/Na^+ and HCO_3^-/Cl^- varied from 0.79 to 11.00 and 1.5 to 16.00, respectively; whereas for "good" groundwater, 0.89 to 1.42 (TDS ranges, 501–1,500 mg/l). For unsuitable groundwater, the range is 0.85 to 0.01 (TDS ranges, 1,501–3,000 mg/l) (Fig. 5). The high TDS is closely associated with high Cl^- in the groundwater which is mainly attributed to saline water intrusion, here in this case from the Arabian Sea. The bivariate plots of Cl^- and NO_3^- concentration also uncover three groups of groundwater quality. This plot also indicates that majority of the groundwater quality comes under "very good" that are not affected by seawater. The rest of the two groups is under the influence of seawater (Fig. 6). The present results show that groundwater was extensively damaged in the coastal area due to the saline aerosols from the Arabian Sea.

Bayesian neural network modeling criterion and results

The BNN model was applied to classify groundwater quality index using the real geochemical and geophysical data. The following criteria have been chosen for assessing the groundwater quality. The output of the BNN modeling produces groundwater quality index in a range 0.00–1.00. Ideally, if the output of the model produces 1.0, then it may be considered as "very good" water. Similarly, if the model outputs display 0.5, then it may be considered that the groundwater quality is "good". However, for the output value 0.0 or nearly so, the quality of the groundwater will be considered as "unsuitable".

The performance of the BNN model was verified on two independent data sets. The entire data sets were shuffled uniformly and checked for the population density of each pattern. In order to check the robustness of the results, we computed various statistical parameters, e.g., correlation coefficient (r), coefficient of determination (r^2), root mean square error (RMSE), reduction of error (RE), and index of agreement (IA) for training, validation and testing. These results are presented in Table 5 and in Appendix Fig. 13. Theoretically, IA ranges from 0 to 1, with a value 1 indicating a perfect association. The IA statistic specifies the degree to which the observed deviations about the observed mean in magnitude and sign correspond to the predicted/reconstructed deviations about the observed mean. This also

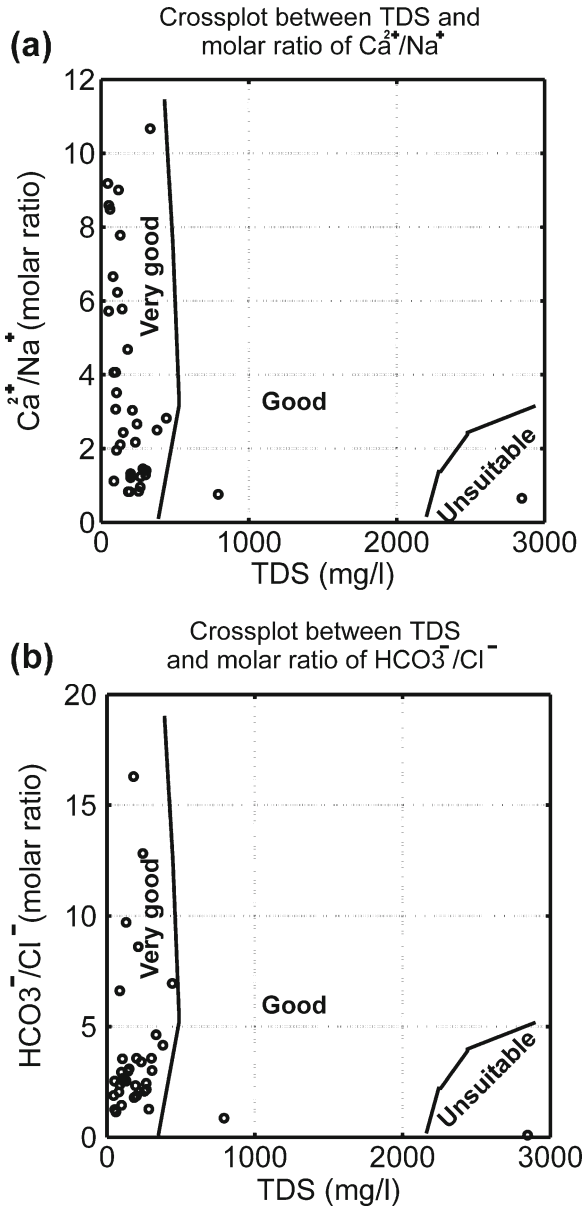


Fig. 5 Relationship between total dissolved solids (TDS) and molar ratios of **a** Ca^{2+}/Na^{+} and **b** $HCO_3^{-1}/C1^{-}$

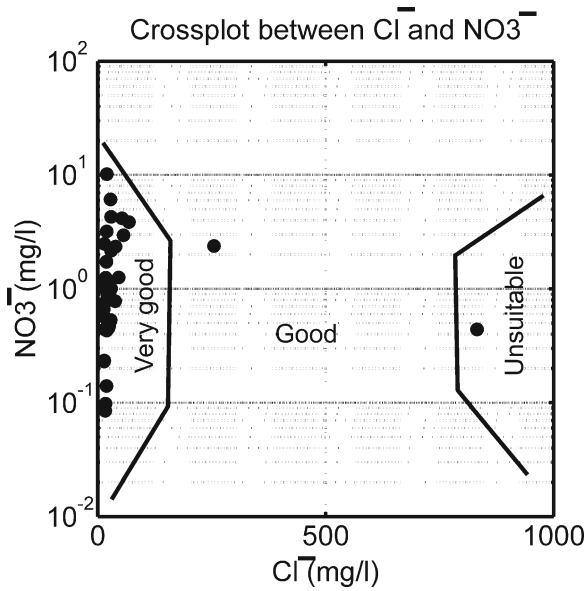


Fig. 6 Cross plot between Cl^{-} and NO_3^{-}. The graph is of semi-log type

reflects sensitivity to both the differences in observed and predicted means and changes in proportionality (Willmott 1981). The coefficient of determination (r^2) is a measure of how well regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it would be able to explain all of the variation. The farther the line from the points, lesser the explanation. The coefficient of determination (r^2) between estimated GWG and target GWQ is 0.84 for validation set, which means that 84 % of the total variation in estimated GWG can be explained by the linear relationship between estimated GWG and target GWQ (Appendix Fig. 12). The coefficient of determination (r^2) between estimated GWG and target GWQ is 0.82 for test data set, which means that 82 % of the total variation in estimated GWG can be explained by the linear relationship between estimated GWG and target GWQ (Appendix Fig. 12). Performance of the model on training, validation, and test set is given in Table 5. In the present case, IA value is above 0.9 for training, validation, and testing case, which is of excellent quality (Table 5). RE value above 0.9 suggests that the predictive result is reliable for the present case (Table 5). RMSE value of training, validation, and test sets are satisfactory ranging between 0.70 and 0.80 (Table 5).

Table 5 Performance of Bayesian neural network model for groundwater quality classification

Data set	r	RMSE	RE	IA	r^2
Training data	0.99	0.79	0.99	0.99	0.98
Validation data	0.92	0.74	0.95	0.94	0.84
Testing data	0.91	0.73	0.94	0.93	0.82

The results of the applications of the algorithm on the real data are presented in Fig. 7. Figure 7 represents BNN-based classification model of groundwater quality (GWQ) of the study area. The results indicate that in most of the places in the study area, GWQ falls in the range of "good" to "very good" category except for Kelus and Shiroda where the BNN output is of the order of 0.30 for both places. Viewing collinear characteristics which are not uncommon on the geochemical components, it can be noted that the present approach is based on equal weightage assigned to all geochemical and geophysical attributes. Thus, the solution derived by this approach is somewhat more unbiased. Further, in order to augment authenticity of the present results, we performed uncertainty analysis. The result of the uncertainty in prediction of the network model is presented in Fig. 8. Results show that uncertainty value varies between 0.00 (meaning that more confidence level) and 1.00 (meaning that less confidence level) (Fig. 8). Prediction results are reliable because the uncertainty value found at most of the places is in the range between 0.00 and 0.30, which is quite satisfactory. Note that uncertainty results are satisfactory at both Kelus and Shiroda where GWQ is found "unsuitable". The uncertainty value is maximum at Nerur (~1.00) followed by Matond (~0.98). This could be attributed either due to the under sampling of model space or due to the presence of

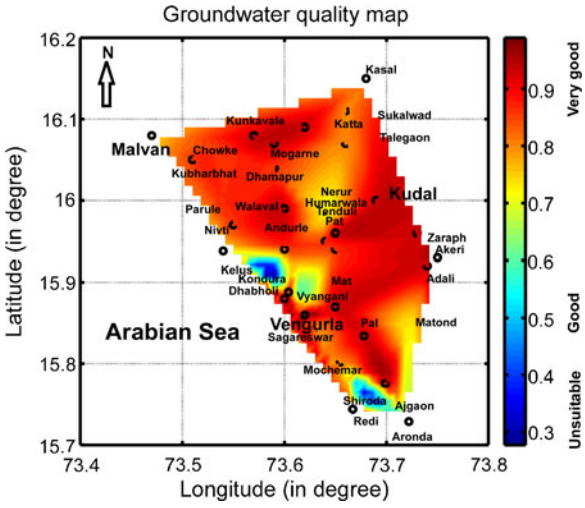


Fig. 7 Bayesian neural network (BNN) classification of groundwater quality (GWQ) map of the study area. The surface maps were constructed using triangle-based cubic interpolation in Matlab

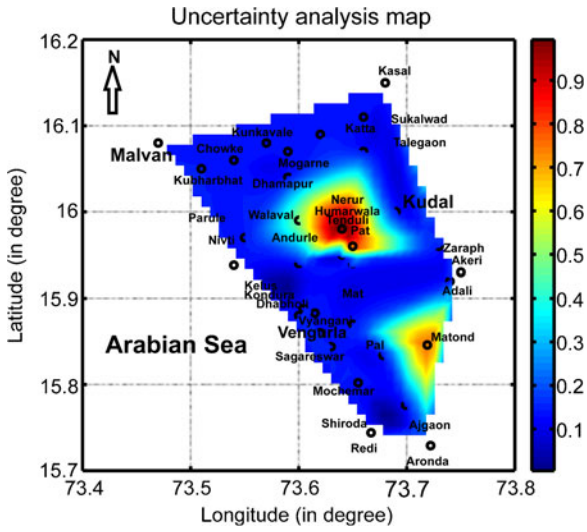


Fig. 8 Standard deviation (STD) map of the predictive distribution of the network output of the study area. The surface maps were constructed using triangle-based cubic interpolation in Matlab

inescapable red noise in real data. Alternatively, this may be also attributed due to the intrinsic noise embodied in the real data and/or theoretical error described by the posterior distribution of the network weight vector (Bishop 1995). Whatever may be the reason, it is significant to mention here that the present method based on the Bayesian framework provides genuine scope to measure uncertainty limits at the network output which is an additional advantage to justify our results. This uncertainty limit in the result may be used as a helpful guide for future model prediction.

Conclusions

A new ANN algorithm is developed utilizing the concept of Bayesian statistics and Hybrid Monte Carlo global optimization scheme for assessing the groundwater quality. The modeling strategy uses the fusion of geochemical and geophysical information, which essentially leads to multi-parameter optimization. The new method reduces the uncertainty caused by the data gaps (for example, sampling well is absent but VES is possible or vice versa), which is generally not possible in non-integrative conventional approaches. This particular analysis, in fact, gives us more confidence to appropriately making the data interpretation. We mention,

however, that uncertainty measure is a significant statistical parameter and, therefore, appropriate care should be taken while calculating and assigning the confidence limit as discussed earlier and demonstrated in Fig. 8 for any geochemical data analysis and their interpretation.

Prior to the application of the method on real data, the algorithm is tested with two independent data (validation and testing) sets which provide very good and consistent results with accuracy greater than 82.5 %. The method is then successfully applied to the actual data collected from the parts of coastal region of Maharashtra, Western India. The BNN model results suggest that groundwater quality is very good in most of the regions of the study area, except for Kelus and Shiroda, and thus may be recommended for domestic uses.

Gibb's diagram indicates that the groundwater quality is mainly controlled by rock–water interactions (rock dominance). The cross plots of hydrochemical parameters support the groundwater quality results, which are further confirmed by the BNN-based classification scheme. The model is developed by considering all possible physical and chemical attributes with a wide range of parameterization. The trained model provides a viable solution to the new data. The results obtained here may provide constraints for its implementation in future management, planning, remedial measures, and monitoring of natural groundwater where groundwater qualities are under threat due to various reasons including anthropogenic disturbances.

Keeping in view the seasonal changes of groundwater chemistry and also to establish a more authentic and sound chemical integrity of the entire study area, it may be proposed that more water samples data in different seasons and their repeat analysis is required in future work.

Acknowledgments We are thankful to the Directors, Indian Institute of Geomagnetism, New-Panvel and NGRI, Hyderabad for their kind permission to publish the work.

Appendix

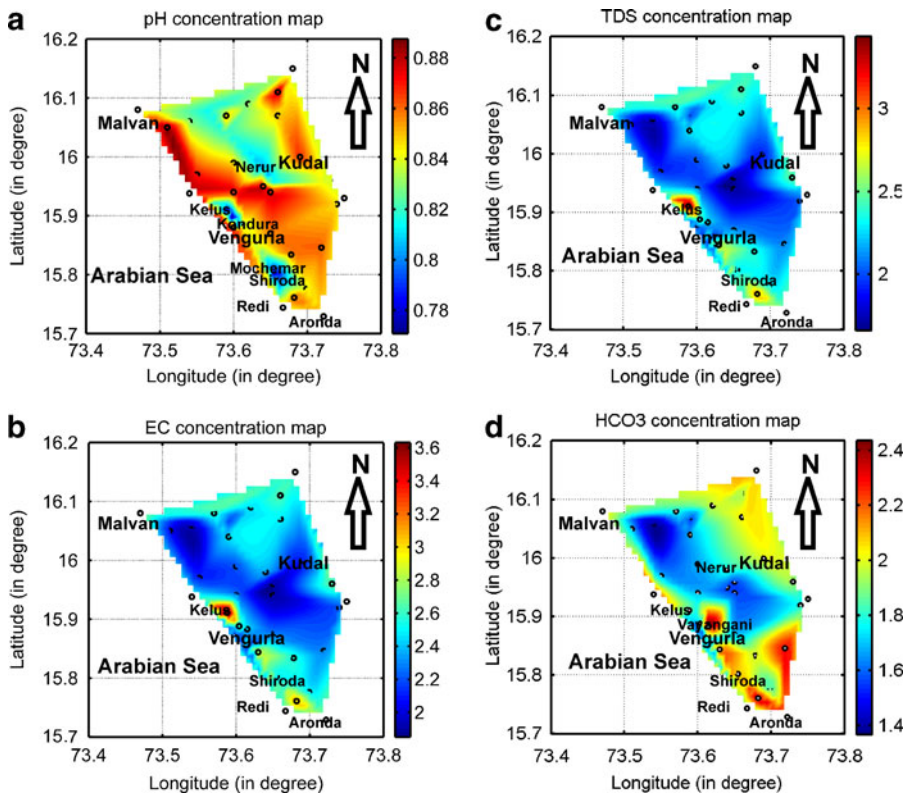


Fig. 9 Concentration map of **a** pH, **b** electrical conductivity (EC), **c** total dissolved solids (TDS), and **d** bicarbonate $\left(HCO_3^-\right)$. Units are expressed in log10. The surface maps were constructed using triangle-based cubic interpolation in Matlab

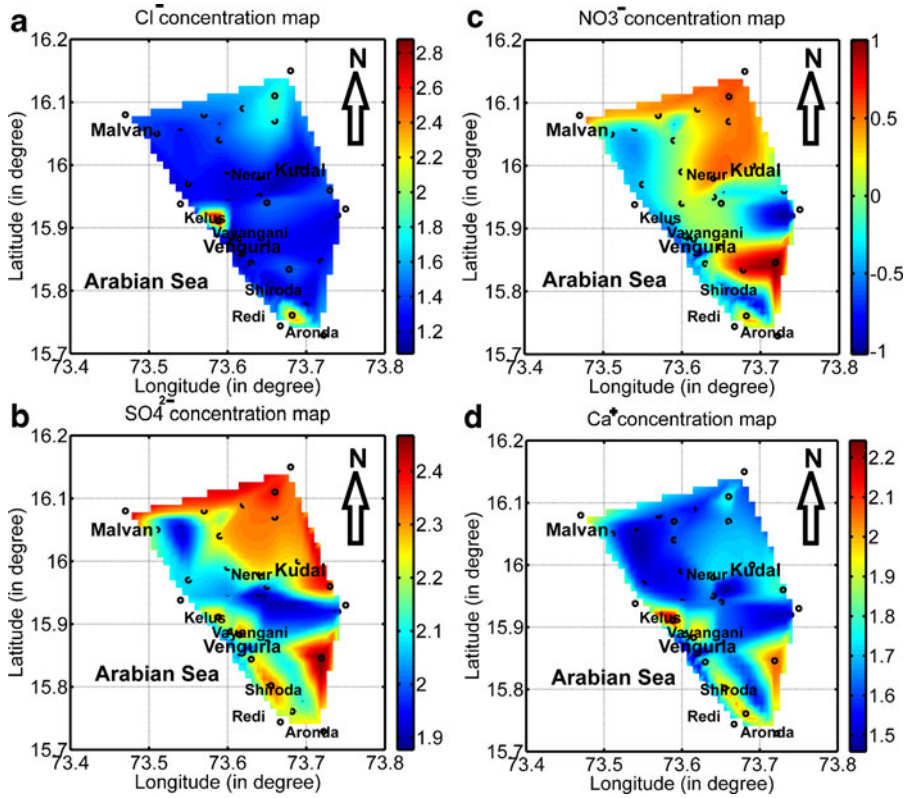


Fig. 10 Concentration map of **a** chloride (Cl^-), **b** sulfate (SO_4^{2-}), **c** nitrate (NO_3^-), and **d** calcium (Ca^{2+}). Units are expressed in log10. The surface maps were constructed using triangle-based cubic interpolation in Matlab

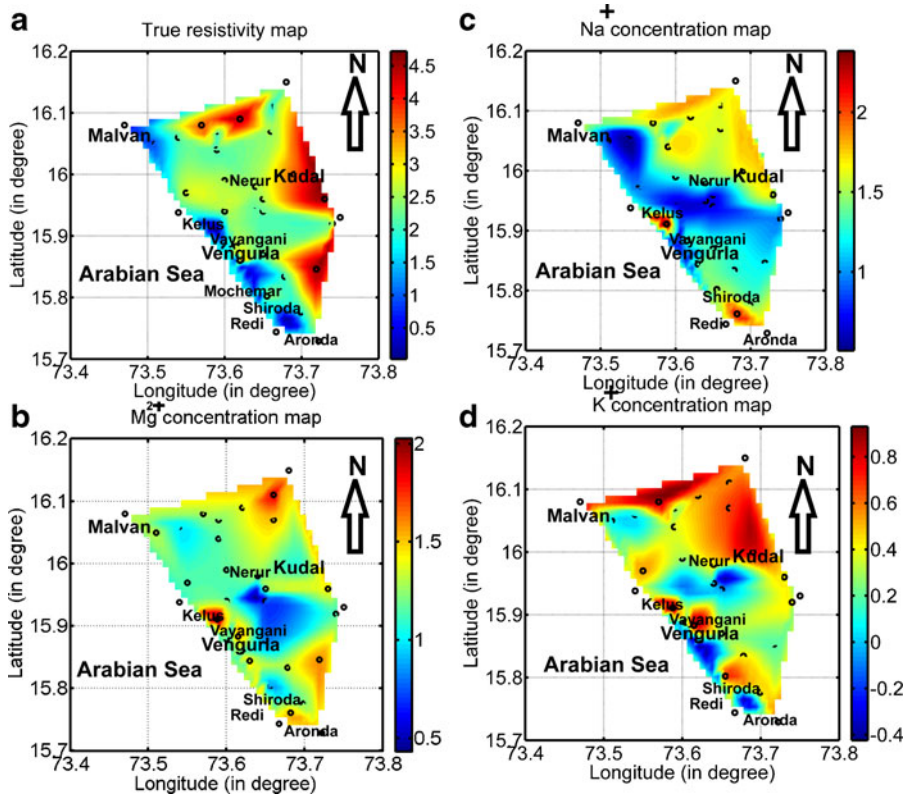


Fig. 11 Concentration map of **a** true resistivity (ρ), **b** magnesium (Mg^{2+}), **c** sodium (Na^+), and **d** potassium (K^+). Units are expressed in log10. The surface maps were constructed using triangle-based cubic interpolation in Matlab

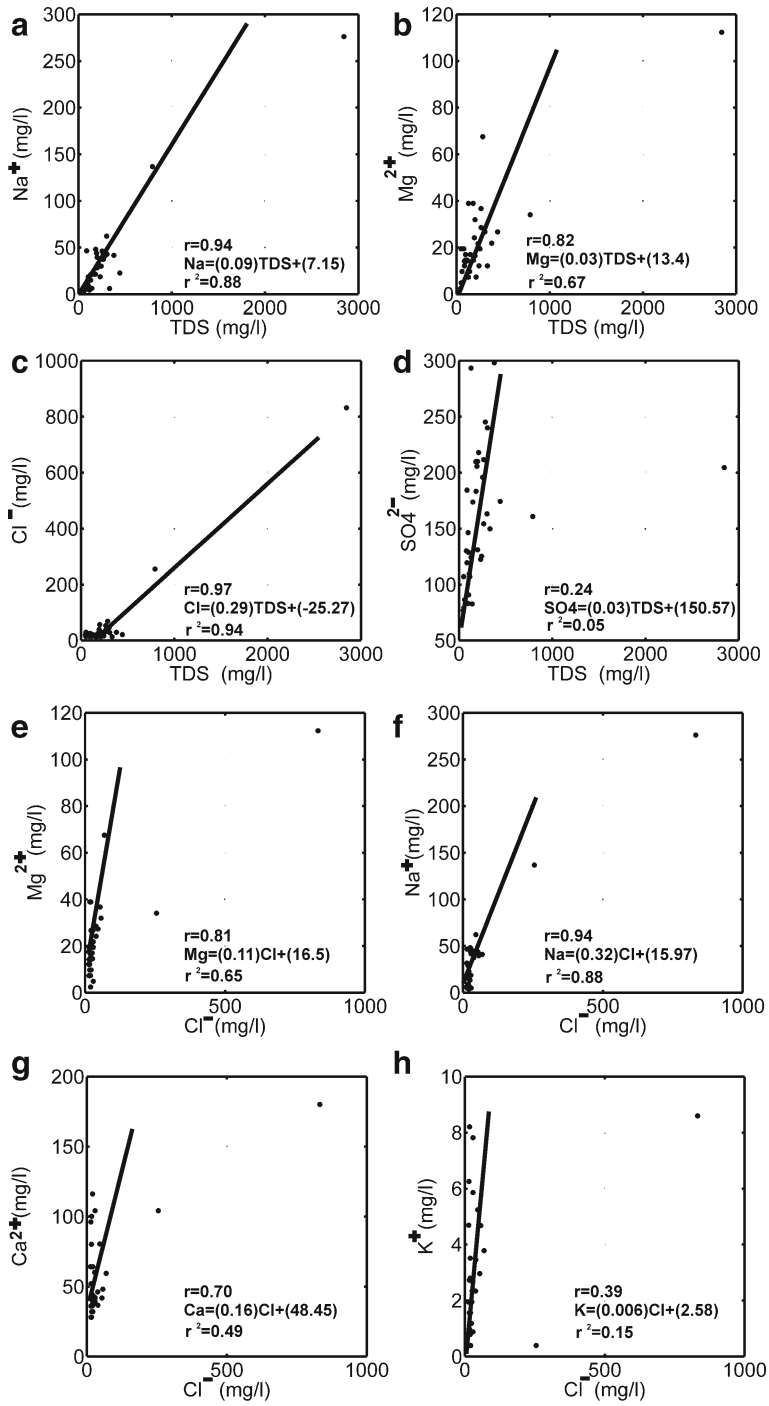


Fig. 12 Cross plots of **a**–**b** Na^{+} and Mg^{2+} with TDS, **c**–**d** Cl^{-} and SO_4^{2-} with TDS, **e**–**f** Mg^{2+} and Na^{+} with Cl^{-}, and **g**–**h** Ca^{2+} and K^{+} with Cl^{-}

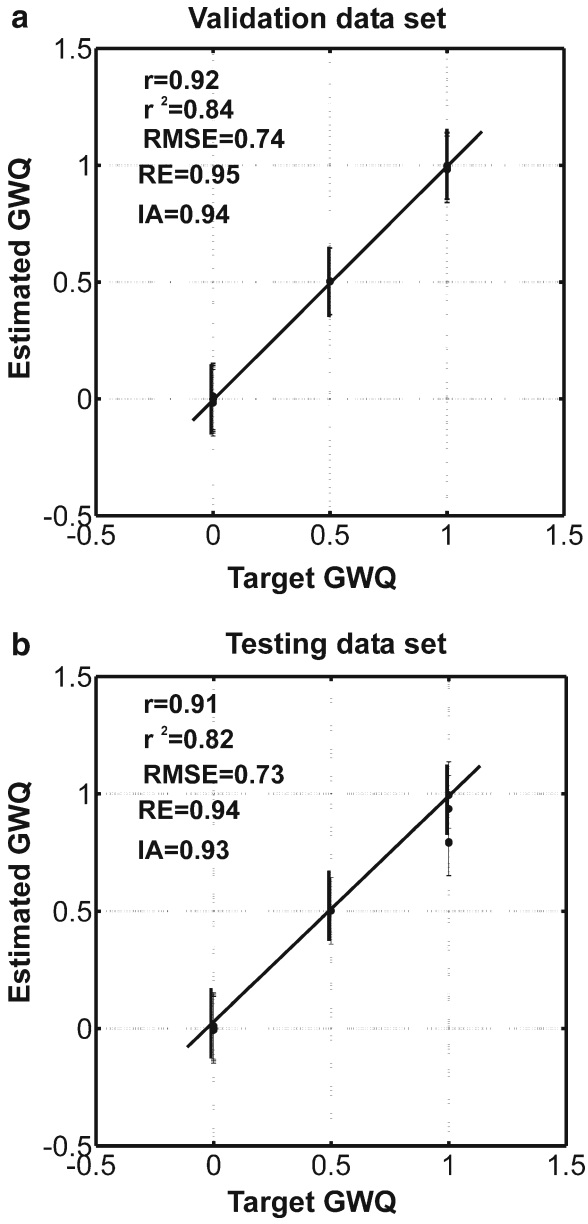


Fig. 13 Comparison of the results of the best Bayesian neural network (BNN) model and observed data. *Error bar* shows 90 % confidence level

References

American Public Health Association (APHA). (1985). *Standard methods for the examination of water and waste* (16th ed., p. 100). Washington: American Public Health Association.

Appelo, C. A. J., & Postma, D. (2005). *Geochemistry, groundwater and pollution* (2nd ed., p. 404). Leiden: Balkema.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Bohlke, J. K. (2002). Groundwater recharge and agricultural contamination. *Hydrogeology Journal, 10*, 153–179.

Brown, C. H. (1998). *Applied multivariate statistics in geohydrology and related sciences*. Berlin: Springer.

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

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

Duane, S., Kennedy, A. D., Pendleton, B., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B, 195*, 216–222.

Duraiswami, R., & Patankar, U. (2011). Occurrence of fluoride in the drinking water sources from Gad River basin, Maharashtra. *Journal of Geological Society of India, 77*, 167–174.

Fetter, C. W. (1988). *Applied hydrogeology*. Columbus: Merrill.

Gibbs, R. J. (1970). Mechanisms controlling world water chemistry. *Science, 170*, 1088–1090.

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

Kannan, N., & Joseph, S. (2010). Quality of groundwater in the shallow aquifers of a paddy dominated agricultural river basin, Kerala, India. *International Journal of Civil Environmental Engineering, 2*, 160–178.

Long, A. J., & Valder, J. F. (2011). Multivariate analyses with end-member mixing to characterize groundwater flow: wind cave and associated aquifers. *Journal of Hydrology, 409*, 315–327.

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. *Journal of Geophysical Research, 115*, B10208. doi:10.1029/2010JB000864.

Maiti, S., Gupta, G., Erram, V. C., & Tiwari, R. K. (2011). Inversion of Schlumberger resistivity sounding data from the critically dynamic Koyna region using Hybrid Monte Carlo-based neural network approach. *Nonlinear Processes in Geophysics, 18*, 179–192. doi:10.5194/npg-18-179-2011.

Maiti, S., Gupta, G., Erram, V. C., & Tiwari, R. K. (2012). Delineation of shallow resistivity structure around Malvan, Konkan region, Maharashtra by neural network inversion using vertical electrical sounding measurements. *Environmental Earth Sciences*. doi:10.1007/s12665-012-1779-8.

Matlab version 7.10.0. (2010). Natick, Massachusetts: The Math Works Inc., (http://www.mathworks.com).

Menke, W. (1984). *Geophysical data analysis: discrete inverse theory*. New York: Academic.

Mondal, N. C., Singh, V. S., Saxena, V. K., & Singh, V. P. (2011). Assessment of seawater impact using major hydrochemical ions: a case study from Sadras, Tamilnadu, India. *Environmental Monitoring and Assessment, 177*, 315–335.

Nabney, I. T. (2004). *Netlab algorithms for pattern recognition*. New York: Springer.

Naik, P., Awashti, A. K., Anand, A. V. S. S., & Behera, P. N. (2009). Hydrogeochemistry of the Koyna River basin, India. *Environmental Earth Sciences, 59*, 613–629.

Panda, U. C., Sundaray, S. K., Rath, P., Nayak, B. B., & Bhatta, D. (2006). Application of factor and cluster analysis for characterization of river and estuarine water systems—a case study: Mahanadi River (India). *Journal of Hydrology, 331*, 434–445.

Park, S. C., Yun, S., Chae, G. T., Yoo, I. S., Shin, K. S., & Heo, C. H. (2005). Regional hydrochemical study on salinization of coastal aquifers, western coastal area of South Korea. *Journal of Hydrology, 313*, 182–194.

Pawar, N. J., Pondhe, G. M., & Patil, S. F. (1998). Groundwater pollution due to sugar-mill effluent at Sonai, Maharashtra, India. *Environmental Geology, 34*(213), 151–158.

Poulton, M. (2001). *Computational neural networks for geophysical data processing*. New York: Pergamon.

Pradhan, B., & Pirasteh, S. (2011). Hydro-chemical analysis of the ground water of the basaltic catchments: Upper Bhatsai Region, Maharashtra. *The Open Hydrology Journal, 5*, 51–57.

Prasanna, M. V., Chidambaram, S., Shahul, A., Hameed, A., & Srinivasamoorthy, K. (2011). Hydrogeochemical analysis and evaluation of groundwater quality in the Gadilam river basin, Tamil Nadu, India. *Journal of Earth System Science, 120*, 85–98.

Raghunath, R., Sreedhara Murthy, T. R., & Raghavan, B. R. (2001). Spatial distribution of pH, EC and total dissolved solids of Nethravathi river basin, Karnataka state, India. *Pollution Research, 20*, 413–418.

Rajmohan, N., & Elango, L. (2006). Hydrogeochemistry and its relation to groundwater level fluctuation in the Palar and Cheyyar river basins, southern India. *Hydrological Process, 20*, 2415–2427.

Rao, P. M., Sekhar, P., & Yadav, Y. S. (2004). Water quality studies on Kolleru lake and its infalling drains of A.P., India. In A. Kumar (Ed.), *Water pollution* (p. 171). New Delhi: APH.

Schoeller, H. (1959). *Arid zone hydrology, recent developments*. Paris: UNESCO.

Schoeller, H. (1967). *Methods and techniques of groundwater investigation and development* (Water Resources Series 33). Paris: UNESCO.

Song, S. H., Lee, J. Y., & Park, N. (2007). Use of vertical electrical soundings to delineate seawater intrusion in a coastal area of Byunsan, Korea. *Environmental Geology, 52*, 1207–1219.

Subba Rao, N. (2006). Seasonal variation of groundwater quality in a part of Guntur district, Andhra Pradesh, India. *Environmental Geology, 49*, 413–429.

Sun, T., Pan, S. B., & Li, Y. J. (2004). Application of artificial neural network model to groundwater quality assessment and classification. *Hydrogeology and Engineering Geology, 3*, 58–61.

Tarantola, A. (1987). *Inverse problem theory*. New York: Elsevier.

Tiwari, R. K., & Maiti, S. (2011). Bayesian neural network modeling of tree-ring temperature variability record from the Western Himalayas. *Nonlinear Processes in Geophysics, 18*, 515–528. doi:10.5194/npg-18-515-2011.

Todd, D. K. (1959). *Groundwater hydrology* (pp. 277–294). New York: Wiley.

Van der Bann, M., & Jutten, C. (2000). Neural networks in geophysical applications. *Geophysics, 65*, 1032–1047.

Willmott, C. J. (1981). On the validation of models. *Physical Geography, 2*, 184–194.

World Health Organization (WHO). (1984). *Guideline of drinking quality* (pp. 333–335). Washington: World Health Organization.

Zou, Z. H., & Wang, H. (2007). Application of BP modeling based on random samples to assessment on natural water quality. *Environmental Engineering, 25*, 69–71.

Zou, Z. H., & Wang, H. (2010). Adaptive neuro fuzzy inference system for classification of water quality status. *Journal of Environmental Sciences, 22*, 1891–1896.