



Groundwater quality assessment of urban Bengaluru using multivariate statistical techniques

Mohammad Shahid Gulgundi¹ · Amba Shetty¹

Received: 22 June 2017 / Accepted: 16 February 2018 / Published online: 24 February 2018
© The Author(s) 2018. This article is an open access publication

Abstract

Groundwater quality deterioration due to anthropogenic activities has become a subject of prime concern. The objective of the study was to assess the spatial and temporal variations in groundwater quality and to identify the sources in the western half of the Bengaluru city using multivariate statistical techniques. Water quality index rating was calculated for pre and post monsoon seasons to quantify overall water quality for human consumption. The post-monsoon samples show signs of poor quality in drinking purpose compared to pre-monsoon. Cluster analysis (CA), principal component analysis (PCA) and discriminant analysis (DA) were applied to the groundwater quality data measured on 14 parameters from 67 sites distributed across the city. Hierarchical cluster analysis (CA) grouped the 67 sampling stations into two groups, cluster 1 having high pollution and cluster 2 having lesser pollution. Discriminant analysis (DA) was applied to delineate the most meaningful parameters accounting for temporal and spatial variations in groundwater quality of the study area. Temporal DA identified pH as the most important parameter, which discriminates between water quality in the pre-monsoon and post-monsoon seasons and accounts for 72% seasonal assignment of cases. Spatial DA identified Mg, Cl and NO₃ as the three most important parameters discriminating between two clusters and accounting for 89% spatial assignment of cases. Principal component analysis was applied to the dataset obtained from the two clusters, which evolved three factors in each cluster, explaining 85.4 and 84% of the total variance, respectively. Varifactors obtained from principal component analysis showed that groundwater quality variation is mainly explained by dissolution of minerals from rock water interactions in the aquifer, effect of anthropogenic activities and ion exchange processes in water.

Keywords Multivariate statistical techniques · Groundwater quality · Cluster analysis · Discriminant analysis · Principal component analysis/factor analysis

Introduction

The dependence on groundwater has gone up over the years in most of the urban areas due to inadequacy of surface water resources to meet the water requirements. Majority of Indian states are withdrawing groundwater for both agricultural and industrial purposes at a rate more than what can be recharged (Jat et al. 2008). Groundwater problems to a great extent are the consequence of human activities like uncontrolled withdrawal of borewell water at a high rate compared to recharge rate (Kazi et al. 2009). In any area, the characteristics of

groundwater are due to natural and anthropogenic processes, which have the capability to alter these systems by contaminating them or modifying the hydrological cycle (Helena et al. 2000; Kumar et al. 2017). When pollution of groundwater in aquifers happens, it perseveres for a long time as a result of slow movement of water in them. The harmful impacts of rural and modern exercises and urban advancement on adjoining groundwater have incited examinations on the nature of these sources (Dawoud and Raouf 2009). It is consequently desirable to ensure that groundwater quality is secured for its utilization in different purposes (Jammel and Hussain 2003; Tirkey et al. 2017). To protect groundwater quality for drinking, periodical monitoring of its quality is essential in urban regions. But testing of different water quality parameters is costly, takes lot of time and is a tedious process. Also measurement of all the parameters at a consistent interim is not required since this will not give extra

✉ Mohammad Shahid Gulgundi
shahid.env@gmail.com; amba@nitk.edu.in

¹ Department of Applied Mechanics and Hydraulics,
National Institute of Technology Karnataka,
Surathkal, Srinivasanagara, P.O. 575 025, India

data on the water quality aspect (Mustapha and Aris 2012a). In order to aid the administration to prioritize and to make informed decisions so as to improve the groundwater quality, it is very important to reduce the apprehensions involved in the dataset by interpreting the spatial and temporal variations in water quality (Wang et al. 2008) and also to locate hidden pollution sources (Zhang et al. 2009).

In recent years, few data-driven approaches like the projection pursuit technique and neural networks have been used for assessing the water quality (Salman and Ruka'h 1999). Water Quality Index (WQI) is regarded as one of the most effective way to communicate water quality (Sadat-Noori et al. 2014). Horton (1965) suggested that various water quality data could be aggregated into an overall index. However, multivariate statistical techniques can be employed for analyzing huge water quality datasets with minimal loss of important information (Juahir et al. 2011; Samson and Elangovan 2017; Shrestha and Kazama 2007). Multivariate statistical techniques, such as cluster analysis (CA), principal component analysis (PCA), factor analysis (FA) and discriminant analysis (DA) can interpret complex data matrices for improved understanding of water quality and other environmental systems by allowing the identification of possible factors/sources thus serving as a worthy tool for quickly solving pollution problems (Vega et al. 1998; Lee et al. 2001; Wunderlin et al. 2001; Reghunath et al. 2002; Simeonov et al. 2003, 2004; Ravikumar and Somashekar 2017). Principal component analysis (PCA) has been utilized to take out the noise from huge data matrix and classify the variables into measurable components, discriminant analysis (DA) recognizes the most segregating measurable

element/variable according to goodness and cluster analysis (CA) chooses the identical group inside a specific data set. Characterization and evaluation of surface and freshwater quality performed by multivariate statistical techniques has proved to be useful in verifying spatial and temporal variations caused naturally and due to human induced factors also (Helena et al. 2000; Singh et al. 2004, 2005; Hassen et al. 2016).

Bengaluru has suddenly overgrown its size after the Information Technology boom. Consequent to this the city and the district administration is struggling to provide necessary infrastructure. The demand for water supply in particular requires scientific planning and effective management of water resources, especially the groundwater in the district (CGWB 2012). In this study, groundwater quality data measured during pre and post monsoon on 14 parameters from 67 sites distributed across the western half of the Bengaluru city were subjected to different multivariate statistical approaches (CA, DA, PCA/FA) in order to evaluate the temporal and spatial variations in groundwater quality caused by parameters and to recognize the likely factors causing variation in groundwater quality.

Materials and methods

Study area

The study area (Fig. 1) is situated in the northwestern and southwestern corner of Bengaluru city, between $12^{\circ}48'24.52''$ and $12^{\circ}53'59.85''$ North latitude and

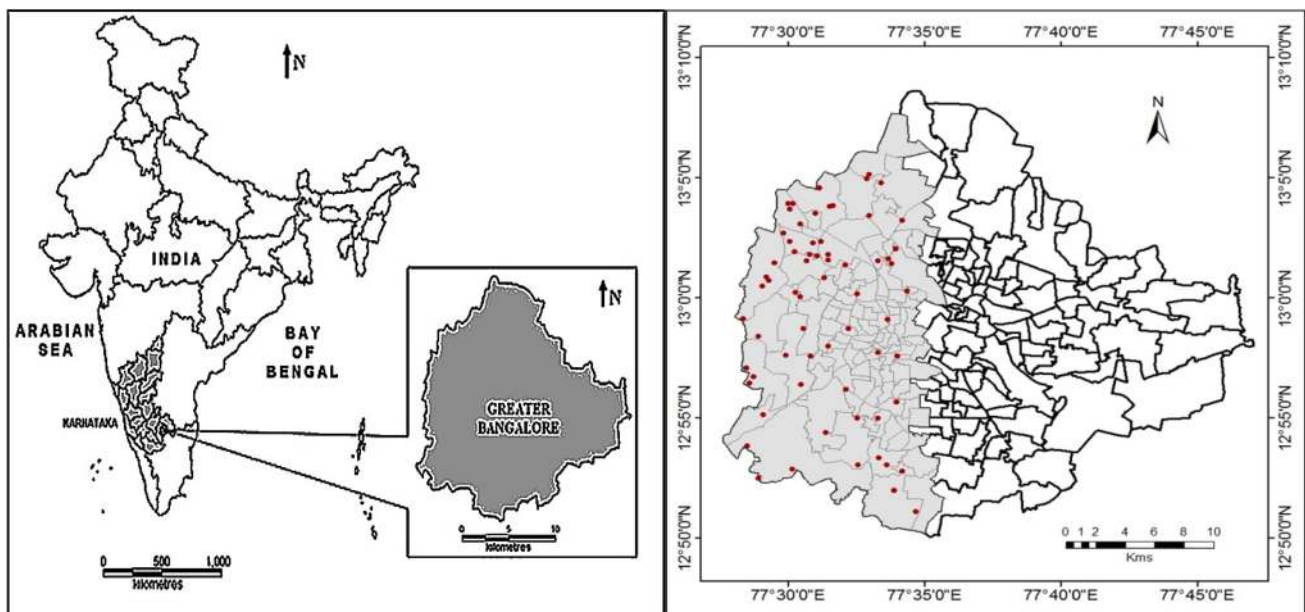


Fig. 1 Location map of the study area

77°24'59.95" to 77°30'6.72" East Longitude and spreads over a region of 241 km². It gets precipitation from both upper east and the southwest storms with yearly aggregate precipitation of around 900 mm. Bengaluru city is for the most part depleted by part of the Arkavathi river catchment toward the west and South Pennar river toward the east. The versatility, presence and aquifer refill of groundwater event are dominated by the measure of weathering, fracture pattern, geomorphological setup and rainfall. The Bangalore urban district contains crystalline storm cellar, fundamentally gneisses and rocks meddled by essential dykes. These arrangements have been modified to laterite along the eastern edge of the city. The city is intensely reliant on groundwater for its household and commercial needs. The appraisal of groundwater asset demonstrates that the asset is over misused. As a result of this overuse, groundwater quality has additionally disintegrated (DMG 2003, 2011).

Hydrogeology

Granites and Gneisses of peninsular gneissic group form the primary aquifers in the study region. Laterites of tertiary age occur as isolated patches capping crystalline rocks. Alluvium of limited thickness and aerial extent 20–25 m thick occur along the river courses possessing substantial groundwater potential. Groundwater occurs in phreatic conditions or unconfined conditions in the weathered zone and under semi-confined to confined conditions in fractured and jointed rock formations. Groundwater movement and recharge of aquifers are controlled by various factors like fracture pattern, degree of weathering, geo-morphological setup and amount of rainfall received. The resistivity examinations uncovered the presence of an exceedingly weathered rock (permeable) reaching out up to a depth of 30 m. The principal aquifer exists between 25 and 30 m depth. There are aquifers even past 60 m depth. The area is sloping towards west. Streams of various watersheds start from this area. Significant piece of the study zone is possessed by streams streaming towards west from this region (DMG 2011; CGWB 2012).

Monitored parameters

A total of 67 groundwater samples were collected in the month of March (2014) for pre-monsoon and November (2014) for post-monsoon seasons. The sampling locations were selected with a view to cover residential, industrial and commercial areas so as to achieve a good sampling representation over the study area. The samples were collected from bore wells after 10 min of pumping in pre-cleaned sterilized plastic bottles and stored in an ice box. The samples collected were analyzed for 14 physico-chemical parameters, namely pH, total dissolved solids (TDS), electrical conductivity

(EC), nitrate (NO₃⁻), chloride (Cl⁻), sulfate (SO₄²⁻) magnesium (Mg²⁺), sodium (Na⁺), calcium (Ca²⁺), potassium (K⁺), iron (Fe), alkalinity (HCO₃⁻), total hardness (TH) and fluoride (F⁻). Electrical conductivity and pH were measured in the field immediately after sampling and the remaining parameters were determined in laboratory within 24 h.

Analysis methods

The sampling, preservation, transportation and analysis of water samples were performed according to standard methods (APHA 2005). The analytical data quality was ensured through careful standardization, procedural blank measurements and spiked and duplicate samples. Calcium and magnesium were determined by EDTA titrations method, sodium and potassium by flame emission photometry, iron by phenanthroline spectrophotometry, bicarbonate and carbonate by titrimetry, chloride by argnetometric titration, nitrate by UV spectrometry, sulfate by nephelometry, total dissolved solids by gravimetry, total hardness by potentiometry and flouride by ion selection electrode method. pH and electrical conductivity were measured in situ using digital portable water analyser (Systronics –371). pH meter was calibrated by immersing the probe using two standard solutions (pH 4 and 10 buffers) while electrical conductivity meter was calibrated by immersing the probe in standard KCl solution (0.1 N). The accuracy of the chemical analysis was verified by calculating ion-balance errors using Aquachem, where the errors were generally around 5%.

Data pretreatment

The statistical analysis of data was carried out using SPSS software, v 20.0. The methods, such as CA and FA, require variables to conform to a normal distribution. Normal distribution of data is an essential requirement for multivariate statistical analyses because the analyses will be valid only if the standard deviations (variances) are low (very close to 0). Else, the parameters with the highest variances will influence the analysis (Güler et al. 2002; Cloutier et al. 2008; Yidana et al. 2011; Boateng et al. 2016). The raw data indicated that Ca, Na, HCO₃⁻ and SO₄²⁻ were very close to normal distribution, but the distribution pattern of other parameters was not normal. Hence these parameters were log transformed to make the data to have normal distribution (Zhang et al. 2009). The standard z-scores of all the parameters were then used for the multivariate statistical analysis to lessen the effects of differences in the units used for measurement and variance and to render the data dimensionless (Singh et al. 2005; Yidana et al. 2011). The z scores were calculated as in Eq. (1):

$$z = \frac{x - \bar{x}}{s} \quad (1)$$

where x represents the value, \bar{x} represents the mean and s represents the standard deviation of the parameter, at a given sampling site.

Water quality index (WQI)

A WQI is a single number (like a grade) that expresses overall water quality at a certain location and time based on several water quality parameters. The main purpose of WQI is to turn complex water quality data into information that is understandable and usable by the public. WQI is a single unit less number of 100-point scale that provides a pointer to the quality of water source (Pradhan et al. 2001; Pius et al. 2012). According to this water quality index, the maximum permissible value is 100. Values greater than 100 indicate pollution and are unfit for human consumption. The methodology considered for development of the WQI is adopted from Tiwari and Mishra (1985) as in Eq. (2):

$$\text{WQI} = \text{Anti log} \left[\sum_{n=1}^n W_n \log_{10} q_n \right], \quad (2)$$

where weightage factor (W) is computed using $W_n = K/S_n$ and K is proportionality constant derived from Eq. (3):

$$K = \left[1 / \sum_{n=1}^n 1/s_i \right], \quad (3)$$

where S_n and S_i are the WHO/ICMR standard values of the water quality parameter. Quality rating (q) is calculated using $q_n = \{[(V_{\text{actual}} - V_{\text{ideal}})/(V_{\text{standard}} - V_{\text{ideal}})] \times 100\}$, where q_n = quality rating of i th parameter for a total of n water quality parameters, V_{actual} = value of the water quality parameter obtained from laboratory analysis, V_{ideal} = value of water quality parameter that can be obtained from the standard tables, V_{ideal} for pH 7 and for other parameters is equivalent to zero, V_{standard} = WHO/ICMR standard of the water quality parameter. Based on the above WQI values, the ground water quality is rated as excellent, good, poor, very poor and unfit for human consumption (Table 3).

Cluster analysis (CA)

CA is one of the multivariate techniques, which groups the objects based on their characteristics. It arranges the objects, such that every object is same as the others in the cluster according to a predefined selection criterion. The clusters of objects obtained should then display high internal (within-cluster) resemblance and high external (between clusters) diversity. Hierarchical agglomerative clustering is the most commonly used approach (Massart and Kaufman 1983), which supplies with instinctive similarity relationships

between any one sample and the entire data set. It is represented by a dendrogram (tree diagram) (McKenna 2003). The dendrogram displays a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a reasonable lessening in dimensionality of the original data. The Euclidean distance shows the similarity between two samples and a distance can be represented by the difference between analytical values from the samples (Forina et al. 2002; Taoufik et al. 2017).

Using Ward's method on the normalized data set, hierarchical agglomerative cluster analysis was conducted in this study. To measure the similarity squared euclidean distance was used. The ward's method makes use of an analysis of variance approach for evaluating the distances between clusters, in order to minimize the sum of squares (SS) of any two clusters that can be formed at each step (Willet 1987; Adams 1998; Otto 1998; Tziritis et al. 2016). Using the linkage distance, the spatial variability of groundwater quality for the study area was determined from cluster analysis, which is reported as $D_{\text{link}}/D_{\text{max}}$. $D_{\text{link}}/D_{\text{max}}$ represents the quotient between the linkage distances for a particular case divided by the maximal linkage distance. To standardize the linkage distance, which is represented on the y -axis, the quotient is then multiplied by 100 (Simeonov et al. 2003; Singh et al. 2005).

Discriminant analysis (DA)

DA is a supervised pattern recognition technique, which is used for the classification of objects or cases into exhaustive and mutually exclusive groups based on a set of independent variables. It is a suitable statistical technique when the dependent variable is a categorical variable and the independent variables are metric (Mustapha and Aris 2012). The purpose of DA is to increase the similarity between-group relative to the within-group variance. DA finds out the variables that discriminate between two or more expected occurring groups (Johnson and Wichern 1992). It also forms a discriminant function (DF) for each group as in Eq. (4):

$$f(G_i) = k_i + \sum_{j=1}^n w_{ij} p_{ij}, \quad (4)$$

where i is the number of groups (G), k_i the constant inherent to each group, n the number of parameters used to classify a set of data into a given group, w_j the weight coefficient, assigned by DA to a given selected parameter (p_j).

In the present study, DA was carried out on raw data using three different modes: standard, forward stepwise and backward stepwise to construct discriminant functions (DFs) and to assess both temporal and spatial variations in groundwater quality. Temporal DA was done taking the monitoring period (pre-monsoon and

post-monsoon) as the grouping variable and the 14 measured groundwater quality parameters as the independent variables. Spatial DA was done in the same way as temporal DA, by taking the spatial clusters obtained in cluster analysis as the grouping variable and the 14 measured water quality parameters as the independent variables.

Principal component analysis/factor analysis

PCA is a technique, which converts the original variables into new uncorrelated variables (axes), known as principal components, which are linear combinations of the original variables (Sarbu and Pop 2005). The new axes lie in the directions where variance is maximum (Hossain et al. 2015). PCA supplies the details of most significant parameters, which describes the whole data set thereby reducing the data with minimal loss of original information (Helena et al. 2000). The principal component (PC) can be expressed as in Eq. (5):

$$z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + a_{i3}x_{3j} + \dots + a_{im}x_{mj}, \quad (5)$$

where a is the component loading, z the component score, x the measured value of a variable, I the component number, j the sample number and m the total number of variables.

PCA is continued with factor analysis. The objective of factor analysis is to lessen the inputs from unimportant variables in order to further simplify the data structure obtained from PCA (Aris et al. 2012; Noshadi and Ghafourian 2016). This objective can be accomplished by rotating the axis defined by PCA, according to well-established rules, and generating new variables, called varifactors (VF). A principal component is a linear combination of observable water quality variables, whereas varifactor can include unobservable, hypothetical, latent variables (Vega et al. 1998; Helena et al. 2000; Qian et al. 2016). PCA of the normalized variables was carried out to extract significant principal components and to further reduce the contribution of less significant variables. Then the extracted principal components were subjected to varimax rotation (raw) generating varifactors (Brumelis et al. 2000; Love et al. 2004; Abdul-Wahab et al. 2005). As a result, a small number of factors will usually account for approximately the same amount of information as do the much larger set of original observations. In FA, the basic concept is expressed as in Eq. (6):

$$z_{ji} = a_{f1}f_{1i} + a_{f2}f_{2i} + a_{f3}f_{3i} + \dots + a_{fm}f_{mi} + e_{fi}, \quad (6)$$

where z is the measured value of a variable, a the factor loading, f the factor score, e the residual term accounting for errors or other sources of variation, i the sample number, j the variable number and m the total number of factors.

Results and discussion

Groundwater chemistry

Basic statistics of the respective values for all the physico-chemical parameters in the pre and post-monsoon groundwater samples from the study area and corresponding permissible limits as specified by the Bureau of Indian Standards (2012) are presented in Table 1 and as box plot in Fig. 2a–c. The values of pH in groundwater of study area vary from 6.07 to 8.13 in pre-monsoon and 5.8 to 7.7 in post-monsoon, indicating slightly acidic to alkaline nature. This shows that there is little seasonal fluctuation in pH values in the area that is lower than the permissible limit of 6.5–8.5. The electrical conductivity of groundwater varies widely, ranging from 240 to 4230 $\mu\text{S}/\text{cm}$ in pre-monsoon and 254 to 4483 $\mu\text{S}/\text{cm}$ in post-monsoon. The total dissolved solids values varied between 152 and 2242 mg/L in pre-monsoon and 162 and 2869 mg/L in post-monsoon. The electrical conductivity and total dissolved solids values in all the samples were well above their respective desirable limits of 1400 $\mu\text{S}/\text{cm}$ and 500 mg/L indicating the presence and dissolution of higher salt content.

Water hardness is caused primarily by the presence of cations, such as calcium and magnesium and anions, such as carbonate, bicarbonate, chloride and sulfates in water. Water hardness varied between 48 and 1784 mg/L for the pre-monsoon period and 50 and 1873 mg/L during post-monsoon period thereby exceeding the desirable limit of 300 mg/L in many samples. Among the alkaline earths, the concentration of calcium is in the range of 6–312 mg/L in pre-monsoon and 6–316 mg/L in post-monsoon, while magnesium content ranges between 8–3244 mg/L in pre-monsoon and 8–268 mg/L in post-monsoon seasons, their higher concentrations indicating hardness in groundwater. Bicarbonate is the predominant anion in both pre and post-monsoon seasons, whose concentration varied from 88 to 505 mg/L in pre-monsoon and 92 to 530 mg/L in post-monsoon. Higher concentration of bicarbonate may be attributed to leaching of mineral substances in the soil and atmosphere during natural filtration of water from sewage (Ravikumar et al. 2012).

Chlorides are in the range of 19–607 mg/L and 20–667 mg/L, respectively, during pre-monsoon and post-monsoon, indicating that there is not much difference in chloride concentration between seasons. Presence of chloride in the groundwater of the study area is due to seepage from sewers, septic tanks and industrial effluents. The nitrate concentration in the study area ranges from 2 to 252 mg/L in pre-monsoon and 2 to 262 mg/L in post-monsoon seasons. Majority of the samples among

Table 1 Basic statistics of groundwater quality data of the study region

Parameter	Season	Min	Max	Mean	Skewness	Kurtosis	Standard IS 10500:2012
pH	Pre-monsoon	6.07	8.13	6.94	0.96	1.26	6.5–8.5
	Post-monsoon	5.80	7.70	6.60	0.96	1.26	
EC ($\mu\text{S}/\text{cm}$)	Pre-monsoon	240	4230	1300	1.92	7.82	1400
	Post-monsoon	254	4483	1378	1.92	7.82	
Total Hardness (mg/L)	Pre-monsoon	48	1784	444	2.55	13.08	300
	Post-monsoon	50	1873	466	2.55	13.08	
TDS (mg/L)	Pre-monsoon	152	2422	770	1.83	7.17	500
	Post-monsoon	162	2869	882	1.92	7.82	
F (mg/L)	Pre-monsoon	0.11	1.38	0.38	1.80	4.42	1
	Post-monsoon	0.11	1.40	0.39	1.80	4.42	
NO ₃ (mg/L)	Pre-monsoon	2	252	51	2.08	7.06	45
	Post-monsoon	2	262	53	2.08	7.06	
Cl (mg/L)	Pre-monsoon	19	607	165	1.23	2.56	250
	Post-monsoon	20	667	181	1.23	2.56	
SO ₄ (mg/L)	Pre-monsoon	3	187	82	0.67	−0.01	200
	Post-monsoon	3	201	88	0.67	−0.01	
Fe (mg/L)	Pre-monsoon	0.05	4.30	1.10	1.23	0.29	0.3
	Post-monsoon	0.05	4.60	1.20	1.25	0.32	
K (mg/L)	Pre-monsoon	0.30	16	5.11	0.95	0.59	–
	Post-monsoon	0.40	17	5.62	0.95	0.59	
Na (mg/L)	Pre-monsoon	22	205	89	0.53	−0.09	100
	Post-monsoon	23	223	97	0.53	−0.09	
Mg (mg/L)	Pre-monsoon	8	244	48	3.37	18.75	30
	Post-monsoon	8	268	53	3.37	18.75	
Ca (mg/L)	Pre-monsoon	6	312	108	0.87	3.08	75
	Post-monsoon	6	336	117	0.87	3.08	
HCO ₃ (mg/L)	Pre-monsoon	88	505	325	−0.10	0.03	–
	Post-monsoon	92	530	341	−0.10	0.03	

Bold-faced values represent mean value of parameters exceeding desirable limit

pre-monsoon samples showed nitrate concentration above the permissible limit of 45 mg/L, which can be attributed to contamination from septic tank and sewage effluent as there is no agricultural activity nor application of nitrogenous fertilizers as it is an urban area. Further, the fluoride concentration was found to vary from 0.11 to 1.38 mg/L in pre-monsoon and 0.11 to 1.40 in post-monsoon, which is exceeding the desirable limit of 1 mg/L in the study area. The geology of the study area is predominated by granites/gneisses with intensive presence of pegmatites, which contributes to the occurrence of fluoride in bore wells.

Estimation of water quality index

In the present study, 12 water quality parameters, pH, TDS, Hardness, F, Fe, Na, SO₄, NO₃, Cl, Na, Ca, Mg were considered for computing WQI. It is well known that the more harmful a given pollutant is, the smaller is its permissible value for the standard recommended for drinking water.

So, the “weights” for various water quality parameters are assumed to be inversely proportional to the recommended standards for the corresponding parameters (Pius et al. 2012). Calculated relative weight (W_i) values of each parameter are given in Table 2.

Water quality types were determined on the basis of WQI. The computed WQI values range from 19 to 145 and 24 to 164 for pre-monsoon and post-monsoon, respectively. The WQI range, type of water and calculation of WQI for percentage samples are classified in Table 3. It can be observed that out of 67 groundwater quality data points 24 stations (35%) fall in the “excellent” category, 16 stations (23%) in “good” category, 18 stations (26%) in “poor” category, 7 stations (10%) in “very poor” category and 3 stations (4%) in unfit category for pre-monsoon season. During post-monsoon, 20 stations (30%) fall in the “excellent” category, 16 stations (23%) in “good” category, 16 stations (23%) in “poor” category 9 stations (13%) in “very poor” category and 5 stations (7%) in unfit category

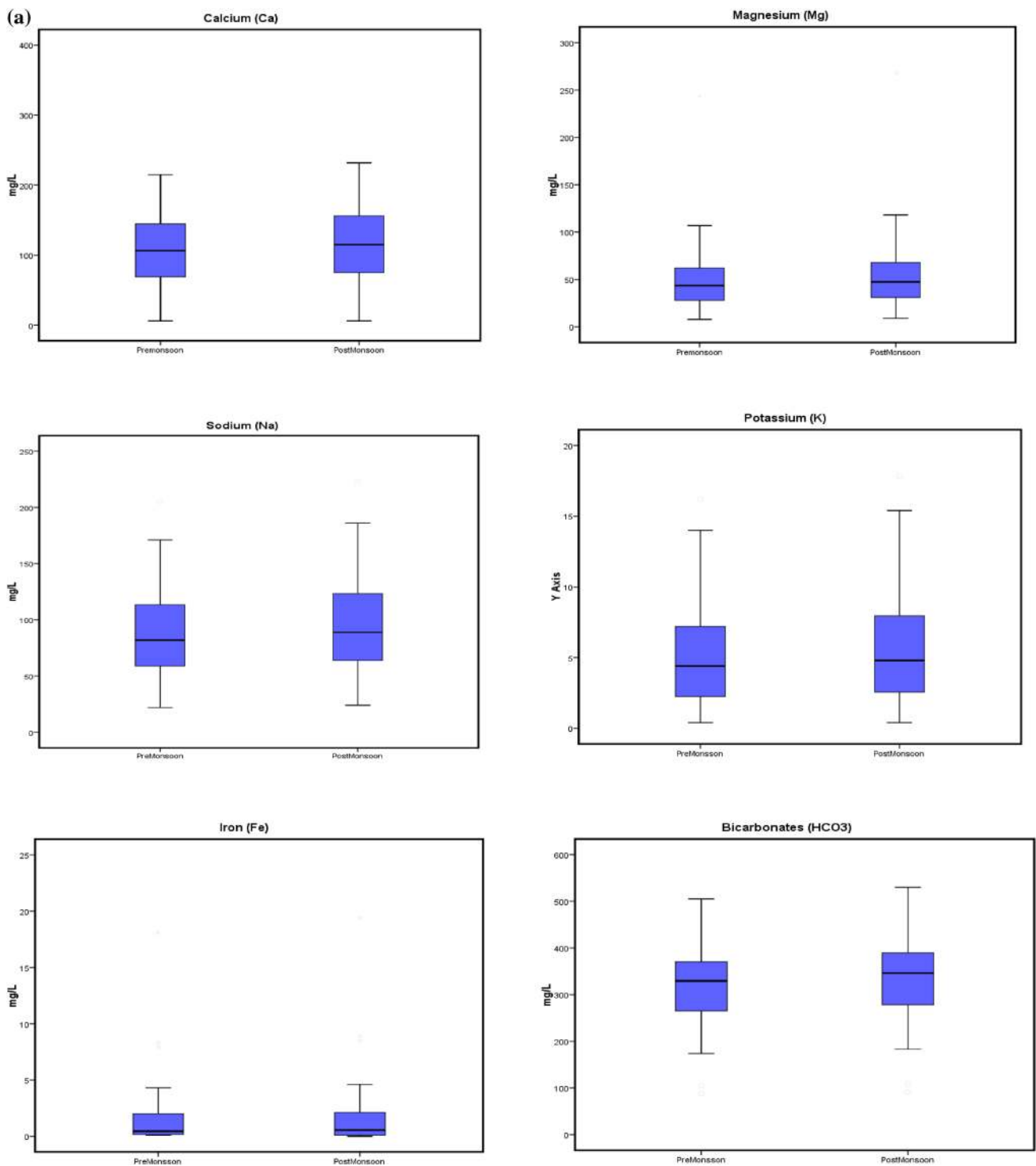


Fig. 2 a Box plots of groundwater quality parameters Ca, Mg, Na, K, Fe, HCO₃. **b** Box plots of groundwater quality parameters Cl, Fe, NO₃, SO₄, TDS, EC. **c** Box plots of groundwater quality parameters TH, pH, F

for post-monsoon dataset. The post-monsoon samples show signs of poor quality in drinking purpose compared to pre-monsoon. Rainfall data published by Indian Meteorology Department (IMD) revealed that Bangalore region

received comparatively higher pre-monsoon rainfall and normal monsoon rainfall while the rainfall in the post-monsoon season was about 31% deficient, when the data were collected. Thus the dilution effect of rainfall recharge

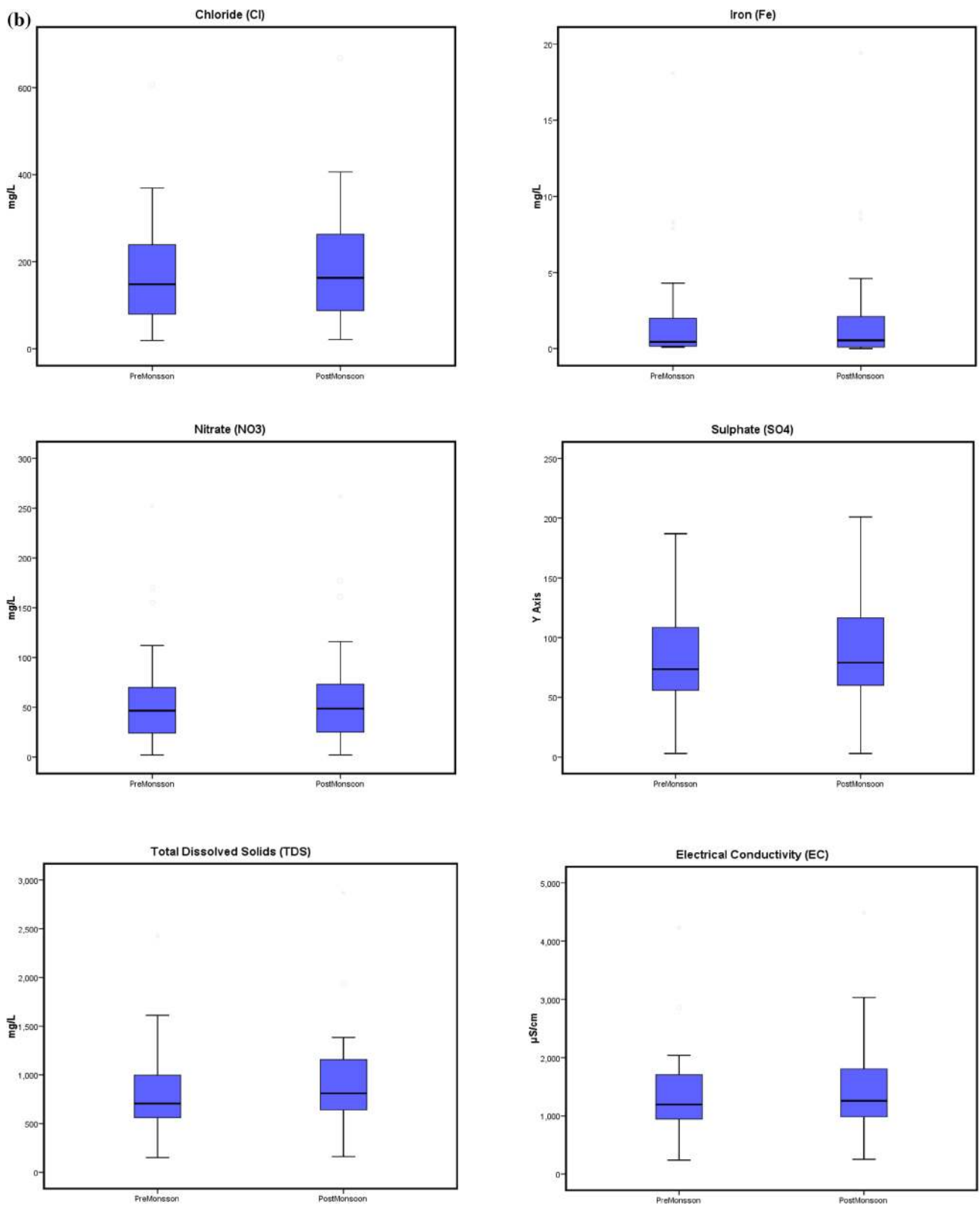


Fig. 2 (continued)

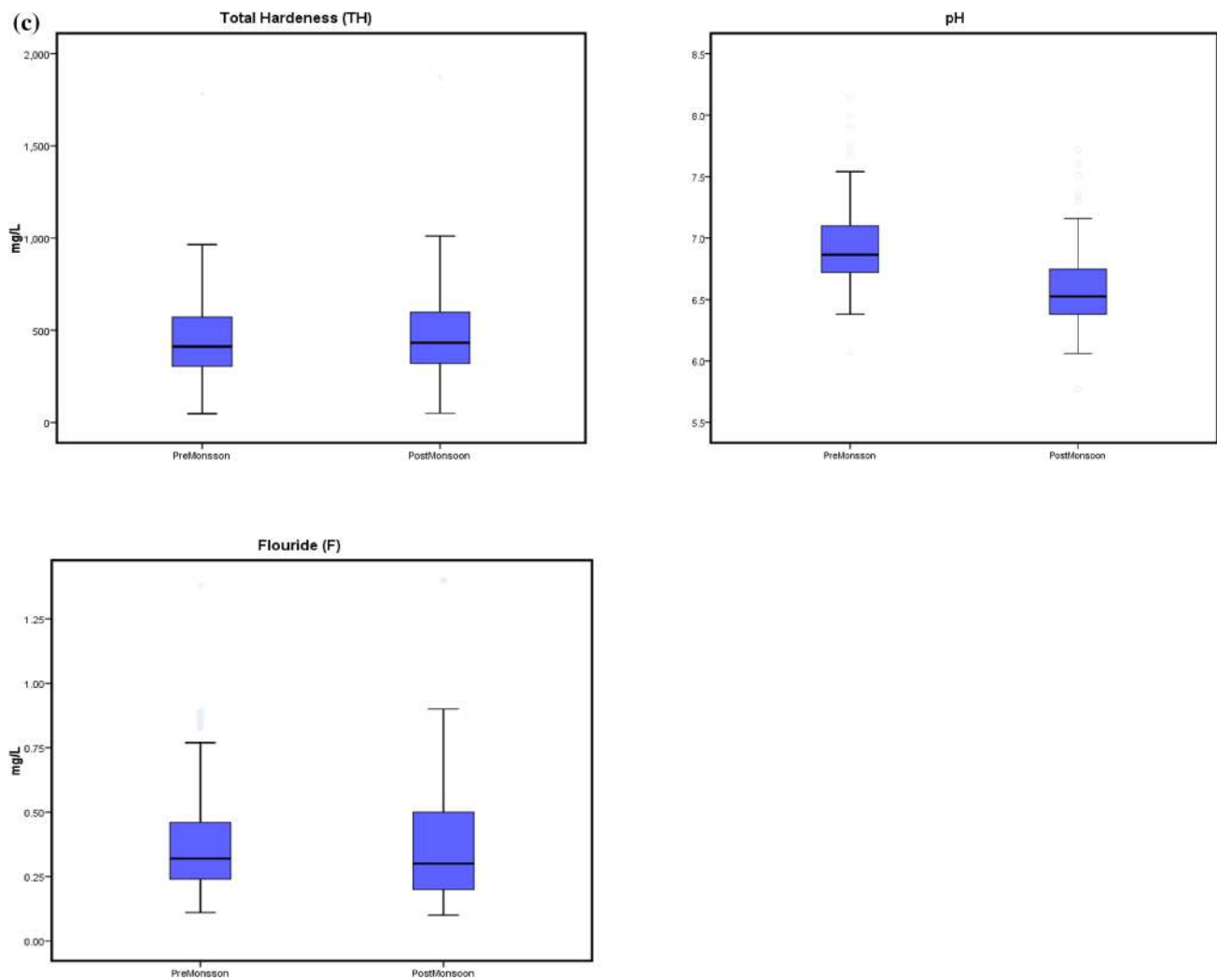


Fig. 2 (continued)

Table 2 Water quality parameters, their standards and unit weights

Parameter	Standard	Weightage
pH	6.5–8.5	0.0509
Hardness	500	0.0008
TDS	500	0.0008
F	1	0.4328
NO ₃	45	0.0096
Cl	250	0.0017
SO ₄	200	0.0021
Fe	1	0.4328
K	10	0.0432
Na	100	0.0043
Mg	30	0.0144
Ca	75	0.0057

Table 3 The WQI range, type of water and percentage wise water quality index area distribution

Water quality index	Description	Percentage of the samples	
		Pre-monsoon	Post-monsoon
0–25	Excellent	35	30
26–50	Good	23	23
51–75	Poor	26	23
76–100	Very poor	10	13
>100	Unfit for drinking (UFD)	4	7

is observed to be higher in the pre-monsoon season. Also, when the rainfall is deficient, there is a risk of higher concentration of surface pollutants getting infiltrated into the groundwater.

Spatial similarity and site grouping

CA for pre-monsoon and post-monsoon data provided a dendrogram grouping the 67 sampling sites into two statistically important clusters (cluster 1 and cluster 2), containing 36 and 31 sites for cluster 1 and 34 and 33 sites for cluster 2, respectively, at $(D_{link}/D_{max}) \times 100 < 25$ as shown in Fig. 3a, b. From the cluster characteristics given in Table 4 it was

observed that, for both pre-monsoon and post-monsoon data, the classification of sampling sites in cluster 1 showed higher level of pollution as compared to cluster 2. While the parameter concentrations in cluster 2 are comparatively lower, some parameters still exceeded the desirable limits. Thus cluster 1 represents high pollution sites and cluster 2 represents low pollution sites. It can be seen that the CA technique is helpful in giving out valid classification of

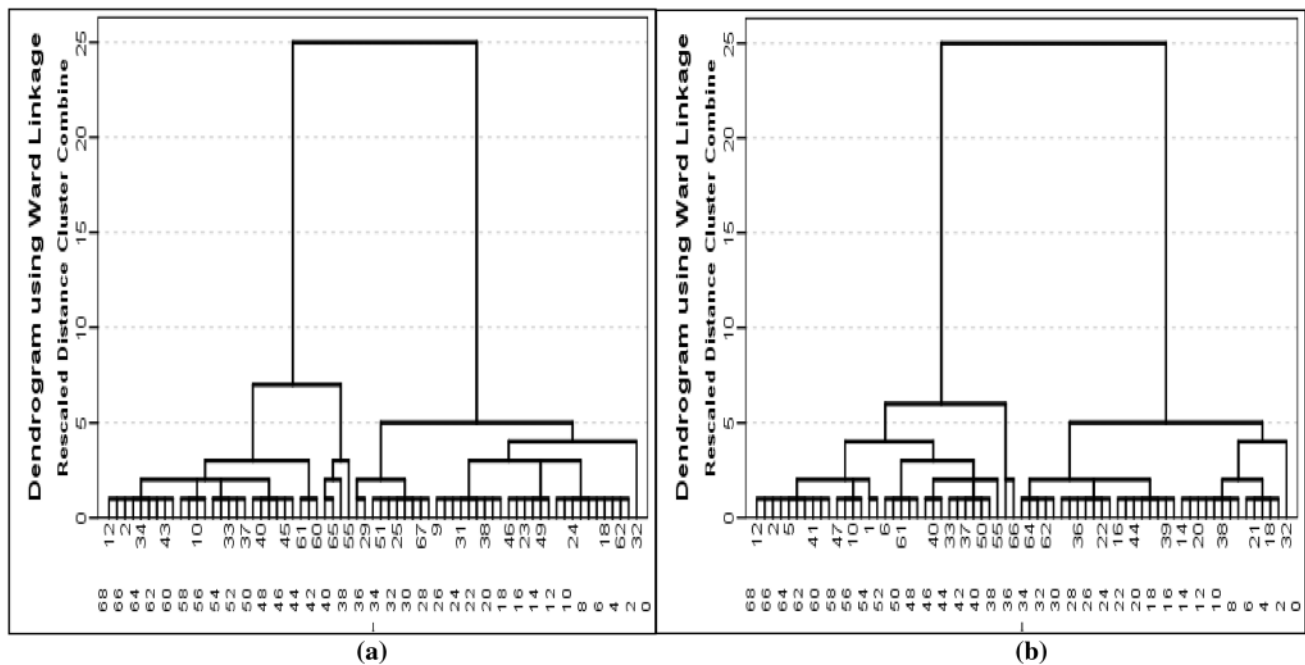


Fig. 3 Dendrogram showing sampling site clusters for a pre-monsoon data and b post-monsoon data

Table 4 Cluster characteristics

Ward's method	Pre-monsoon		Post-monsoon		Pre-monsoon		Post-monsoon	
	Cluster 1 (N = 36)		Cluster 2 (N = 31)		Cluster 1 (N = 34)		Cluster 2 (N = 33)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Ca	139.94	42.23	73.58	29.51	155.62	42.83	79.20	30.88
Mg	66.11	34.66	27.65	11.42	74.61	38.42	31.03	12.42
Na	107.87	38.45	68.42	31.81	116.48	42.86	78.31	36.79
K	6.69	3.46	3.37	2.62	7.47	3.83	3.80	2.90
Fe	0.93	1.24	1.27	1.25	1.06	1.36	1.30	1.36
HCO ₃	371.72	81.24	270.48	70.87	385.88	85.73	295.02	84.45
Cl	232.67	103.14	88.90	47.38	265.39	109.13	97.63	51.30
NO ₃	68.67	48.82	31.35	21.08	75.25	49.61	31.02	22.16
SO ₄	106.44	34.51	54.84	18.39	115.15	37.13	61.57	23.45
TDS	993.67	313.73	517.97	129.01	1166.05	374.28	595.86	156.43
EC	1688.89	550.22	858.87	223.94	1821.95	584.81	931.03	244.42
TH	587.67	239.79	280.29	101.60	634.57	248.07	295.81	103.49
pH	6.92	0.34	6.96	0.46	6.55	0.33	6.63	0.43
F	0.35	0.19	0.43	0.27	0.33	0.16	0.46	0.29

groundwater in the entire region. This will help in designing a future spatial sampling strategy in an optimal manner reducing the number of sampling sites in the monitoring network, which will reduce the cost without affecting the significance of the outcome.

Spatial and temporal variations in groundwater quality

Discriminant analysis was used in order to identify the most important parameters influencing the spatial and temporal variations in groundwater quality. Only 12 parameters were considered for DA excluding TDS and EC to avoid multicollinearity. Discriminant functions (DFs) and classification matrices (CMs) were derived from the standard, forward stepwise and backward stepwise modes of DA. Temporal DA was performed on raw data taking season (pre-monsoon and post-monsoon) as the grouping variable and the measured parameters as the independent variables. The classification functions obtained are given in Table 5 and the classification matrix is given in Table 6.

Standard mode DA constructed DFs using all 12 parameters to give 76% correct assignment of cases in the CM. The forward stepwise mode used only six parameters (K, Fe, HCO₃, NO₃, pH and F) giving 74% correct assignment and the backward stepwise mode gave 72% correct assignment of cases using only one parameter (pH). Thus temporal DA indicated that pH is the most important parameter, which discriminates between the water quality in the pre-monsoon and post-monsoon seasons, followed by K, Fe, HCO₃, NO₃ and F.

Spatial DA was performed on raw data taking cluster (1 and 2) as the grouping variable and the measured parameters as the independent variables. The classification functions obtained are given in Table 7 and the classification matrix

Table 6 Classification matrix for temporal DA

Group	Percent correct	Pre-mon	Post-mon
Standard mode			
Pre	72.06	49	19
Post	79.41	14	54
Total	75.74	63	73
Forward stepwise			
Pre	64.71	44	24
Post	82.35	12	56
Total	73.53	56	80
Backward stepwise			
Pre	67.65	46	22
Post	76.47	16	52
Total	72.06	62	74

is given in Table 8. Standard mode DA constructed DFs using all 12 parameters to give 91% correct assignment of cases in the CM. The forward stepwise mode used seven parameters (Mg, K, HCO₃, Cl, NO₃, SO₄ and F) giving 90% correct assignment and the backward stepwise mode gave 89% correct assignment of cases using only three parameters (Mg, Cl and NO₃). Thus spatial DA identified Mg, Cl and NO₃ as the three most important parameters, which cause the discrimination between the two clusters, followed by K, HCO₃, SO₄ and F.

Data structure determination and source identification

Principal component analysis was applied to standardized datasets separately for the two clusters delineated by CA in order to identify and compare the factors influencing the high and low pollution clusters. Before carrying out PCA,

Table 5 Classification functions for temporal DA

Variables	Standard mode		Forward stepwise mode		Backward stepwise mode	
	Pre-mon	Post-mon	Pre-mon	Post-mon	Pre-mon	Post-mon
Ca	- 0.12	- 0.06				
Mg	- 0.29	- 0.21				
Na	0.04	0.04				
K	- 1.06	- 1.00	- 0.95	- 0.88		
HCO ₃	0.04	0.03	0.05	0.05		
Fe	9.20	8.97	8.83	8.62		
NO ₃	0.23	0.22	0.24	0.23		
Cl	0.04	0.04				
SO ₄	- 0.08	- 0.07				
TH	0.06	0.04				
pH	69.08	65.66	65.01	61.91	46.70	44.36
F	- 10.33	- 9.34	- 14.00	- 12.79		
Constant	- 255.42	- 233.08	- 240.80	- 220.37	- 162.70	- 146.90

Table 7 Classification functions for spatial DA

Variables	Standard mode		Forward stepwise mode		Backward stepwise mode	
	Cluster 1	Cluster 2	Cluster 1	Cluster 2	Cluster 1	Cluster 2
Ca	0.13	0.16				
Mg	0.07	0.15	0.03	0.08	0.04	0.09
Na	0.07	0.06				
K	-0.73	-0.53	-0.005	0.19		
HCO ₃	0.03	0.03	0.03	0.03		
Fe	8.13	7.94				
Cl	0.03	0.04	0.00007	0.01	0.01	0.03
NO ₃	0.17	0.19	0.02	0.04	0.02	0.05
SO ₄	-0.02	0.00	0.04	0.07		
TH	-0.03	-0.04				
pH	53.58	53.40				
F	-6.50	-10.60	5.35	1.19		
Constant	-198.78	-206.95	-9.08	-18.94	-2.27	-10.23

the Kaiser–Meyer–Olkin (KMO) and Bartlett’s sphericity tests were performed on the parameter correlation matrix in order to examine the validity of the PCA (Mustapha et al. 2012). For cluster 1, KMO value of 0.697 > 0.6 and Bartlett’s Sphericity test significance $p < 0.05$ confirmed suitability for PCA. The parameters K, NO₃, Fe and pH were excluded from the analysis due to communalities < 0.5. PCA with varimax rotation was applied to the standardized datasets of the remaining ten parameters. For cluster 2, KMO value of 0.691 > 0.6 and Bartlett’s Sphericity test significance $p < 0.05$ confirmed suitability for PCA. The parameters K, NO₃, Fe and pH were excluded from the analysis due to communalities < 0.5. PCA with varimax rotation was applied to the standardized datasets of the remaining ten parameters.

PCA of the high and low pollution cluster datasets (cluster 1 and cluster 2) yielded three PCs for both the high and low pollution sites with eigenvalues greater than 1,

explaining 85 and 84% of the total variance in the respective groundwater quality data sets. Eigenvalue is important in measuring the significance of the factor, i.e factors with the greater eigenvalues are considered to be most significant. Eigenvalues of 1.0 or greater are considered significant (Kim and Mueller 1987). Same number of VFs were obtained for two clusters by performing FA on the PCs. Variable loadings, explaining variance and corresponding VFs, are presented in Table 9. (Liu et al. 2003) designated the factor loadings as ‘weak’, ‘moderate’ and ‘strong’, with respect to the absolute loading values of 0.50–0.30, 0.75–0.50 and > 0.75, respectively.

For the data set pertaining to cluster 1, VF1, which explained 47.4% of the total variance had strong positive loadings on Ca, Mg, TDS, EC and TH and moderate positive loading on Na. Thus VF1 mainly accounts for calcium and magnesium salts in water resulting in high hardness. Also it can be inferred that the high electrical conductivity and high dissolved solids’ content in the water samples are predominantly contributed by calcium and magnesium and to a lesser extent by sodium. VF2 explaining 20.9% of the total variance had strong positive loadings on Cl, Na and SO₄. Thus VF2 indicates Na–Cl water type and also the presence of sodium sulfate in groundwater. VF3 explaining 17.1% of the total variance had strong positive loadings on HCO₃ and F, moderate positive loading on Na and moderate negative loading on SO₄. The strong positive loading on F and HCO₃ indicates that dissolution of fluoride occurring in groundwater is favorable in alkaline environment.

For the data set representing cluster 2, among the three VFs, VF1, which explained 46.3% of the total variance had strong positive loadings on Ca, Cl, TDS, EC and TH; and moderate positive loadings on Mg and HCO₃. Thus VF1 indicates that the presence of high hardness, electrical conductivity and dissolved solids in the groundwater is mainly

Table 8 Classification matrix for spatial DA

Group	Percent correct	Cluster 1	Cluster 2
Standard mode			
Cluster 1	98.59	70	1
Cluster 2	83.08	11	54
Total	91.18	81	55
Forward stepwise mode			
Cluster 1	97.18	69	2
Cluster 2	81.54	12	53
Total	89.71	81	55
Backward stepwise mode			
Cluster 1	98.59	70	1
Cluster 2	78.46	14	51
Total	88.97	84	52

Table 9 Varimax rotated factor loading on significant PCs of cluster 1 and cluster 2

Parameter	Varimax rotated component (cluster 1)			Varimax rotated component (cluster 2)		
	VF1	VF2	VF3	VF1	VF2	VF3
TH	0.98	0.06	- 0.13	0.91	0.16	- 0.27
EC	0.95	0.26	- 0.02	0.90	0.40	0.06
TDS	0.94	0.29	- 0.02	0.89	0.41	0.09
Mg	0.93	- 0.15	- 0.16	0.37	0.46	- 0.62
Ca	0.90	0.25	- 0.003	0.92	0.03	- 0.17
Cl	0.21	0.81	- 0.11	0.86	- 0.19	0.18
Na	0.32	0.78	0.35	0.15	0.21	0.90
SO ₄	- 0.13	0.75	- 0.47	0.25	0.80	- 0.14
HCO ₃	0.01	- 0.01	0.86	0.54	0.66	0.001
F	- 0.29	- 0.09	0.73	- 0.22	0.75	0.24
Eigen value	4.73	2.09	1.71	4.62	2.33	1.45
% Total variance	47.36	20.92	17.09	46.26	23.29	14.50
Cumulative % variance	47.36	68.29	85.39	46.26	69.56	84.06

Bold-faced values represent strong loadings

due to the presence of calcium chlorides and bicarbonates and to a lesser extent due to magnesium chlorides and bicarbonates. VF2 explaining 23.3% of the total variance had strong positive loading on F, HCO₃ and SO₄; and moderate positive loadings on EC, TDS and Mg. Thus PC2 indicates higher dissolution of Fluoride in alkaline environment. Also, the presence of sodium and magnesium sulfates is indicated. PC3 explaining 14.5% of the total variance had strong positive loading on Na and strong negative loading on Mg. Thus PC3 indicates salinity due to sodium ion possibly from sodium-containing rock formations. Also the low loading on HCO₃ along with the negative loading on Mg may be due to removal of Mg from the groundwater in the form of Magnesium bicarbonate precipitate.

Conclusions

- The present study demonstrated the importance of multivariate statistical analysis in groundwater studies. Basic statistics showed that most of the parameters were found to exceed the specified desirable limits while few parameters exceeded the permissible limits as well. The WQI calculated showed that the number of samples rated as poor, very poor and unfit constitute about 50% of the total samples thereby pointing out to the fact that the groundwater of these needs some degree of treatment before consumption, and it also needs to be protected from the perils of contamination. The results of WQI agree with the fact that many parameters exceeded the desirable limits as observed from basic statistical analysis.

- Different multivariate statistical techniques were applied to evaluate spatial and temporal variations in groundwater quality of Bengaluru city. Hierarchical cluster analysis was useful in classifying the 67 sampling sites into two main clusters as high- and low-pollution areas. This helps in the identification of problematic zones in the area where remedial actions need to be focused. Also, grouping the areas having similar groundwater condition may be used to determine the number of sampling sites required for regular monitoring of groundwater quality.
- DA was useful in identifying a few indicator parameters responsible for significant variations (spatial and temporal) in groundwater quality of the study area. pH was identified as the most important parameter, which discriminates between the groundwater quality in the pre-monsoon and post-monsoon seasons and accounts for 72% seasonal assignation of cases. Mg, Cl and NO₃ were identified as the three most important parameters discriminating between the two clusters and accounting for 89% spatial assignation of cases.
- Grouping of the measured parameters to identify the underlying factors or processes influencing the groundwater quality in the study region was achieved through PCA. Three principal components (PCs) each were identified for the two clusters. Dissolution of hardness causing Ca and Mg from bed rock and anthropogenic sources, fluoride dissolution from bedrock in alkaline environment and salinity from natural and anthropogenic sources were identified to be the main factors influencing the ground water quality in both the clusters.

- Thus, the usefulness of multivariate statistical techniques for analysis and interpretation of complex data sets was illustrated in this study for groundwater quality assessment. The grouping information extracted from cluster analysis can be used to design optimal sampling strategy, which; could reduce the number of sampling stations and associated costs. DA provided with data reduction, by identifying the most important parameters, which; needs to be monitored in order to study the spatial and temporal variations in water quality. While PCA served as a means to identify those parameters, which; have greatest contribution to temporal variation in the groundwater quality and suggested possible sets of pollution sources. Overall the multivariate statistical techniques helped in understanding the temporal/spatial variations in groundwater quality, identification of pollution sources/factors as an effort towards a more effective groundwater quality management.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abdul-Wahab SA, Bakheit CS, Al-Alawi SM (2005) Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environ Model Softw* 20(10):1263–1271
- Adams MJ (1998) The principles of multivariate data analysis. In: Ashurst PR, Dennis MJ (eds) *Analytical methods of food authentication*. Blackie Academic and Professional, London, p 350
- APHA (2005) *Standard methods for the examination of water and wastewater*, 20th edn. American Public Health Association, Washington, DC
- Aris AZ, Praveena SM, Abdullah MH, Radojevic M (2012) Statistical approaches and hydrochemical modeling of groundwater system in a small tropical island. *J Hydroinf* 14:206–220
- Avannavar SM, Shrihari S (2008) Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India. *Environ Monit Assess* 143:279–290
- BIS (2003) *Drinking water specifications*, Bureau of Indian Standards, 1991, IS:10500 (revised 2003)
- Boateng TK, Opoku F, Acquah SO, Akoto O (2016) Groundwater quality assessment using statistical approach and water quality index in Ejisu-Juaben municipality, Ghana. *Environ Earth Sci* 75(6):1–14
- Brümelis G, Lapiņa L, Nikodemus O, Tabors G (2000) Use of an artificial model of monitoring data to aid interpretation of principal component analysis. *Environ Model Softw* 15(8):755–763
- Central Ground Water Board. (2012) *Ground water information booklet-Dakshina Kannada District Karnataka*, Ministry of water resources, Government of India
- Cloutier V, Lefebvre R, Therrien R, Savard MM (2008) Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J Hydrol* 353(3):294–313
- Dawoud MA, Raouf ARA (2009) Groundwater exploration and assessment in rural communities of Yobe State, Northern Nigeria. *Water Resour Manag* 23(3):581–601
- Department of Mines and Geology (2003) *Status of ground water quality in Bangalore and its environs*. District groundwater brochure, Bangalore
- Department of Mines and Geology (2011) *Groundwater hydrology and groundwater quality in and around Bangalore city*. District groundwater brochure, Bangalore
- Forina M, Armano C, Raggio V (2002) Clustering with dendrograms on interpretation variables. *Anal Chim Acta* 454:13–19
- Güler C, Thyne GD, McCray JE, Turner KA (2002) Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol J* 10(4):455–474
- Hassen I, Hamzaoui-Azaza F, Bouhlila R (2016) Application of multivariate statistical analysis and hydrochemical and isotopic investigations for evaluation of groundwater quality and its suitability for drinking and agriculture purposes: case of Oum Ali-Thelepte aquifer, central Tunisia. *Environ Monit Assess* 188(3):1–20
- Helena B, Pardo R, Vega M, Barrado E, Fernández JM, Fernández L (2000) Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga river, Spain) by principal component analysis. *Water Res* 34:807–816
- Horton RK (1965) An index number system for rating water quality. *J Water Pollut Control Fed* 37:300–305
- Hossain MA, Ali NM, Islam MS, Hossain HZ (2015) Spatial distribution and source apportionment of heavy metals in soils of Gebeng industrial city, Malaysia. *Environ Earth Sci* 73(1):115–126
- Jammel A, Hussain AZ (2003) Impact of sewage on the quality of Uyakandan channel water of River Cauvery at Tiruchirapalli. *Indian J Environ Prot* 23(6):660–662
- Jat MK, Garg PK, Khare D (2008) Monitoring and modelling of urban sprawl using remote sensing and GIS techniques. *Int J Appl Earth Obs Geoinf* 10(1):26–43
- Jerry AN (1986) *Basic environmental technology (water supply, waste disposal and pollution control)*. Wiley, New York
- Johnson RA, Wichern DW (1992) *Applied multivariate statistical analysis*, 3rd edn. Prentice-Hall, New Jersey
- Juahir H, Zain SM, Yusoff MK, Hanidza TT, Armi AM, Toriman ME, Mokhtar M (2011) Spatial water quality assessment of Langat River basin (Malaysia) using environmetric techniques. *Environ Monit Assess* 173(1–4):625–641
- Kazi TG, Arain MB, Jamali MK, Jalbani N, Afridi HI, Sarfraz RA, Baig JA, Shah AQ (2009) Assessment of water quality of polluted lake using multivariate statistical techniques a case study. *Ecotoxicol Environ Saf* 72:301–309
- Kim JO, Mueller CW (1987) *Introduction to factor analysis: what it is and how to do it*. Quantitative applications in the social sciences series
- Kumar M, Ramanathan AL, Tripathi R, Farswan S, Kumar D, Bhattacharya P (2017) A study of trace element contamination using multivariate statistical techniques and health risk assessment in groundwater of Chhaprola Industrial Area, Gautam Buddha Nagar, Uttar Pradesh, India. *Chemosphere* 166:135–145
- Lee JY, Cheon JY, Lee KK, Lee SY, Lee MH (2001) Statistical evaluation of geochemical parameter distribution in a ground water system contaminated with petroleum hydrocarbons. *J Environ Qual* 30(5):1548–1563
- Liu CW, Lin KH, Kuo YM (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci Total Environ* 313(1):77–89
- Love D, Hallbauer D, Amos A, Hranova R (2004) Factor analysis as a tool in groundwater quality management: two southern African case studies. *Phys Chem Earth* 29(15–18):1135–1143

- Massart DL, Kaufman L (1983) The interpretation of analytical data by the use of cluster analysis. Wiley, New York
- McKenna JE (2003) An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environ Model Softw* 18(3):205–220
- Mustapha A, Aris AZ (2012) Multivariate statistical analysis and environmental modeling of heavy metals pollution by industries. *Pol J Environ Stud* 21:1359–1367
- Mustapha A, Aris AZ, Juahir H, Ramli MF (2012) Surface water quality contamination source apportionment and physicochemical characterization at the upper section of the Jakara basin, Nigeria. *Arab J Geosci*
- Noshadi M, Ghafourian A (2016) Groundwater quality analysis using multivariate statistical techniques (case study: Fars province, Iran). *Environ Monit Assess* 188(7):1–13
- Otto M (1998) Multivariate methods. In: Kellner R, Mermet JM, Otto M, Widmer HM (eds) *Analytical chemistry*. Wiley, Weinheim
- Pius A, Jerome C, Sharma N (2012) Evaluation of groundwater quality in and around Peenya industrial area of Bangalore, South India using GIS techniques. *Environ Monit Assess* 184(7):4067–4077
- Pradhan SK, Patnaik D, Rout SP (2001) Water quality index for the ground water around a phosphatic fertilizer plant. *Indian J Environ Prot* 21(4):355–358
- Qian J, Wang L, Ma L, Lu Y, Zhao W, Zhang Y (2016) Multivariate statistical analysis of water chemistry in evaluating groundwater geochemical evolution and aquifer connectivity near a large coal mine, Anhui, China. *Environ Earth Sci* 75(9):1–10
- Ravikumar P, Somashekar RK (2012) Assessment and modelling of groundwater quality data and evaluation of their corrosiveness and scaling potential using environmetric methods in Bangalore south taluk, Karnataka state, India. *Water Resour* 39(4):446–473
- Ravikumar P, Somashekar RK (2017) Principal component analysis and hydrochemical facies characterization to evaluate groundwater quality in Varahi river basin, Karnataka state, India. *Appl Water Sci* 7(2):745–755
- Reghunath R, Murthy TS, Raghavan BR (2002) The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Res* 36(10):2437–2442
- Sadat-Noori SM, Ebrahimi K, Liaghat AM (2014) Groundwater quality assessment using the water quality index and GIS in Saveh-Nobaran aquifer, Iran. *Environ Earth Sci* 71(9):3827–3843
- Salman SR, Ruka'h YA (1999) Multivariate and principal component statistical analysis of contamination in urban and agricultural soils from north Jordan. *Environ Geol* 38(3):265–270
- Samson S, Elangovan K (2017) Multivariate statistical analysis to assess groundwater quality in Namakkal district, Tamil Nadu, India
- Sarbu C, Pop HF (2005) Principal component analysis versus fuzzy principal component analysis: a case study: the quality of Danube water (1985–1996). *Talanta* 65(5):1215–1220
- Shrestha S, Kazama F (2007) Assessment of surface water quality using multivariate statistical techniques: a case study of the Fuji river basin, Japan. *Environ Model Softw* 22:464–475
- Simeonov V, Stratis JA, Samara C, Zachariadis G, Voutsas D, Anthemidis A (2003) Assessment of the surface water quality in northern Greece. *Water Res* 37(4119):4124
- Simeonov V, Simeonova P, Tzimou-Tsitouridou R (2004) Chemometric quality assessment of surface waters: two case studies. *Chem Inz Ekol* 11(6):449–469
- Singh KP, Malik A, Mohan D, Sinha S (2004) Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti river (India): a case study. *Water Res* 38:3980–3992
- Singh KP, Malik A, Sinha S (2005) Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques: a case study. *Anal Chim Acta* 538:355–374
- Taoufik G, Khouni I, Ghrabi A (2017) Assessment of physico-chemical and microbiological surface water quality using multivariate statistical techniques: a case study of the Wadi El-Bey river, Tunisia. *Arab J Geosci* 10(7):181
- Tirkey P, Bhattacharya T, Chakraborty S, Baraik S (2017) Assessment of groundwater quality and associated health risks: a case study of Ranchi city, Jharkhand, India. *Groundwater for Sustainable Development*
- Tiwari TN, Mishra M (1985) A preliminary assignment of water quality index of major Indian rivers. *Ind J Environ Prot* 5(4):276–279
- Tziritis E, Skordas K, Kelepertsis A (2016) The use of hydrogeochemical analyses and multivariate statistics for the characterization of groundwater resources in a complex aquifer system. A case study in Amyros river basin, Thessaly, central Greece. *Environ Earth Sci* 75(4):1–11
- Vega M, Pardo R, Barrado E, Deban L (1998) Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res* 32:3581–3592
- Wang J, Da L, Song K, Li B (2008) Temporal variations of surface water quality in urban, suburban and rural areas during rapid urbanization in Shanghai. *China* 152:387–393
- Willet P (1987) *Similarity and clustering in chemical information systems*. Research Studies Press, Wiley, New York
- Wunderlin DA, Diaz MP, Ame MV, Pesce SF, Hued AC, Bistoni MA (2001) Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality, a case study: Suquia river basin (Cordoba, Argentina). *Water Res* 35:2881–2894
- Yidana SM, Yiran GB, Sakyi PA, Nude PM, Banoeng-Yakubo B (2011) Groundwater evolution in the Voltaian basin, Ghana—an application of multivariate statistical analyses to hydrochemical data. *Nat Sci* 3(10):837
- Zhang Y, Guo F, Meng W, Wang XQ (2009) Water quality assessment and source identification of Daliao river basin using multivariate statistical methods. *Environ Monit Assess* 152:105–121

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.