

Water quality analysis of the Rapur area, Andhra Pradesh, South India using multivariate techniques

A. Nagaraju¹ · Y. Sreedhar¹ · A. Thejaswi² · Mohammad Hossein Sayadi³

Received: 25 June 2015 / Accepted: 18 November 2016 / Published online: 28 November 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract The groundwater samples from Rapur area were collected from different sites to evaluate the major ion chemistry. The large number of data can lead to difficulties in the integration, interpretation, and representation of the results. Two multivariate statistical methods, hierarchical cluster analysis (HCA) and factor analysis (FA), were applied to evaluate their usefulness to classify and identify geochemical processes controlling groundwater geochemistry. Four statistically significant clusters were obtained from 30 sampling stations. This has resulted two important clusters viz., cluster 1 (pH, Si, CO₃, Mg, SO₄, Ca, K, HCO₃, alkalinity, Na, Na + K, Cl, and hardness) and cluster 2 (EC and TDS) which are released to the study area from different sources. The application of different multivariate statistical techniques, such as principal component analysis (PCA), assists in the interpretation of complex data matrices for a better understanding of water quality of a study area. From PCA, it is clear that the first factor (factor 1), accounted for 36.2% of the total variance, was high positive loading in EC, Mg, Cl, TDS, and hardness. Based on the PCA scores, four significant cluster groups of sampling locations were detected on the basis of similarity of their water quality.

Keywords Groundwater · Multivariate statistical analysis · Rapur

Introduction

Water quality is controlled by many factors including climate, soil topography, and water rock interaction (Farnham et al. 2003; Love et al. 2004; Li et al. 2016a, b). These factors lead to a variation in hydrogeochemical process of groundwater using statistical techniques. Detailed hydrochemical research is needed to evaluate the different processes and mechanisms involved in polluting water (Helena et al. 1999). Earlier studies have been focused in hydrogeochemical process of groundwater using statistical analysis (Ashley and Llyod 1978; Reghunath et al. 2002; Liu et al 2003; Monjerezi et al. 2008; Nagaraju et al. 2014a, b; Wu et al. 2014; Li et al. 2016a, b). These techniques constitute a useful tool for groundwater quality characterization for identification of the regional groundwater flow pattern and investigation of groundwater contamination like trace elements (Voudouris et al. 1997; Cloutier et al. 2008; Belkhiri et al. 2010; Kumar et al. 2013). The multivariate statistical analysis methods have the advantage of explaining complex water quality monitoring data. Earlier workers have been successfully applied to a number of hydrogeochemical studies (Singh et al. 2005; Kowalkowskia et al. 2006; Boyacioglu 2008; Wu et al. 2014; Nagaraju et al. 2016a, b). Further, these studies have shown that multivariate statistical analysis can help to interpret the complex datasets, and it is useful in verifying temporal and spatial variations caused by natural and anthropogenic factors. Surface water, groundwater quality assessment, and environmental research employing multi-component techniques are well described in the literature

✉ A. Nagaraju
arveti@yahoo.com

¹ Department of Geology, Sri Vekateswara University, Tirupati, Andhra Pradesh 517 502, India

² Department of Environmental Sciences, Kakatiya University, Warangal, Telangana 506 009, India

³ Department of Environment, University of Birjand, Birjand, Iran

(Prau 2005). Multivariate statistical approaches allow deriving hidden information from the dataset about the possible influences of the environment on water quality (Spanos et al. 2003).

This multivariate method was used here to obtain information about the most relevant characteristics of the physico-chemical variables with a minimal loss of original data (De Bartolomeo et al. 2004; Altun et al. 2008; Kazi et al. 2009), to create an entirely new set of factors much smaller in number when compared to the original dataset of variables focused on reducing the contribution of the less significant variables to simplify even more the data structure coming from the principal component analysis (İşçen et al. 2008).

Factor analysis attempts to explain the correlations between the observations in terms of the underlying factors, which are not directly observable (Yu et al. 2003). There are three stages in factor analysis (Gupta et al. 2005): for all the variables a correlation matrix is generated, factors are extracted from the correlation matrix based on the correlation coefficients of the variables, and to maximize the relationship between some of the factors and variables, the factors are rotated. Cluster Analysis was used to explore the similarities between water samples (Kotti et al. 2005) and grouping the sites according to the similarity of contaminants. Hierarchical agglomerative CA was performed on the normalized dataset using squared Euclidean distances as a measure of similarity. The CA technique is a classification procedure that involves measuring either the distance or the similarity between the objects to be clustered.

Therefore, the main objectives of the present study are the following: (i) to assess the status of water quality in relation to physico-chemical parameters; (ii) to assess the correlations between the different water quality parameters (iii) to find out the similarities and dissimilarities among the different sampling sites, and (iv) to ascertain the influence of the pollution sources on the water quality variables.

Area of study

This study area is about 40 sq.km and is located in the Rapur Taluk of Nellore District, Andhra Pradesh. It forms part of the Survey of India toposheet No. 57 N/11 and lies between 14°16' 30'' forms and 14°19'N latitude and 79°38' and 79°41' east longitude (Fig. 1). The area is accessible by the Nellore Rapur road which passes through North–Western portion of the area. The area is characterized by hot and sub-humid climate and is in the tropical region. In general, the climate is good and is not subjected to sudden variations in temperature. The maximum, average, and minimum temperatures are 44, 31, and 18 °C respectively.

They receive rainfall during the months of July, August, September, and October with maximum precipitation in October. The heavy rainfall is limited to a few days in a year due to depressions in Bay of Bengal which leads to flash floods of high discharge. The annual normal rainfall of this area is about 1084 mm. The dry climate, the atmospheric dust, and low intensity of precipitation affect the quality of precipitation water.

Hydrogeological setting

This area lies in the semi-arid region of Andhra Pradesh and is susceptible to various threats such as growing urban areas as well as developing agricultural areas. This area is underlain by variety of geological formations comprising from the oldest archaeans to recent alluvium. Hydrogeologically, these formations are classified as consolidated (hard), semi-consolidated (soft), and unconsolidated (soft) formations. The consolidated formations include mainly migmatized high grade metamorphics (essentially garnetiferous amphibolites and pelitic schist), low grade metamorphics (essentially amphibolites and pelitic schists) of Nellore schist belt, and granitic gneiss and Cuddapahs (quartzites and shales) of Pre-cambrian period.

Ground water occurs in almost all geological formations, and its potential depends upon the nature of geological formations, geographical set up, incidence of rainfall, recharge, and other hydrogeological characters of the aquifer. Among the consolidated formations, gneisses are relatively good aquifers. Schistose formations also form potential aquifers when the wells tapping the contact zones with intrusives. Quartzites and shales of Cuddapah group are of little significance from the ground water point of view as they are restricted to the hilly terrain in the western margin of the district. In the consolidated formations, ground water occurs under unconfined to semiconfined conditions. Ground water is developed in these formations by dug wells, dug cum bore wells, and bore wells tapping weathered and fractured zones. The yield of the dug wells is in the range of 15–35 m³/day and reduces considerably during peak summer periods. The occurrence of fractures in these formations is limited to 40–60 m bgl and occasionally extends down to 70–80 m bgl. The bore wells in these formations generally tap the weathered and fractured zones.

Materials and methods

Groundwater samples were collected from 30 locations from Rapur area during April 2014 (Fig. 1). The collected water samples were transferred into precleaned polythene

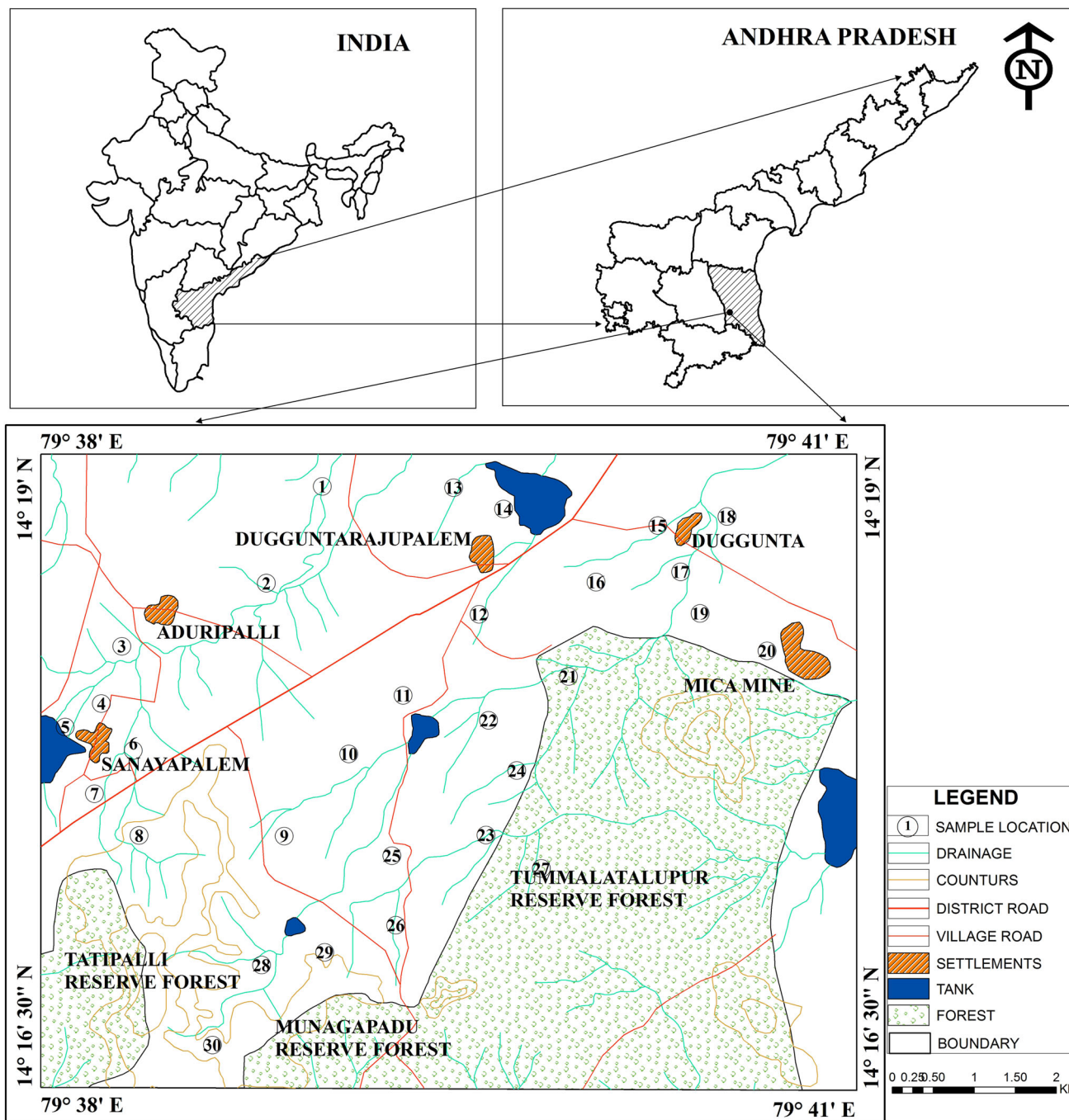


Fig. 1 Map of the study area depicting sample locations

container for analysis of chemical characters. Samples were analyzed in the laboratory for the physico-chemical attributes like pH, electrical conductivity (EC), total hardness (TH), total dissolved solids (TDS), dissolved silica, and major ions (Ca^{2+} , Mg^{2+} , Na^+ , Cl^- , CO_3^{2-} , HCO_3^- , and SO_4^{2-}). All parameters were analyzed by following the standard methods (APHA 2005). The pH and conductivity were measured using Systronics micro pH meter model 361 and Deluxe conductivity meter model 601. Total hardness

(TH), Ca^{2+} , Mg^{2+} , Cl^- , CO_3^{2-} , and HCO_3^- were determined by titration. Na^+ and K^+ were measured by Flame photometry, SO_4^{2-} by Lovibond spectrophotometer.

Statistical analysis

The descriptive statistics (minimum, maximum, mean, standard deviation, standard error, skewness, and kurtosis) were calculated for the physico-chemical variables in water

samples. The Pearson correlation coefficient was obtained to describe the degree of association between two variables under study. Factor analysis was employed to establish the possible relationships between the physico-chemical variables of the sites sampled. Cluster analysis and principal component analysis (PCA) were applied to determine water quality. SPSS 16.0 version statistical program was used for descriptive statistics and multivariate data analysis.

Results and discussion

The descriptive statistics concerning quality of water from 30 stations are presented (Table 1). The range, mean, the standard error, and the standard deviation were shown for all studied parameters as well as other statistical parameters (skewness and kurtosis). From Table 1, it is clear that the 15 analyzed parameter values and standard deviation value range varies considerably. The analyzed parameter values and standard deviation value range vary considerably (Table 1; Fig. 2). From this, it is clear that there is a wide variation of studied parameters in the water. Skewness values of Ca, K, CO₃, and hardness are bigger than the unit which means these parameters positively skewed toward lower concentrations, as can also be confirmed by the fact that their median concentrations are lower than their mean concentrations. Generally, the skewness and kurtosis values should be in the range of ± 2 , otherwise it is considered as extreme (Reimann et al. 2008). In the studied area, the K (skewness 2.539 and kurtosis 5.431), CO₃ (skewness 2.062

and kurtosis 4.605), and hardness (skewness 1.645 and kurtosis 2.838) are considered to be extreme. This is due to the presence of outliers, which correspond to samples with elevated values of these concentrations.

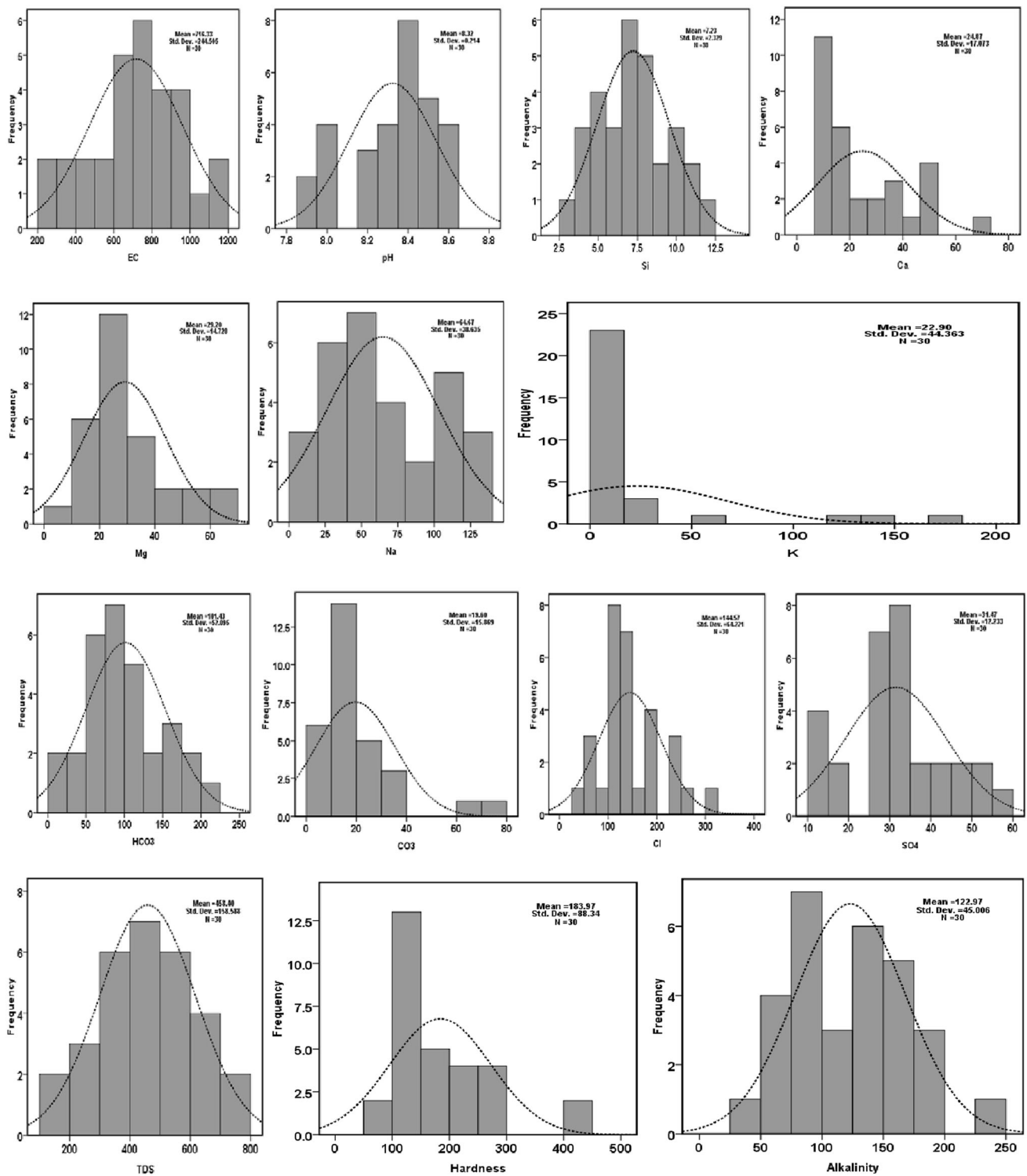
Further, the Pearson's correlation coefficients for physico-chemical parameters of water in the study area are presented in Table 2. A significantly positive correlation at $P < 0.01$ was found between the elemental pairs EC–Mg (0.592), EC–Na + K (0.733), EC–K (0.548), EC–HCO₃ (0.542), EC–Cl (0.810), EC–TDS (0.935), EC–hardness (0.588), and EC–alkalinity (0.542). The wide variety of EC concentrations, reflecting dissolved solutes, is related to lithology, land use, and human activity in the Rapur area (Rezaei and Sayadi 2015). It is obvious that pH and Si have relatively poor correlations with the other studied parameters except Si-alkalinity (0.489). Ca is also significantly positively correlated with hardness (0.552) (Wu et al. 2014) and negatively correlated with Na (–0.551) at $P < 0.01$ (Li et al. 2013a). Mg is also significantly positively correlated with HCO₃, Cl, TDS, hardness, and alkalinity at $P < 0.01$. Na has relatively strong correlations ($P < 0.01$) with Na + K (0.572), Cl (0.546). A high positive correlation between Cl and Na concentrations has been reported (Giridharan et al. 2009; Li et al. 2013b, 2014).

Hierarchical cluster analysis (HCA)

The HCA is a data classification technique. There are different clustering techniques, but the hierarchical clustering is the one most widely applied in Earth sciences

Table 1 Descriptive statistics of the water quality parameters

	Minimum	Maximum	Mean	Std. error	Std. deviation	Skewness	Kurtosis
EC	230.00	1190.00	716.33	44.64	244.50	–0.203	–0.450
pH	7.90	8.60	8.32	0.039	0.21	–0.662	–0.585
Si	3.00	12.00	7.23	0.42	2.32	0.188	–0.642
Ca	7.00	72.00	24.86	3.11	17.07	1.036	0.376
Mg	7.00	64.00	29.20	2.68	14.72	0.849	0.023
Na	8.00	138.00	64.66	7.05	38.63	0.289	–1.095
K	2.00	168.00	22.90	8.09	44.36	2.539	5.431
Na + K	12.00	210.00	87.56	9.70	53.13	0.553	–0.227
HCO ₃	10.00	209.00	101.43	9.51	52.09	0.405	–0.390
CO ₃	2.00	72.00	19.60	2.89	15.86	2.062	4.605
Cl	39.00	306.00	144.57	11.72	64.22	0.645	0.178
SO ₄	10.00	57.00	31.46	2.23	12.23	0.225	–0.294
TDS	147.00	762.00	458.40	28.95	158.58	–0.164	–0.533
Hardness	71.00	446.00	183.97	16.12	88.34	1.645	2.838
Alkalinity	44.00	230.00	122.97	8.21	45.01	0.267	–0.509



(Davis 1986), and often used in the classification of hydrogeochemical data (Steinhorst and Williams 1985; Schot and van der Wal 1992; Ribeiro and Macedo 1995; Güler et al. 2002). Cluster analysis was carried out to investigate the relationship among water quality

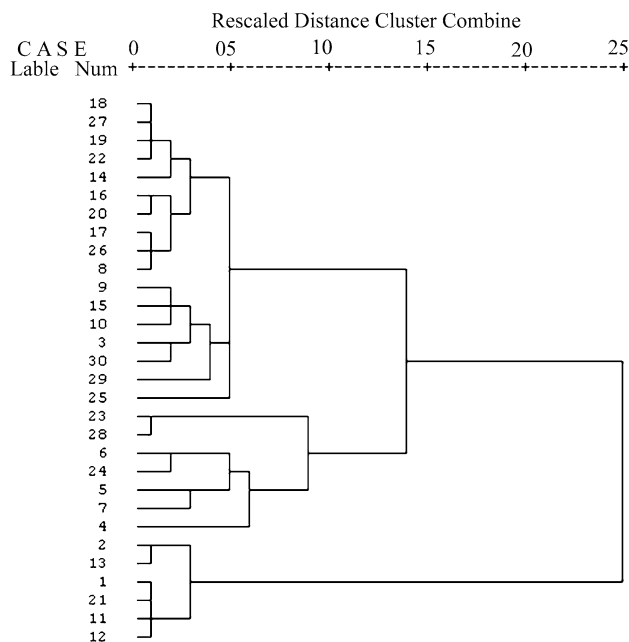
parameters which play significant role in groundwater studies. Cluster analysis is also useful to handle large water quality datasets and identify the dominant mechanisms and factors which control the groundwater chemistry (Kouping et al. 2006). In the present study, the cluster analysis has

Table 2 Pearson correlation coefficient of physico-chemical parameters of water

	EC	pH	Si	Ca	Mg	Na	K	Na + K	HCO ₃	CO ₃	Cl	SO ₄	TDS	Hardness	Alkalinity
EC	1														
pH	-0.063	1													
Si	0.163	0.175	1												
Ca	0.130	0.101	-0.247	1											
Mg	0.592**	-0.040	0.145	0.264	1										
Na	0.378*	0.087	0.294	-0.551**	0.036	1									
K	0.548**	-0.240	-0.085	0.111	0.119	-0.186	1								
Na + K	0.733**	-0.137	0.143	-0.308	0.125	0.572**	0.700**	1							
HCO ₃	0.542**	0.109	0.000	0.300	0.524**	-0.215	0.554**	0.306	1						
CO ₃	0.185	0.140	0.280	-0.102	0.031	0.389*	0.133	0.393*	0.104	1					
Cl	0.810**	-0.142	0.108	0.060	0.631**	0.546**	0.270	0.622**	0.161	0.049	1				
SO ₄	0.223	0.011	0.080	0.241	0.104	0.136	0.104	0.186	-0.155	0.187	0.250	1			
TDS	0.935**	-0.161	0.187	0.100	0.665**	0.364*	0.545**	0.720**	0.535**	0.185	0.840**	0.181	1		
Hardness	0.588**	-0.023	0.099	0.552**	0.841**	-0.187	0.127	-0.030	0.489**	-0.186	0.535**	0.125	0.604**	1	
Alkalinity	0.542**	0.141	0.489**	0.098	0.524**	0.122	0.335	0.369*	0.614**	0.545**	0.263	0.062	0.572**	0.386*	1

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

**Fig. 3** Hierarchical cluster results of dendrogram of sampling sites

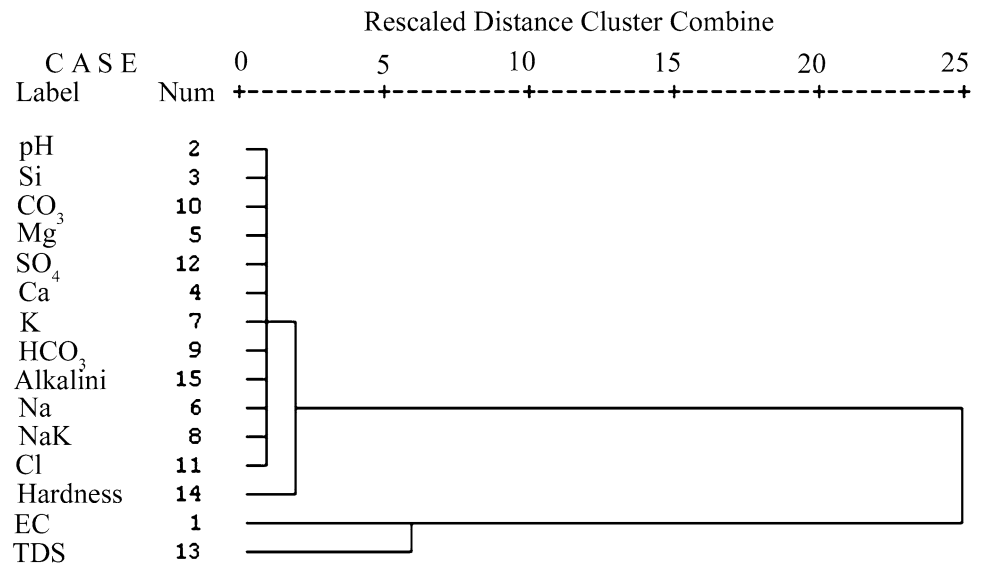
provided sampled groups of similar stations and created a dendrogram (Fig. 3). This has resulted into four statistically significant clusters from 30 sampling stations of Rapur area. Further, the phenon line was drawn across the dendrogram and are scaled at a distance of cluster combination of about 8. Thus, this position of the phenon line allows a division of the dendrogram into 4 clusters of water samples. The first cluster consists of 6 locations (1, 2, 11, 12, 13, and 21), the second cluster is composed of samples from 5 locations (4, 5, 6, 7, and 24), the third one consists only two samples (23 and 28), and the rest samples form the last cluster.

In the present study, Fig. 4 shows that the monitoring data obtained at 30 stations were classified and have produced a dendrogram. The clustering procedure generated two very convincing groups from the data. Cluster 1 consists of pH, Si, CO₃, Mg, SO₄, Ca, K, HCO₃, alkalinity, Na, Na + K, Cl, and hardness, which may be representative of overall effects of water–rock interactions and rock weathering on general groundwater quality, as this cluster includes almost every cation and anion of groundwater. Cluster 2 has two contributive parameters: EC and TDS. These two parameters represent the salinity of groundwater. Thus, cluster 2 can be regarded as a salinity factor. It may represent the overall effects of climatic factors, especially the evaporation, on groundwater salinity. As the groundwater salinity is also influenced by human activities in this area, the values of EC and TDS variation at the stations were greatly shaped by human activities such as industrial and municipal and agricultural activities, in addition to the factor of climate change mentioned above (Sayadi et al. 2014).

Principal component analysis (PCA)

The application of different multivariate statistical techniques, such as principal component analysis (PCA) assists in the interpretation of complex data matrices for a better understanding of water quality of a study area. A matter in the water quality monitoring is the complexity related with analyzing a large number of measured variables. The correlation matrix of variables was generated and factors were extracted by the centroid method, rotated by Varimax. From the results of Table 3, the first five eigenvalues were found to be bigger than 1 (Fig. 5).

Fig. 4 Dendrogram of water quality parameters



From Tables 3 and 4, it is clear that the first factor (factor 1), accounted for 36.218 of the total variance, had high positive loading in EC, Mg, Cl, TDS and Hardness which were 0.730, 0.900, 0.783, 0.769 and 0.909, respectively. This factor (factor 1) represents the contribution of EC to this factor, which can be considered a result of action-exchange processes in soil–water interface (Guo and Wang 2004). This factor also represents the contribution of point pollution and the physico-chemistry of the stream. The point pollution is from domestic wastewater, nonpoint pollution is from agricultural and livestock farms. Mg is a basic metal which increases Hardness of water (Razmkhah

et al. 2010). Factor 2 explains 17.742% of the total variance and is positively correlated with K, Na + K, and HCO₃ which were 0.931, 0.788 and 0.584, respectively. This factor may also be due to anthropogenic activities such as domestic waste water or influents. The loading for factor 3 was 11.470% for Ca and Na which is correlated at –0.788 and 0.902, respectively. Thus, this factor contains hydrogeochemical variable Ca, originating, at a first glance, from mineralization of the geological components of soils. The contribution of Ca to this factor can be considered a result of action-exchange processes in soil–water interface (Guo and Wang 2004).

Table 3 Extracted values of various factor analysis parameters along with total variance

Component	Initial Eigenvalues			Extraction sums of squared loadings			Rotation sums of squared loadings		
	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %	Total	% of variance	Cumulative %
1	5.433	36.218	36.218	5.433	36.218	36.218	3.971	26.476	26.476
2	2.661	17.742	53.960	2.661	17.742	53.960	2.892	19.281	45.757
3	1.721	11.470	65.430	1.721	11.470	65.430	2.489	16.595	62.352
4	1.585	10.565	75.995	1.585	10.565	75.995	1.918	12.784	75.136
5	1.187	7.916	83.911	1.187	7.916	83.911	1.316	8.775	83.911
6	0.856	5.706	89.617						
7	0.647	4.313	93.931						
8	0.368	2.456	96.386						
9	0.213	1.421	97.808						
10	0.140	0.935	98.743						
11	0.110	0.736	99.480						
12	0.046	0.309	99.789						
13	0.032	0.211	100.000						
14	8.403E–7	5.602E–6	100.000						
15	1.937E–16	1.291E–15	100.000						

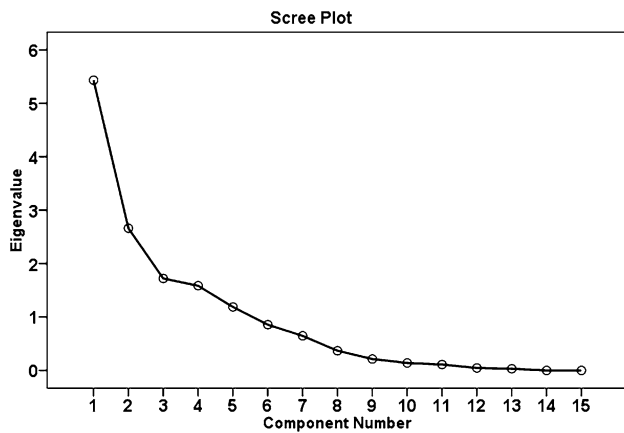


Fig. 5 Screen plot of the eigenvalue and component number

Factor 4 explains 10.565% of the total variance and is positively correlated with pH, Si, CO₃ and alkalinity which were 0.605, 0.657, 0.666, and 0.706, respectively. This factor contains hydrogeochemical variable pH, Si, CO₃, originating, at a first glance, from mineralization of the geological components of soils. The loading for factor 5 was 7.916% for SO₄ which is correlated at 0.916. The sources of dissolved SO₄ in natural waters may include dissolution of sedimentary sulfates, oxidation of both sulfide minerals and organic materials, and anthropogenic inputs.

Piper diagram

Piper (1944) based on the concentration of dominant cations and anions have proposed a trilinear diagram to show the percentages at mill equivalents per liter of cations

and anions in water samples. The Piper diagram was modified by Davis and Dewiest (1967). This is useful to understand the total chemical character of water samples in terms of cation–anion pairs. The Piper diagram reveals similarities and differences among groundwater samples because those with similar qualities will tend to plot together as groups (Todd 2001). This diagram is very useful in bringing out chemical relationships among groundwater in more definite terms (Walton 1970). The Piper diagram consists of two triangular and one intervening diamond-shaped field (Fig. 6). The percentage reacting values at the three cation groups—Ca, Mg, and (Na + K)—are plotted as a single point in the left triangular field and the three anion groups—(HCO₃ + CO₃), SO₄, and Cl—similarly on the right triangular field. The plots suggest that among cations Ca²⁺ do not have a clear dominance, and the other two major cations (Mg²⁺ and Na⁺) have greater dominance than Ca²⁺. Among anions Cl⁻ dominate in the groundwater samples. From Fig. 6, it is clear that the rate of increase in the concentration of Cl⁻ ions is more than SO₄²⁺ ions. It may be due to dissolution of evaporate minerals such as gypsum and halite (Sappa et al. 2014).

Conclusions

Water quality monitoring programs generate complex dataset that needs multivariate statistical methods for interpretation of the underlying information. The present work applied multivariate statistical techniques to study the source/genesis of chemical parameters in groundwaters of

Table 4 Rotated component loading matrix (loadings of 15 experimental variables)

	Component				
	1	2	3	4	5
EC	0.730	0.558	0.212	0.074	0.152
pH	-0.059	-0.241	-0.126	0.605	0.056
Si	0.173	-0.097	0.325	0.657	-0.075
Ca	0.317	-0.014	-0.788	-0.019	0.346
Mg	0.900	0.045	-0.107	0.150	-0.053
Na	0.152	0.015	0.902	0.194	0.185
K	0.101	0.931	-0.144	-0.128	0.026
NaK	0.195	0.788	0.536	0.034	0.156
HCO ₃	0.435	0.584	-0.395	0.259	-0.328
CO ₃	-0.157	0.366	0.224	0.666	0.261
Cl	0.783	0.230	0.429	-0.135	0.264
SO ₄	0.104	0.054	-0.021	0.073	0.916
TDS	0.769	0.549	0.234	0.051	0.093
Hardness	0.909	-0.027	-0.348	0.026	0.022
Alkalinity	0.411	0.435	-0.046	0.706	-0.096

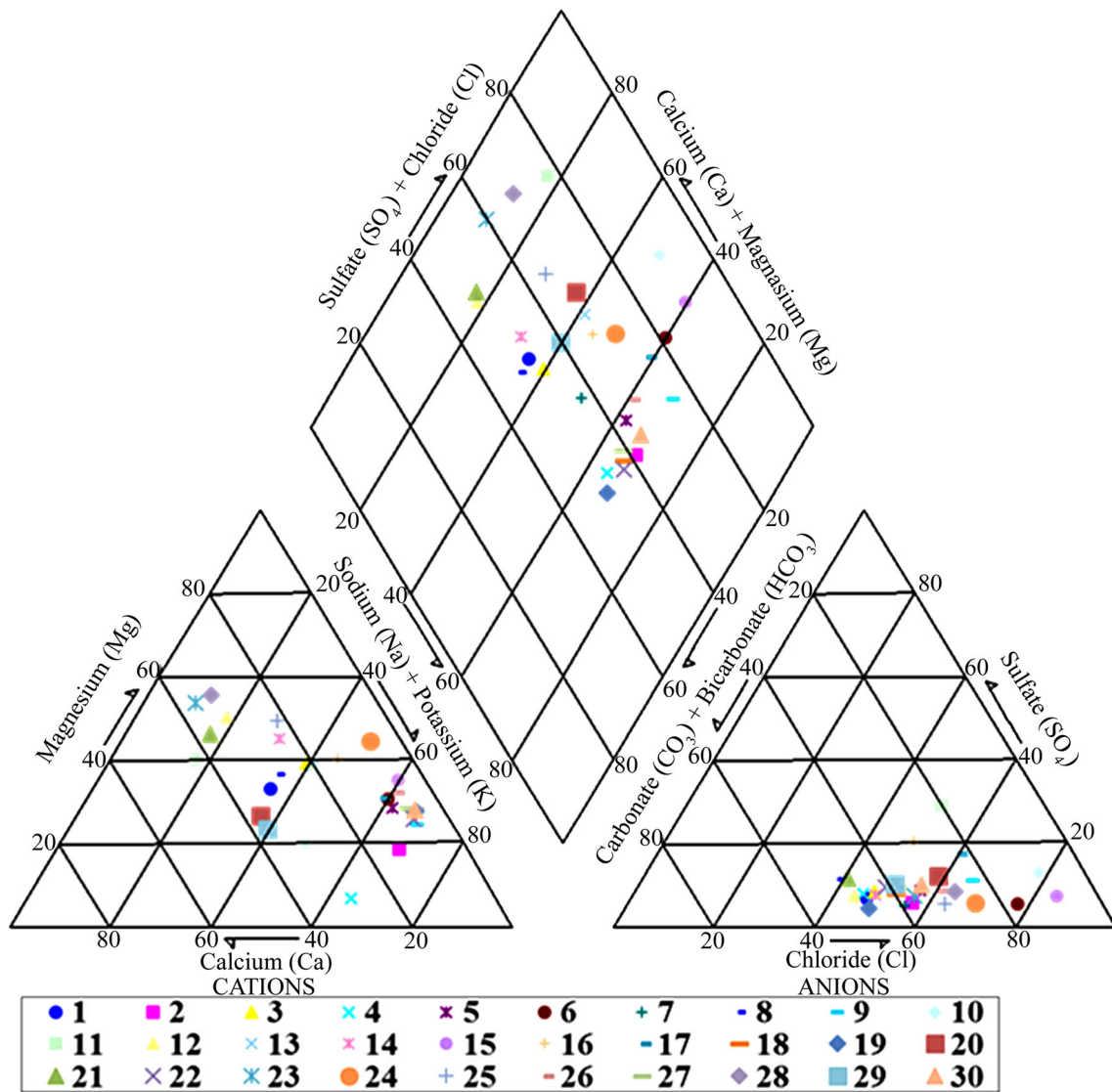


Fig. 6 Piper diagram for representing the analysis of ground water

the Rapur area of Andhra Pradesh including cluster analysis, and factor analysis. The factor analysis (FA) identified five factors responsible for data structure explaining 83.91% of total variance. The cluster analysis classified the locations into 5 clusters which possess different groundwater quality characteristics. It also yielded 2 clusters regarding water quality parameters. This study shows that multivariate analysis is a useful method that could assist decision makers in determining the source and extent of pollution.

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

Altun O, Türker Sacan M, Erdem AK (2008) Water quality and heavy metal monitoring in water and sediment samples of the Küçükçekmece Lagoon, Turkey. *Environ Monit Assess* 151:345–362

APHA (2005) Standard methods for the examination of water and wastewater, 21st edn. American Public Health Association, Washington

Ashley RP, Llyod JW (1978) An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. *J Hydrol* 39:355–364

Belkhir L, Boudoukha A, Mouni L, Baouz T (2010) Application of multivariate statistical methods and inverse geochemical modeling for characterization of groundwater—a case study: AinAzal plain (Algeria). *Geoderma* 159:390–398

Boyacioglu H (2008) Water pollution sources assessment by multivariate statistical methods in the Tahtali Basin, Turkey. *Environ Geol* 54(2):275–282

- 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* 353:294–313
- Davis J (1986) *Statistics and data analysis in geology*. Wiley, Toronto
- Davis SN, Dewiest RJM (1967) *Hydrogeology*. Wiley, New York
- De Bartolomeo A, Poletti L, Sanchini G, Sebastiani B, Morozzi G (2004) Relationship among parameters of lake polluted sediments evaluated by multivariate statistical analysis. *Chemosphere* 55:1323–1329
- Farnham IM, Johannesson KH, Singh AK, Hodge VF, Stetzenbach KJ (2003) Factor analytical approaches for evaluating groundwater trace element chemistry data. *Anal Chim Acta* 490(1–2):123–138
- Giridharan L, Venugopal T, Jayaprakash M (2009) Assessment of water quality using chemometric tools: a case study of river Cooum, South India. *Arch Environ Contam Toxicol* 56(4):654–669. doi:10.1007/s00244-009-9310-2
- Güler C, Thyne GD, Mccray JE, Turner AK (2002) Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol J* 10:455–474
- Guo H, Wang Y (2004) Hydrogeochemical processes in shallow quaternary aquifers from the northern part of the Datong Basin, China. *Appl Geochem* 19(1):19–27
- Gupta AK, Gupta SK, Patil RS (2005) Statistical analyses of coastal water quality for a port and harbour region in India. *Environ Monit Assess* 102(1–3):179–200
- Helena BA, Vega M, Barrado E, Pardo R, Fernandez L (1999) A case of hydrochemical characterization of an alluvial aquifer influenced by human activities. *Water Air Soil Pollut* 112(3):365–387
- İşçen CF, Emiroglu Ö, İlhan S, Arslan N, Yilmaz V, Ahiska S (2008) Application of multivariate statistical techniques in the assessment of surface water quality in Uluabat Lake, Turkey. *Environ Monit Assess* 144:269–276
- Kazi TG, Arain MB, Jamali MK, 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
- Kotti ME, Vlessidis AG, Thanasoulas NC, Evmiridis NP (2005) Assessment of river water quality in Northwestern Greece. *Water Resour Manage* 19:77–94
- Kouping C, Jiu JJ, Jianmin H (2006) Multivariate statistical evaluation of trace elements in ground water in a coastal area in Shenzhen, China. *Environ Pollut* 147:771–780
- Kowalkowska T, Zbytrowska R, Szejnab J, Buszewski B (2006) Application of chemometrics in river water classification. *Water Res* 40(4):744–752
- Kumar A, Rout S, Ravi PM (2013) Spatial geochemical variation of major and trace elements in the marine sediments of Mumbai Harbor Bay. *Environ Earth Sci* 70:3057–3066
- Li P, Qian H, Wu JH, Zhang YQ, Zhang HB (2013a) Major ion chemistry of shallow groundwater in the Dongsheng Coalfield, Ordos Basin, China. *Mine Water Environ* 32(3):195–206
- Li PY, Wu JH, Qian H (2013b) Assessment of groundwater quality for irrigation purposes and identification of hydrogeochemical evolution mechanisms in Pengyang County China. *Environ Earth Sci* 69(7):2211–2225. doi:10.1007/s12665-012-2049-5
- Li P, Qian H, Wu J (2014) Origin and assessment of groundwater pollution and associated health risk: a case study in an industrial park, northwest China. *Environ Geochem Health* 36(4):693–712. doi:10.1007/s10653-013-9590-3
- Li P, Zhang Y, Yang N, Jing L, Yu P (2016a) Major ion chemistry and quality assessment of groundwater in and around a mountainous tourist town of China. *Expo Health* 8(2):239–252. doi:10.1007/s12403-016-0198-6
- Li P, Wu J, Qian H, Zhang Y, Yang N, Jing L, Yu P (2016b) Hydrogeochemical characterization of groundwater in and around a wastewater irrigated forest in the southeastern edge of the Tengger Desert, Northwest China. *Expo Health* 8(3):331–348. doi:10.1007/s12403-016-0193-y
- 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* 313:77–89
- 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
- Monjerezi M, Vogt RD, Aagaard P, Saka JDK (2008) Hydrogeochemical processes in an area with saline groundwater in lower Shire River valley, Malawi: an integrated application of hierarchical cluster and principal component analyses. *Appl Geochem* 26:1399–1413
- Nagaraju A, Sunil Kumar K, Thejaswi A, Sharifi Z (2014a) Statistical analysis of the hydrogeochemical evolution of groundwater in the Rangampeta area, Chittoor District, Andhra Pradesh, South India. *Am J Water Resour* 2(3):63–70
- Nagaraju A, Sreedhar Y, Sunil Kumar K, Thejaswi A, Sharifi Z (2014b) Assessment of groundwater quality and evolution of hydrochemical facies around Tummalalalle area, Cuddapah District, Andhra Pradesh, South India. *J Environ Anal Chem* 1:1–6
- Nagaraju A, Thejaswi A, Sun L (2016a) Statistical analysis of high fluoride groundwater hydrochemistry in Southern India: quality assessment and implications for source of fluoride. *Environ Eng Sci* 33:471–477. doi:10.1089/ees.2015.0511
- Nagaraju A, Sharifi Z, Balaji E (2016b) Statistical and analytical evaluation of groundwater quality of Tirupati Area, Chittoor District, Andhra Pradesh, South India. *J Geol Soc India* 88:222–234
- Piper AM (1944). A graphic procedure in the geochemical interpretation of water analysis. *Trans Am Geophys Union Trans Washington, D.C.* 25:914–923
- Praus P (2005) Water quality assessment using SVD-based principal component analysis of hydrological data. *Water SA* 31(4):417–422
- Razmkhah H, Abrishamchi A, Torkian A (2010) Evaluation of spatial and temporal variation in water quality by pattern recognition techniques: a case study on Jajrood River (Tehran, Iran). *J Environ Manag* 91(4):852–860
- Reghunath R, Sreedhara MTR, Raghavan BR (2002) The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Res* 36:2437–2442
- Reimann RC, Filzmoser P, Garrett RG, Dutter R (2008) *Statistical data analysis explained: applied environmental statistics*. Wiley, Chichester
- Rezaei A, Sayadi MH (2015) Long-term evolution of the composition of surface water from the River Gharasoo, Iran: a case study using multivariate statistical techniques. *Environ Geochem Health*. doi:10.1007/s10653-014-9643-2
- Ribeiro L, Macedo ME (1995) Application of multivariate statistics, trend and cluster analysis to groundwater quality in the Tejo and Sado aquifer. In: *Groundwater quality: remediation and protection*. Proceedings of the Prague Conference, May 1995. IAHS Publ. No. 225, pp 39–47
- Sappa G, Ergul S, Ferranti F (2014) Geochemical modeling and multivariate statistical evaluation of trace elements in arsenic contaminated groundwater systems of Viterbo Area, (Central Italy). *SpringerPlus* 3(1):237
- Sayadi MH, Rezaei A, Rezaei MR, Nourozi K (2014) Multivariate statistical analysis of surface water chemistry: a case study of Gharasoo River, Iran. *Proc Int Acad Ecol Environ Sci* 4(3):114–122

- Schot PP, Van der wal J (1992) Human impact on regional groundwater composition through intervention in natural flow patterns and changes in land use. *J Hydrol* 134:297–313
- 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(1–2):355–374
- Spanos T, Simeonov V, Stratis J, Kristina X (2003) Assessment of water quality for human consumption. *Microchim Acta* 141(1–2):35–40
- Steinhorst RK, Williams RE (1985) Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis. *Water Resour Res* 21:1149–1156
- Todd DK (2001) *Groundwater hydrology*. Wiley, New York
- Voudouris KS, Lambrakis NJ, Papatheothourou G, Daskalaki P (1997) An application of factor analysis for the study of the hydrogeological conditions in Plio-Pleistocene aquifers of NW Achaia (NW Peloponnesus, Greece). *Math Geol* 29(1):43–59
- Walton WC (1970) *Groundwater resources evaluation*. McGraw-Hill Book Company, New York
- Wu J, Li P, Qian H, Duan Z, Zhang X (2014) Using correlation and multivariate statistical analysis to identify hydrogeochemical processes affecting the major ion chemistry of waters: a case study in Laohebaphosphorite mine in Sichuan, China. *Arab J Geosci* 7(10):3973–3982
- Yu S, Shang J, Zhao J (2003) Factor analysis and dynamics of water quality of the Songhua River Northeast China. *Water Air Soil Pollut* 144(1–4):159–169