International Conference "Water, Environment, Energy and Society" (WEES–2009)
New Delhi, 12–16 January 2009

Cluster Analysis of Groundwater Quality Data of Trombay Region, Maharashtra

K. Tirumalesh[1] and K. Shivanna
Isotope Applications Division, Bhabha Atomic Research Centre
Mumbai - 400 085, INDIA
E-mail: [1]tirumal@barc.gov.in

A.K. Sriraman
Water and Steam Chemistry Division, Bhabha Atomic Research Centre
Mumbai - 400 085, INDIA

ABSTRACT: This paper summarizes the findings obtained in a monitoring study to understand the sources and processes affecting the quality of groundwater in Trombay region using cluster analysis. Three distinct types of groundwater clusters were found, viz., Na–HCO_3–EC–pH–SO_4–Cl, Mg–Ca and SiO_2–K. These three different clusters indicate that different geochemical mechanisms control the chemical characteristics of the groundwater. The groundwater chemistry is mainly influenced by dissolution of minerals in the deep zone whereas base exchange dominates in the shallow zone. Enhanced mineralization in shallow zone compared to deep zone lindicate faster circulation of groundwater in deep zone preferably through fissures and fractures whereas groundwater flow is sluggish in shallow zone.

INTRODUCTION

Groundwater is important as a source of water for human consumption, agricultural and industrial use. Identifying the source and origin of groundwater, and understanding the geochemical processes affecting groundwater quality are crucial for the sustainable water supply. Groundwater quality depends on natural processes such as precipitation inputs, soil erosion and water-rock interaction, biota interrelationships and also anthropogenic influences such as urban, industrial and agricultural activities.

Multivariate analyses, such as cluster, factor and discriminant analysis, are useful for interpreting the governing processes through data reduction and classification (Robert, 1962). The use of these methods to water quality monitoring and assessment has increased in the last decade, mainly due to the need to obtain appreciable data reduction for analysis and decision (Vega *et al.*, 1998). Multivariate treatment of environmental data is widely used to characterize and evaluate surface waters (Reisenhofer *et al.*, 1995) and groundwater quality (Helena *et al.*, 2000; Kim *et al.*, 2005) and it is useful for evidencing temporal and spatial variations caused by natural and human factors linked to seasonality (Lingeswara Rao, 2003; Singh *et al.*, 2005).

Cluster analysis is a method for searching similar pair of relationship such as correlation in a large symmetric matrix. This analysis organizes large set of data into groups and enables straight forward and logical comparison of various chemical constituents (James, 1966).

In the present paper groundwater from shallow and deep zones of Trombay region was monitored in order to characterize the chemical nature of groundwater and to provide an understanding of its geochemical evolution using cluster analysis.

LOCATION, GEOLOGY AND HYDROGEOLOGY

This study area bound by hill on the north and west side, Mumbai Harbor Bay in the south and east side. The geographical coordinates are 19°00′–19°05′ N latitude and 72°54′–72°56′ E longitude. The high tide coastal line is about 200m towards south. The study area with sampling points is shown in Figure 1.

This area is covered by volcanic traps of Tertiary period. The subsurface geology reveals that the top zone consists of an over burden of 1 to 2 mere in thickness comprising clayey soil. This zone of overburden is followed by highly weathered basalt up to a depth of 9–10 meters. Below this zone fresh basalt with lateral joints and vertical cracks was observed.

The common rock forming minerals present in this area are plagioclase feldspars, pyroxene, augite and altered minerals like biotite and chlorite (Godse, 1967; Rakesh, 2005).

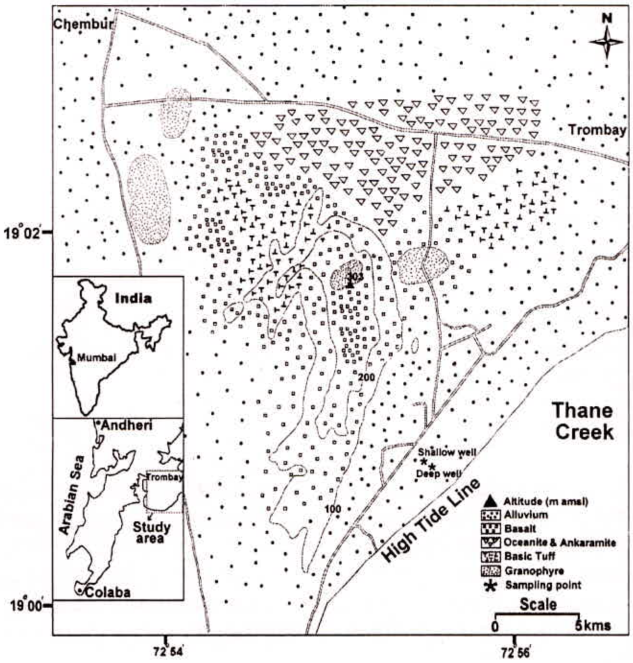

Fig. 1: Location map of the Trombay region with geological units and sampling points

The average annual rainfall is about 2700 mm. Groundwater is generally confined to the weathered strata, which forms the main aquifer zone and seasonal variation in water table is noticed. The groundwater flow direction is towards the Bay i.e., in the SE direction.

SAMPLING AND MEASUREMENTS

Water samples from bore wells tapping shallow (2 m bgl) and deep zones (10 m bgl) were collected monthly from the three bore wells for a period of 25 months. Samples collected were filtered using 0.45 μm pore size membrane filters and stored in polyethylene bottles, which were initially washed with concentrated HNO_3 and rinsed thoroughly with distilled water. For cation measurement a duplicate set was collected and acidified to pH< 2 by adding ultra pure concentrated HNO_3.

Physical parameters like pH, Temperature, Dissolved Oxygen (DO) and Electrical Conductivity (EC) were measured in situ using portable pH/DO/Temp meter (Corning, model 313) and conductivity meter (Orion model 130). Alkalinity was measured by titrating 10 mL of water sample with 0.02 N H_2SO_4. A mixed indicator (Bromocresol green—Methyl red) was used to mark the end point of the reaction at pH 4.3. Permanganate Demand (PD), Total Hardness (TH), E-coli and Total Dissolved Solids (TDS) were measured as per standard methods (APHA 1995). Anions viz., Cl^-, SO_4^{2-}, NO_3^-, F^- and Br^- and cations viz., Na^+, Ca^{2+}, Mg^{2+}, K^+ and NH_4^+ were analyzed using ion chromatography (DX-500) employing electrochemical detector (ED 40) in conductivity mode. Dissolved silica was measured by spectrophotometer.

RESULTS AND DISCUSSION

Quality Assurance

The general quality controls adopted in this laboratory were: calibration with known standards, analysis of reagent blanks, recovery test, analysis of duplicates and analysis of control samples. In order to assess the quality of results, the laboratory participated in the United Nations (UN) GEMS/Water Performance Evaluation (PE) studies no 5 and 6 organized by International Atomic Energy Agency (IAEA) under RAS/8/097 during 2004 and 2005 respectively. In these tests the laboratory performance was evaluated by way of a Z-score, $Z = (X_{lab}–X_V)/s$, where X_{lab} is the laboratory value, X_V is the value accepted as the true one and s is the standard deviation of the laboratories mean value. Laboratory performance is evaluated as point of excellence if Z-score is between 0 to 0.15, good, between 0.15 to 0.32, satisfactory between 0.32 to 1.645, point of opportunity between 1.645 to 2.258 and point of immediate concern above 2.58. The Z-score values obtained for each parameter are presented in Figure 2. All the results fall under point of excellence or good or satisfactory category and none of them in point of opportunity or point of immediate concern category.

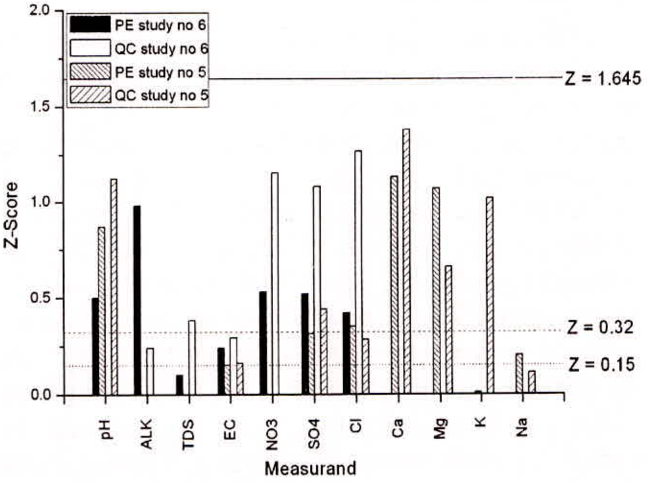

Fig. 2: Z-score values for inter laboratory evaluation

The limit of detection and the limit of quantification are also characteristics for the evaluation of any analytical method as they allow the determination of the minimum quantities that a method can detect or quantify respectively. These parameters are necessary to check if the method is performing with adequate accuracy at the required concentration levels. These limits were calculated from the standard deviation of the blank multiplied by three for the limit of detection and by ten for the limit of quantification. The precision of the methods was estimated from duplicate analyses. The relative standard deviations for all the measured parameters are less than 5%.

The accuracy of the chemical data was also checked by calculating Charge Balance Error (CBE) given by,

$$\text{CBE (\%)} = \frac{\text{meq (cations)} - \text{meq (anions)}}{\text{meq (cations)} + \text{meq (anions)}} \times 100 \quad \dots (1)$$

The CBE values were found to range from –5.7 to +5.6% but most of the groundwater samples (70 out of 75) show CBE within allowed limits i.e. ±5%.

Basic statistics of hydro-chemical data of the groundwater samples are given in Table 1 to give an estimate of the variation in chemistry.

Cluster Analysis

For cluster analysis, A 10 × 10 matrix of correlation coefficient was computed using MegaStat (Micrsoft-XL add-ins) software. Linear correlations of different chemical constituents of groundwater samples are given in Table 2. Similar pair of chemical constituents was linked based on the correlation coefficients and then the next most pair and so on until all the chemical constituents have been clustered. In the first step, mutually highest correlation value in each column of the matrix was identified as shown in Table 2 in bold face type. Similar highest correlation coefficients of chemical constituents such as $Na–HCO_3$ and Mg–Ca were clustered first. The new correlation coefficients between $Na–HCO_3$ and Mg–Ca clusters and independent constituents were recalculated by arithmetic averaging method (Davis 1973) and the rest of the correlation coefficients were retained as such. Clustering steps of chemical constituents is shown in the dendrogram (Figure 3).

The cluster analysis suggests that three main inter-related types of chemical constituents are responsible for the Hydrogeochemical variability in the quality of groundwater. Cluster-I ($Na–HCO_3–EC–pH–SO_4–Cl$), cluster-II (Mg–Ca) and cluster-III ($SiO_2–K$).

Table 1: Statistical Parameters for the Groundwater Quality of the Study Site

	Bore Well No. 1				*Bore Well No. 2*				*Bore Well No. 3*			
	Min.	*Max.*	*AVG*	*Median*	*Min.*	*Max.*	*AVG*	*Median*	*Min.*	*Max.*	*AVG*	*Median*
EC	120	380	259	280	265	553	3894	360	250	440	313	300
TDS	66	312	174	182	181	335	257	255	145	323	228	226
pH	7.2	8.6	7.9	8.1	7.7	8.7	8.2	8.2	7.2	8.6	7.8	7.6
Na^+	6.1	51.8	18	17.5	14.2	115	58.3	58.4	10.9	52.2	18.3	13.5
K^+	0.4	3.2	1.2	1.1	0.6	3.5	1.3	1	0	1	0.4	0.4
NH_4^+	0	0.7	0.1	0	0	0	0	0	0	2	0.4	0
Ca^{2+}	13.4	47.5	23.4	23	9.4	31.2	16.2	14.7	11.8	34	23.8	23.2
Mg^{2+}	4.6	13	10.1	11	5.9	13	7.7	7.1	8.6	14.1	12.4	12.5
HCO_3^-	55	218	128	125	150	232	198	200	130	180	152	150
F^-	0.1	0.4	0.2	0.2	0.2	0.8	0.3	0.2	0.1	0.3	0.2	0.2
Cl^-	8.7	37.9	15.5	14.5	12.4	46.2	17.5	15.9	9.3	19.3	12.3	11
NO_3^-	0	1.4	0.5	0.5	0.2	2.4	0.8	0.6	0	2.4	0.75	0.5
SO_4^{2-}	5	23.1	14.7	15.3	14.8	28.7	19.1	18.5	11.6	24.3	16.1	15
SiO_2	3.2	24.6	17.604	19	14.2	46.6	20.1	18.9	3.2	24.6	17.6	19
TH	52.3	133	99	102	56.3	137	95	88.3	72	172	135	136
PD	0.65	1.3	0.99	1	0.7	1.6	1.1	1	1	3.6	1.6	1.5
DO	4	7	5.2	5	4	7	4.9	5	4	7	5.3	5

Note: EC is in μS/cm and others in mg/L

Table 2: Different Step Correlation Matrices (highest correlation of a column are shown in bold)

First Step Correlation Matrix

	pH	EC	Na	K	Ca	Mg	HCO_3	Cl	SO_4	SiO_2
pH	1									
EC	0.3	1								
Na	0.41	0.53	1							
K	0.08	-0.11	0.23	1						
Ca	-0.12	-0.22	-0.61	-0.16	1					
Mg	-0.3	-0.02	-0.59	-0.42	0.64	1				
HCO_3	0.33	0.7	0.77	0.06	-0.17	-0.17	1			
Cl	0.32	0.25	0.38	0.16	-0.08	-0.13	0.22	1		
SO_4	0.17	0.46	0.63	0.1	-0.31	-0.05	0.48	0.53	1	
SiO_2	0.02	0.21	0.21	0.14	-0.15	-0.1	0.17	0.17	0.17	1

Second Step Correlation Matrix

	Na-HCO_3	Mg-Ca	pH	EC	K	Cl	SO_4	SiO_2
Na-HCO_3	1							
Mg-Ca	-0.39	1						
pH	0.37	-0.21	1					
EC	0.62	-0.12	0.3	1				
K	0.15	-0.29	0.08	-0.11	1			
Cl	0.3	-0.1	0.32	0.25	0.16	1		
SO_4	0.55	-0.18	0.17	0.46	0.1	0.53	1	
SiO_2	0.19	-0.13	0.02	0.21	0.14	0.17	0.17	1

Third Step Correlation Matrix

	Na-HCO_3-EC	Mg-Ca	pH	K	Cl	SO_4	SiO_2
Na-HCO_3-EC	1						
Mg-Ca	-0.25	1					
pH	0.33	-0.21	1				
K	0.02	-0.29	0.08	1			
Cl	0.28	-0.1	0.32	0.16	1		
SO_4	0.51	-0.18	0.17	0.1	0.53	1	
SiO_2	0.2	-0.13	0.02	0.14	0.17	0.17	1

Fourth Step Correlation Matrix

	Na-HCO_3-EC	Mg-Ca	SO_4-Cl	pH	K	SiO_2
Na-HCO_3-EC	1					
Mg-Ca	-0.25	1				
SO_4-Cl	0.39	-0.14	1			
pH	0.33	-0.21	0.24	1		
K	0.02	-0.29	0.13	0.08	1	
SiO_2	0.2	-0.13	0.17	0.02	0.14	1

Fifth Step Correlation Matrix

	Na-HCO_3-EC-SO_4-Cl	Mg-Ca	pH	K	SiO_2
Na-HCO_3-EC-SO_4-Cl	1				
Mg-Ca	-0.2	1			
pH	0.29	-0.21	1		
K	0.07	-0.29	0.08	1	
SiO_2	0.19	-0.13	0.02	0.14	1

Sixth Step Correlation Matrix

	Na-HCO_3-EC-SO_4-Cl-pH	Mg-Ca	K	SiO_2
Na-HCO_3-EC-SO_4-Cl-pH	1			
Mg-Ca	-0.2	1		
K	0.08	-0.29	1	
SiO_2	0.1	-0.13	0.14	1

Seventh Step Correlation Matrix

	Na-HCO_3-EC-SO_4-Cl-pH	Mg-Ca	K-SiO_2
Na-HCO_3-EC-SO_4-Cl-pH	1		
Mg-Ca	-0.2	1	
K-SiO_2	0.09	-0.21	1

Eighth Step Correlation Matrix

	Na-HCO_3-EC-SO_4-pH-K-SiO_2	Mg-Ca
Na-HCO_3-EC-SO_4-pH-K-SiO_2	1	
Mg-Ca	-0.21	1

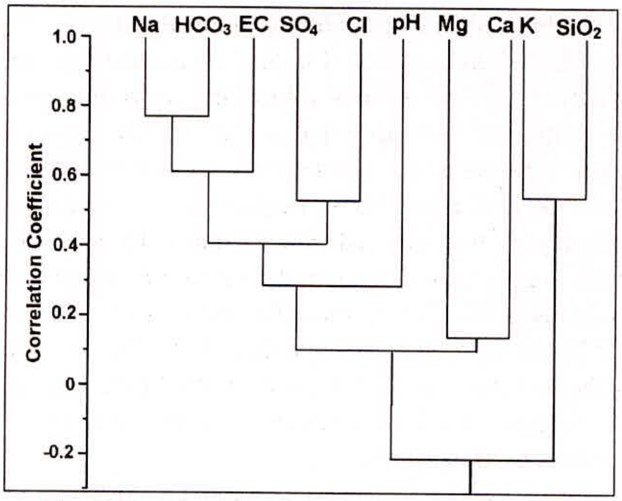

Fig. 3: Dendrogram showing various clusters of chemical constituents

In cluster-I, sodium and bicarbonate are correlated positively at higher level. The Na-HCO_3 cluster is further linked with EC, SO_4, Cl and pH at lower level of correlation. Moreover, they are clustered with positive correlation. Generally shallow zone show Na–HCO_3 type of water. The principle cause for this correlation is base exchange reaction with the clays. The base exchange with clay can be given as:

$$Na_2\text{–clay} + Ca^{2+} (Mg^{2+}) \rightarrow 2Na^+ + Ca^{2+} (Mg^{2+})\text{– clay}$$

To ascertain the possibility of base exchange reaction occurring in the groundwater, $(Ca^{2+} + Mg^{2+})–(HCO_3^- + SO_4^{2-})$ was plotted against Na–Cl values (Figure 4). Waters undergoing base exchange would fall on a line with slope –1 while waters plotting close to the zero value on the x-axis are not influenced by base exchange (Benony 2007). Figure 3 shows that the slope of the best fit equation is close –1 indicating the source of Na^+ is base exchange and not albite or other sea sources.

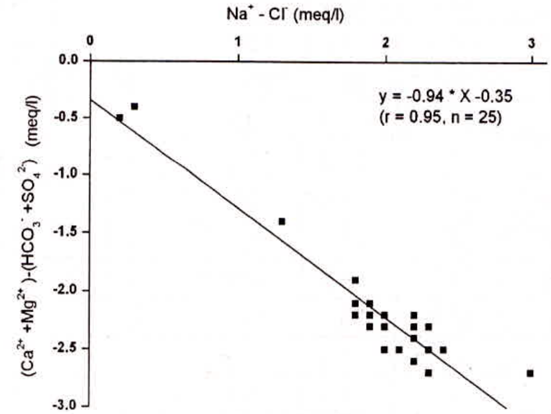

Fig. 4: Plot of $Ca^{2+} + Mg^{2+} – HCO_3^- – SO_4^{2-}$ versus $Na^+–Cl^-$ of shallow zone samples

The shallow zone water show more TDS and EC values (Figure 5) which further point to long residence time of groundwater facilitating higher mineralization and effective base exchange reaction.

Cluster-II represents contribution of alkaline earth minerals present in the subsurface. This cluster is dominated by Mg^{2+} and Ca^{2+} with a correlation coefficient of 0.64 and it is always negatively correlated to other constituents and clusters. This cluster is predominant in deep zone samples. The common sources for these ions in basalts are generally calcic plagioclase (anorthite) and pyroxene (olivine). The weathering reactions can be represented as follows:

$$CaAl_2Si_2O_8 + 2CO_2 + 3H_2O \rightarrow Al_2Si_2O_5\,(OH)_4 + Ca^{2+} + HCO_3^{-1} \text{ (Anorthite weathering)}$$

$$Mg_2SiO_4 + 4CO_2 + 4H_2O \rightarrow 2Mg^{2+} + 4HCO_3^{-} + 2H_4SiO_4 \qquad \text{(Olivine weathering)}$$

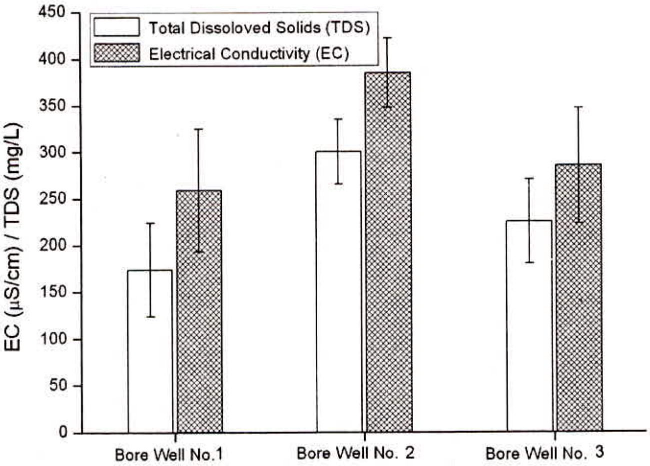

Fig. 5: Histogram of average TDS and EC of three wells with standard deviation (mean ±1 σ)

Groundwater derived from basalt formation has a distinguishable higher Mg^{2+}/Ca^{2+} ratio which is evident in these groundwaters. The $Mg^{2+}/(Ca^{2+} + Mg^{2+})$ ratio is >0.5 and $HCO_3^-/SiO_2 < 10$ indicating silicate weathering (Hounslow 1995).

Cluster-III comprises of SiO_2 and K with poor but positive correlation. This indicates common source of potassium and silica possibly biotite mineral present in the subsurface. Both shallow and deep zones contain this type of water. The hydrolysis of potash feldspar is given as follows;

$$KAlSi_3O_8 + 8H_2O \rightarrow Al(OH)_3 + 3H_4SiO_4 + K^+ + OH^- \qquad \text{(K-feldspar weathering)}$$

Since K-feldspar are comparatively less soluble in water, other sources for potassium and silica affects their correlation leading to poor correlation ($r = 0.14$). The poor correlation between potassium and silica can also be due to removal of potassium through absorption and bioaccumulation by thick and dense vegetation spread over entire study area.

CONCLUSIONS

An attempt has been made to understand the sources and processes affecting the quality of groundwater in Trombay region using cluster analysis. Three distinct types of groundwater point to heterogeneity in the subsurface geology and different geochemical mechanisms controlling the chemical characteristics of the groundwater. The distribution of chemical species in groundwater is mainly controlled by dissolution of anorthite, olivine and K-feldspar rock forming minerals in the deep zone and base exchange with clayey soil in the shallow zone. Low TDS and EC in deep zone indicate fast circulation of groundwaters through fissures and fractures whereas groundwater flow is sluggish in shallow zone leading to comparatively high TDS and EC values.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Gursharan Singh (Head, Isotope Applications Division), Dr. S.V. Narasimhan (Head, water and steam Chemistry Division), Shri Ramkishen (Head, Engineeing services section, Trombay) and Dr. A.K. Tyagi (Chemistry Division) for their active support at various stages of the above programme. Thanks are also due Shri S.B. War and Shri P.S. Shetty for their valuable assistance and guidance during the project.

REFERENCES

APHA, AWWA, WEF (1995). "Standard Methods for the Examination of Water and wastewater", 19th edition, American public Health Association, Washington.

Benony, IK. Kortatsi (2007). "Hydrochemical framework of groundwater in the Ankobra Basin, Ghana". *Aqua. Geochem.* 13, 41–74.

Davis John, C. (1973). "Statistics and data analysis in Geology", John Wiley & sons, Inc. New York, pp. 456–473.

Godse, V.B., Mohan, A.L., Mahesh Singh, Amalraj, R.V. and Thomas, K.T. (1967). 'Characterization of Trombay soils from disposal of radioactive wastes". Sym. Disposal of Radioactive Wastes into the Ground, held at IAEA/ENEA Vienna.

Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J.M. and Fernandez, L. (2000). "Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principle component analysis". Wat. Res. 34, 807–816.

Hounslow, W.A. (1995). "Water quality data- analysis and interpretation", Lewis Publishers, New York.

James, M. Parks (1966). "Cluster analysis applied to multivariate geological problems". *Jour. Geol.* 74, 703–715.

Kim, J.H., Kim, R.H., Leė, J., Cheong, T.J., Yum, B.W. and Chang, H.W. (2005). "Multivariate statistical analysis to identify the major factors governing groundwater quality of in coastal area of Kimje South Korea". *Hydrol. Proces.* 19, 1261–1276.

Lingeswara Rao, S.V. (2003). "Cluster analysis of groundwater quality data of Venkatagiri Taluq, Nellore Distrit, Anfhra Pradesh". *Jour. Geol. Soc. Ind.* 62, 447–454.

Rakesh, R.R., Yadav, D.N, Narayan, P.K. and Nair, R.N. (2005). B.A.R.C, Report—B.A.R.C/2005/I/010.

Reisenhofer, E., Adami, G. and Barbieri, P. (1998). "Using chemical and physical parameters to define the quality of karstic freshwaters (Timavo River North-eastern Italy); *a chemometric approach*". Wat. Res. 32, 1193–1203.

Robert, L. Miller and James Steven Kahn (1962). "Statistical analysis in the geological sciences", John Wiley& sons, Inc. New York, pp. 259–299.

Singh, K.P., Malik, A. and 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.

Vega, M., Pardo, R., Barrado, E. and Deban, L. (1998). "Assessment of seasonal and polluting effects on quality of river water by exploratory data analysis". *Wat.Res.* 32, 3581–3592.