Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

Vol. 8	No. 4	pp. 657-664	2009

Original Research Paper

Interpretation of Water Quality Data by Multivariate Statistical Tools: A Study in Mysore District, Karnataka, India

Nilufer Arshad and G. S. Gopalakrishna
DOS in Geology, University of Mysore, Mysore-570 006, Karnataka, India

Key Words:
Water quality
Groundwater
Multivariate statistical analysis
Mysore district

ABSTRACT

In the management of water resources, variables which control the quality of water, are just as important as its quantity. Hydrochemical assessment of water quality of surface and groundwater for 58 samples was made during postmonsoon of 2007 from strategic locations in Husnur Taluk. Water quality data collected from different localities are used in conjunction with multivariate statistical technique to identify key variables. In surface, four components were extracted which account for 84.46% of the total variance. The first component shows that the EC and TDS play an important role in the hydrochemical constituents of the surface water. In groundwater samples, 5 components were extracted, which account for 95 % of the total variance. The maximum number of variables, i.e., Na, Cl, SO_4, TDS and EC were characterized by the first component and show that the hydrochemical constituents of groundwater are mainly controlled by the first component. The 'single dominance' nature fourth and fifth components in PCA indicate non-mixing or partial mixing of different types of groundwaters. The findings of the cluster analysis are presented in the form of dendogram of the sampling stations (cases) which produced three major groups.

INTRODUCTION

Multivariate analysis techniques are very useful in the analysis of data corresponding to a large number of variables. Analysis via these techniques produces easily interpretable results. A unique feature of adopting these techniques is, some deal with the relationships between the variables and the others are primarily concerned with relationship between samples. Some of the recent reports, which have utilised the multivariate techniques in water quality studies, are of Reghunath et al. (2002), Nicolaos Lambrakis et al. (2004), Bernard Parinet et al. (2004), Debasis Deb et al. (2008) and Pathak et al. (2008). In this multivariate analysis study, principal component analysis (PCA) and hierarchical cluster analysis (HCA) were employed to investigate the factors which cause variations in the observed quality data in the investigated area.

STUDY AREA

Western part of Hunsur Taluk, Mysore district, Karnataka has been chosen for the present investigation on the water resources to assess the differences in the quality of water resources and their relationship together (Fig. 1). The study region covers an area of 633.77 sq. km based on the toposheets of survey of India numbers 57D/3, 57D/4, 57D/7 and 57D/8 on a scale of 1:50000 and lies between latitudes 12°25' to 12°15'N and longitudes 76°5' to 76°25'E**.** The Lakshmantirtha river is a tributary of Cauvery basin and originates in the western ghats of Kodogu district, passes through Hunsur and confluences with River Cauvery at KRS dam. The rainfall is highly variable in its distribution over time and space. The annual average rainfall is 680-750 mm. The southwestern monsoon

contributes more that 60% of annual rainfall from June-September. During October and December, the northeastern monsoon brings rainfall to the area. The lithology is composed of the Gneiss complex which comprises the basement rock of the major part of the study area and shows foliation in various degrees, which later were intruded by acid and basic Palaeoproterozoic rocks in dyke form which are confined mostly in the central part of the study area. Most of the intrusive rocks are of Quartz vein, Dolerite, Gabro and Pegmatite vein (Geological Society of India 2006).

MATERIALS AND METHODS

Fifty eight water samples from three different sources (groundwater, lakes and reservoirs and Lakshmantirtha river) were collected during postmonsoon period (December) in the year 2007. Fig. 1 shows the locations of the water samples. Table 1 gives sample No. and sample type. The samples collected were analysed for pH, electrical conductivity (EC), Na, Ca, Mg, K, HCO_3, Cl and SO_4 by standard procedures prescribed by APHA (1995).

Data standardization: The suitability of data for carrying out the analysis should be determined by using Keiser-Meyer-Olkin (KMO) and Bartlett's tests (Dennis Child 2006). KMO is a measure of sample adequacy. If only KMO value is greater than 0.5, the PCA can be used. If KMO value is less than 0.5 performing PCA/factor analysis will not be appropriate. Bartlett's test measures the relationship between the variables at a significance level. A significant relationship should exist among variables to carry out the analysis.

Principal component analysis: PCA is a statistical method used to determine components that are linear combinations of the original variables. In PCA, the first principal component is the linear combination of the variables with maximal variance and represents the largest variability of the

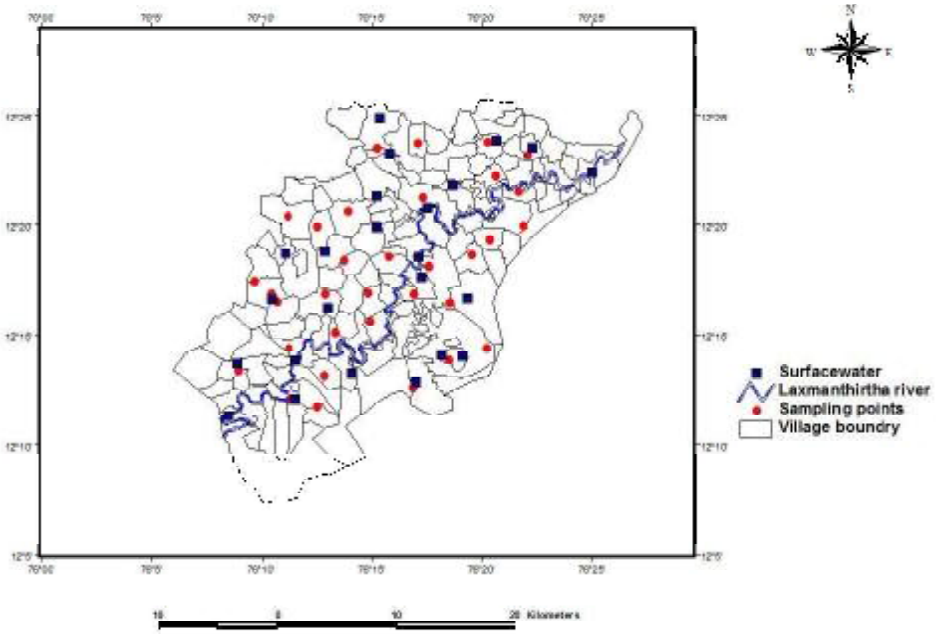

Fig 1: Study area with location of sampling points.

Table 1: Sample no. and sample type.

Sample No.	Sample Type	Sample No.	Sample Type	Sample No.	Sample Type
S1	Surface water	S21	Groundwater	S41	Groundwater
S2	Groundwater	S22	Surface water	S42	Surface water
S3	Surface water	S23	Groundwater	S43	Groundwater
S4	Surface water	S24	Groundwater	S44	Groundwater
S5	Groundwater	S25	Surface water	S45	Groundwater
S6	Surface water	S26	Surface water	S46	Groundwater
S7	Groundwater	S27	Groundwater	S47	Groundwater
S8	Surface water	S28	Groundwater	S48	Surface water
S9	Surface water	S29	Surface water	S49	Surface water
S10	Groundwater	S30	Surface water	S50	Surface water
S11	Groundwater	S31	Surface water	S51	Surface water
S12	Surface water	S32	Groundwater	S52	Groundwater
S13	Groundwater	S33	Groundwater	S53	Groundwater
S14	Surface water	S34	Surface water	S54	Groundwater
S15	Surface water	S35	Groundwater	S55	Groundwater
S16	Groundwater	S36	Groundwater	S56	Groundwater
S17	Surface water	S37	Groundwater	S57	Groundwater
S18	Groundwater	S38	Groundwater	S58	Groundwater
S19	Surface water	S39	Groundwater		
S20	Surface water	S40	Groundwater		

original data set. The second component is the linear combination with the next largest variability that it is orthogonal to the first component, and so on. The principal components are used to discover and interpret the dependence that exist among the variables and to examine relationships that may exist among them . The following procedure is adopted in applying the principal components analysis for the study.

***Correlation matrix*:** Correlation and covariance matrix are the two different matrixes which are used in PCA. In this study correlation matrix has been used. The sums of squares and sums of products of the normalized scores constitute the correlation matrix (R) (Hope 1986). This means that the variables have been standardized to have unit variance. The use of the R matrix for analysing involves a decision that variables have been considered equally important (Chatfield & Collins 1980). Karpuzcu & Sene (1987) stated that if parameters (variables) are in widely different units (mg/L, pH, m_3/min, etc.), then standard variates and correlation matrix should be used.

***Identification of important components*:** By using correlation matrix, the variances of the variable (eigen value) and principal components (eigen vectors) will be computed.

Rotation of principal components: The most important principal components selected are rotated and a new set of components will be generated which can be more easily interpreted. A variety of rotation techniques (varimax, equamax, quartimax) may be used for this purpose. Varimax rotation is the most widely used rotation in principal component analysis.

This technique tends to eliminate medium-range correlations between the components and the original variables, thus, simplifying the decision as to which of the original variables to include in the components extracted (Chatfield & Collins 1980).

RESULTS AND DISCUSSION

Principal component analysis was performed on some of the water quality indicators obtained by the

Tabel 2: KMO and Bartlett's test (Surface water).

Kaiser-Meyer-Olkin measure of sampling adequacy.		0.550
Bartlett's Test of Sphericity	Approx. Chi-Square	293.664
	df	66
	Sig.	0.000

Table 3: Total variance explained (Surface water).

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.728	39.403	39.403	3.505	29.210	29.210
2	2.718	22.653	62.056	3.254	27.117	56.327
3	1.530	12.750	74.805	2.136	17.800	74.127
4	1.159	9.656	84.461	1.240	10.334	84.461

Extraction Method: Principal Component Analysis
Rotation Method: Varimax with Kaiser Normalization

physicochemical analysis of the water samples. With using tools of the selection variables in factor analysis menu of SPSS software (V.17) the sample type was chosen separately for the surface and groundwater resources and results are discussed as below.

Surface Water

Since KMO value was greater than 0.5, it indicates the existence of a statistically acceptable factor solution representing relations among the parameters. The Bartlett's test showed that there was significant relationship among the variables (Table 2). The PCA was applied and 9 components were formed which equalled the 9 variables.

Component selection: According to Cattell (1966), with the help of scree plot the number of components can be reduced which helps in better interpretation. By examining the scree plot it is noticed that the line starts to level off (elbow) at the point 4 (Fig. 2) so the number of components were eliminated from the point 4 onwards. Hence, number of components was restricted to 4, which include 84.4% of the total variance (Table 3).

By using the extraction method in PCA, five components were extracted (Table 4). The first component with an eigen value of 4.72 (accounting for 39.4% of the total variance) is characterized by high loading of Mg, TDS, hardness and EC suggesting the water chemistry to be mainly controlled by TDS, hardness and EC. The second component (accounting for 22.65% of the total variance) is associated with high loading of Cl, Na and SO_4 and a negative loading of F. The third factor (accounting for 12.75% of the total variance) includes K, PO_3 and Ca variables. The last component (accounting for 9.65% of the total variance) is NO_3 and it is

Table 4: Rotated Component Matrix[a] (Surface water).

Parameters	Component			
	1	2	3	4
EC	.885			
TDS	.890			
Ca			.771	
Mg	.920			
Hardness	.844			
F		–.819		
NO_3				.941
PO_3			.859	
SO_4		.802		
Chloride		.857		
Na		.813		
K			.687	

a. Rotation converged in 5 interactions.

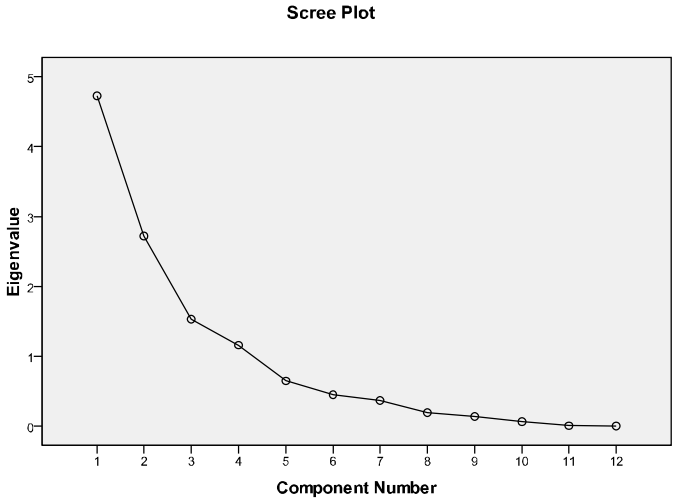

Fig. 2: Scree plot for surface water

characterized by single dominance variable. This could be attributed to the nitrification process, which takes place mainly in the lakes and reservoirs due to addition of domestic discharge and agricultural runoff.

Groundwater

In the selection variable of the factor analysis menu in SPSS software, groundwater was selected for carrying out the PCA and the remaining procedure was the same as adopted for surface water as explained above. As shown in Table 5, KMO is greater than 0.5 and Bartlett's test shows a significant level (0.0) and exist a significance dependence between the variables, hence, the PCA analysis was fit for this study on the groundwater samples.

Table 5: KMO and Bartlett's test (Groundwater).

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.610
Bartlett's Test of Sphericity	Approx. Chi-Square	588.847
	df	66
	Sig.	0.000

Table 6: Total variance explained (Groundwater).

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.289	52.406	52.406	4.328	36.065	36.065
2	2.066	17.218	69.624	2.608	21.736	57.802
3	1.504	12.532	82.155	1.703	14.195	71.997
4	1.018	8.485	90.640	1.541	12.838	84.835
5	.520	4.330	94.970	1.216	10.135	94.970

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Table 7: Rotated component matrix[a] (Groundwater).

	Component				
Parameters	1	2	3	4	5
EC	.951				
TDS	.860				
Ca		.781			
Mg			.962		
Hardness		.697			
F				.934	
NO_3			.626		
SO_4	.868				
Chloride	.868				
Na	.687				
K					.793
Total Alkalanity		.946			

a. Rotation converged in 6 iterations.

By following the Cattelles principal 5 components were extracted (Fig. 3) and included more than 94% of the total variance of the variables (Table 6).

By using the extraction method in PCA, five components were extracted (Table 7). The first component with an eigen value of 6.28 is characterized by 5 variables which accounts for 52% of the total variance of the variables. These variables include high loadings of Na, SO_4 and NO_3 which contribute to the TDS and together account for the high loading of EC with an eigen value of 0.951. This component is associated with a combination of various hydrogeochemical processes that contribute to enrich more mineralized water (high value of TDS), as suggested by Rao et al. (2006). When comparing this component with the first component extracted in surface water, one thing which is very clear, is that the variables which influence EC for both of the water resources differ from one another and implies the quality variation among them. The second component extracted with an eigen value of 2.06 was characterized by high loadings of total alkalinity and moderate to high loading of Ca and hardness. According to Rao et al. (2006) the mineral dissolution during water-soil and water-rock interactions depends upon the amount of CO_2 which originates from HCO_3. The concentration of HCO_3 in groundwater is result of the reaction of soil CO_2 with dissolution of silicate minerals. The association Ca with hardness too represents temporary hardness of water. The third component, which accounts for 12.5% of the total variance, includes Mg which again contributes to the hardness of the water but it is of lesser significance in comparison to Ca. NO_3 too belongs to the third component. Factors 4-5 are characterised by the dominance of only one variable, such as F

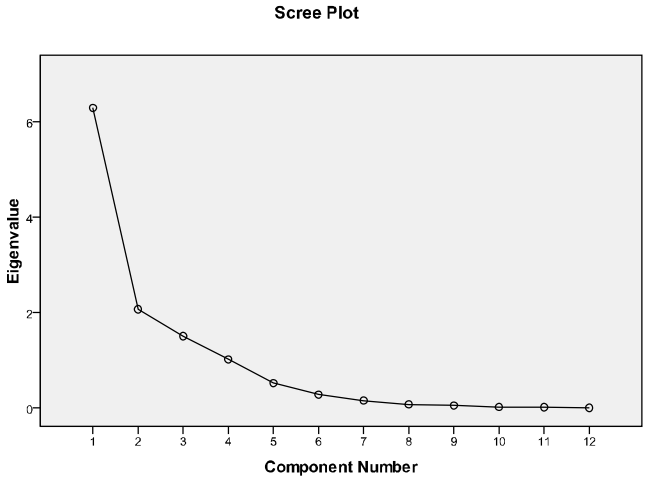

Fig. 3: Scree plot for groundwater.

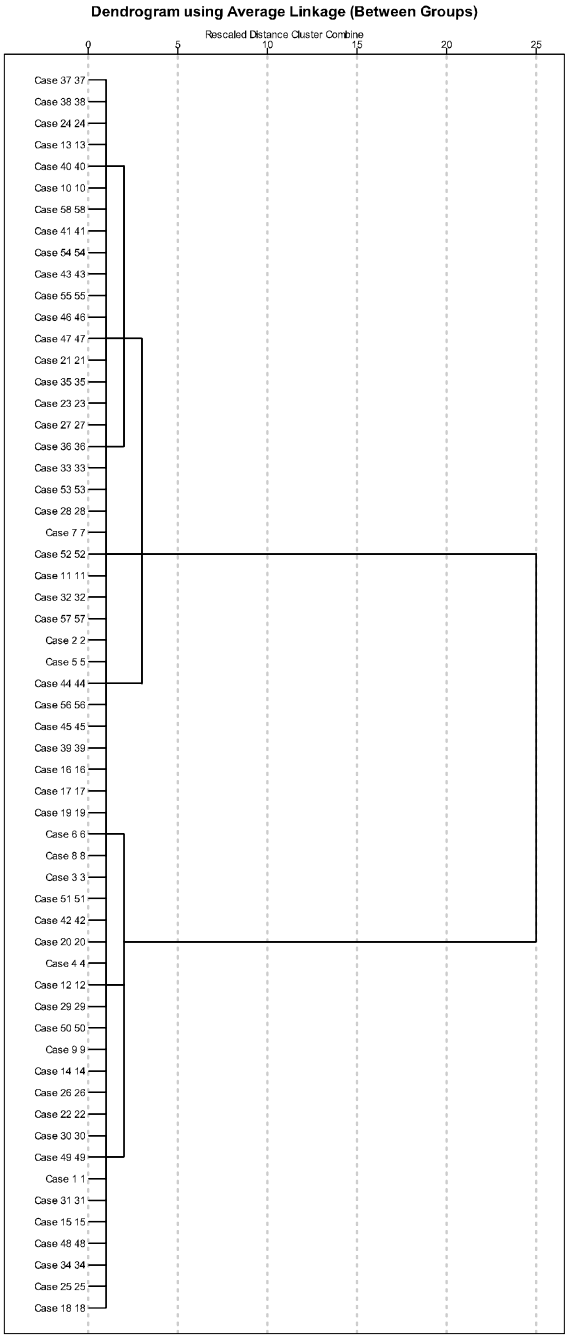

Fig. 4: Dendogram of the location of 58 cases using single linkage.

(component 4) and K (component 5). The single dominance of variables in each factor indicates non-mixing or partial mixing of different types of water.

Cluster Analysis (CA)

Cluster analysis comprises of a series of multivariate methods which are used to find true groups of data or stations. In clustering, the objects are grouped such a way that similar objects fall into the same class (Danielsson et al. 1999). The hierarchical method of cluster analysis, which is used in this study, has the advantage of not demanding any prior knowledge of the number of clusters, which the non-hierarchical method does. A review by Sharma suggests Ward's clustering procedure to be the best, because it yields a larger proportion of correct classified observations than do most other methods (Sharma 1996). Hence, Ward's clustering procedure is used in this study. As a distance measure, the squared euclidean distance was used, which is one of the most commonly adopted measures (Fovell & Fovell 1993). In this method the distance between the clusters was determined by the distance of the two closest objects (nearest neighbour) in the different clusters.

All the 58 samples were subjected to HCA in which 3 major clusters were formed and the output of this cluster analysis is given as a dendogram in Fig. 4. In cluster one, all the surface waters have been grouped into one implying that they possess a similar kind of a quality, while the groundwaters have been clustered in two different groups of 2 and 3. The second cluster includes groundwaters grouped in the southeast and northwestern part of the study area, while the third cluster includes the groundwaters which have been grouped together in the central part of the study area. This implies that the southwestern and

northeastern groundwater samples of the study area have a similar quality which slightly differs from the ground waters clustered in the central part. These differences are attributed to the lithology and interaction of groundwater with varied rocks.

CONCLUSION

The data for both surface and groundwaters were analysed for PCA. In Surface water 4 components were extracted and contributed to the 84.46% of the total variance. The most significance feature observed was that the most dominating factor controlling the water quality in surface water are TDS and harness which influence on the EC. Nitrification process was also another observation made which takes place in the lakes and reservoirs. In the groundwater 5 components were extracted and included 94.4 % of the total variance. This component shows high loading of TDS which implies that the hydrogeochemical process contribute to enrich and salinize the water mainly by Na, NO_3 and SO_4. By comparing the first component loadings between the surface and groundwaters, it is understood that the elements which contribute to the electrical conductivity of the water differ from each other and this implies the quality variation between the surface and ground waters. The non-mixing or partial mixing of different types of groundwater as deduced by the PCA indicates slow movement of groundwater or the absence of interconnected underground fractures. Dendogram for 58 cases were plotted which grouped them into 3 clusters. All the surface waters were clustered into one group, whereas the groundwater samples were clustered into 2 groups. This is mainly due to the varied lithology and rock-water interaction. This study also illustrates the utility of multivariate statistical analyses in hydrogeochemical studies.

REFERENCES

APHA 1995. Standard Methods for Analysis of Water and Wastewater. American Public Health Association, 14th ed., Washington DC, 1457 pp.

Bernard Parinet, Antoine Lhote and Bernard Legube 2004. Principal component analysis: An appropriate tool for water quality evaluation and management-application to a tropical lake system. Ecological Modelling, pp. 295-311.

Cattell, R.B. 1966. The Scree test for number of factors. Multivariate Behavorial Research I, pp. 245-276.

Chatfield, C. and Collins, A.J. 1980. Introduction to Multivariate Analysis. Chapman and Hall in Association with Methuen, Inc., New York.

Danielsson, A., Cato, I., Carman, R. and Rahm, L. 1999. Spatial clustering of metals in the sediments of the Skagerrak/ Kattegat. Appl Geochem., 14: 689-706.

Debasis, Deb, E., Vinayak, N., Deshpande, E. and Kamal Ch. Das 2008. Assessment of water quality around surface coal mines using principal component analysis and fuzzy reasoning techniques. Mine Water Environ., 27: 183-193.

Dennis Child 2006. The Essentials of Factor Analysis. 3rd Edition, Continuum International Publishing Group.

Fovell, R. and Fovell, M.Y. 1993. Climate zones of the conterminous United States defined using cluster analysis. J. Climate, 6: 2103-35.

Geological Society of India, 2006. Annual Report. Bangalore.

Hope, K. 1986. Methods of Multivariate Analysis. University of London Press Ltd., London.

Karpuzcu, M. and Sene, S. 1987. Design of Monitoring Systems for Water Quality by Principal Component Analysis and a Case Study. In: Proceedings of the International Conference.

Nicolaos Lambrakis, Andreas Antonakos and George Panagopoulos 2004. The use of multicomponent statistical analysis in hydrogeological environmental research. Water Research, 1862-1872.

Pathak, J.K., Mohd. Alam and Shikha Sharma 2008. Interpretation of groundwater quality using multivariate statistical technique in Moradabad city, Western Uttar Pradesh State, India. E-Journal of Chemistry, 5(3): 607-619.

Rajesh Reghunath, T.R. Sreedhara Murthy and Raghavan, B.R. 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India. Water Research, 2437-2442.

Rao, N.S., Devedas, D. J. and Rao, K.V.S. 2006. Environmental Geosciences, 13: 239-259.

Sharma, S. 1996. Applied Multivariate Techniques. Wiley, New York.