Nature Environment and Pollution Technology An International Quarterly Scientific Journal	Vol. 8	No. 4	pp. 693-700	2009

Original Research Paper

Hydrogeochemical Studies by Multivariate Statistical Analysis in Upper Thirumanimuthar Sub-basin, Cauvery River, Tamil Nadu, India

M. Suresh, B. Gurugnanam, S. Vasudevan*, B. Rajeshkanna*, K. Dharanirajan and N. Prabhakaran**

Geographic Information Technology Lab, Department of Earth Sciences, Annamalai University, Annamalai Nagar, Tamil Nadu

*Centre for Geographic Information Technologies, Bharathidasan University, Trichy, Tamil Nadu

**Department of Ocean Studies, Pondicherry University, Andaman

Key Words:
Groundwater
Hydrogeochemical studies
Multivariate statistical analysis
Factor analysis
Cluster analysis

ABSTRACT

In the present paper deals with the study of hydrogeochemistry of groundwater by multivariate statistical techniques such as factor and cluster analyses. The upper Thirumanimuthar sub-basin, Cauvery River, hard rock terrain in Salem District covering an area of about 346.40 km^2 has been selected for the study. Fifty one samples were collected during premonsoon season 2007 and analysed for various water quality parameters like pH, EC, TDS, Ca, Mg, Na, K, HCO_3, CO_3, SO_4, Cl and TH. Hydrogeochemical data of 51 groundwater samples were subjected to Q- and R- mode factor and cluster analysis. R-mode analysis reveals the interrelations among the variables studied and the Q-mode analysis reveals the interrelations among the samples studied. The R-mode factor analysis shows that Na and Cl with HCO_3 account for most of the electrical conductivity and total dissolved solids of the groundwater. The 'single dominance' nature of the majority of the factors in the R-mode analysis indicates non-mixing or partial mixing of different types of groundwaters. Both Q-mode factor and R-mode cluster analyses show that there is an exchange between the river water and adjacent groundwater. Cluster classification map reveals that 97.79% of the study area comes under cluster I classification.

INTRODUCTION

The study area is located in the upper Thirumanimuthar sub-basin, Cauvery River, hard rock terrain in Salem district of Tamilnadu (Fig. 1). The chemistry of groundwater is an important factor determining its use for domestic, irrigation or industrial purposes. The quality of groundwater is controlled by several factors, including climate, soil characteristics, manner of circulation of groundwater through the rock types, topography of the area, human activities on the ground, etc. Apart from these factors, charnockite, fissile hornblende-biotite gneiss and contact between them play an important role in determining the quality of groundwater.

In this study, such a situation has been deduced by using multivariate statistical techniques such as factor and cluster analyses. Here, a qualitative study has been attempted to major cations and anions interaction in the groundwater. Multivariate statistical analysis has been successfully applied in a number of hydrogeochemical studies. Steinhorst & Williams (1985) used multivariate statistical analysis of water chemistry data in two field studies to identify groundwater sources. In the application of multivariate analysis to chemical data, Usunoff & Guzma´n-Guzma´n (1989) demonstrated the usefulness of the approach in hydrogeochemical investigations when considering the geological and hydrogeological knowledge of the aquifer.

Multivariate analyses, such as cluster and factor, aim to interpret the governing processes through data reduction and classification, and are widely applied mainly to spatial data in geochemistry (Papatheodorou et al. 1999), hydrochemistry (Voudouris et al. 2000), mineralogy (Seymour et al. 2004) and even in marine geophysics (Papatheodorou et al. 2002). 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, Helena et al. 2000, Lambrakis et al. 2004). Multivariate treatment of environmental data is widely used to characterize and evaluate surface waters (Reisenhofer et al. 1995, Miller et al. 1997, De Ceballos et al. 1998, Momen et al. 1999, Perona et al. 1999, Lau & Lane 2002, Simeonov et al. 2003, Yu et al. 2003) and groundwater quality (Vengosh & Keren 1996, Suk & Lee 1999, Panagopoulos et al. 2004, Vincent Cloutier et al. 2008) and it is useful for evidencing temporal and spatial variations caused by natural and human factors linked to seasonality.

STUDY AREA

The upper Thirumanimuthar sub-basin of Central Tamilnadu has been selected for the present investigation. The study area lies between latitudes 11°31’57” N to 11°48’05” N and longitudes 78°02’33” E to 78°21’13” E covering an area of 442.78 km^2. In these, plain area covers an area of 346.40 km^2 (Fig. 1). The major source for recharge of water in this area is rainfall during monsoon season. The average annual rainfall is 852 mm (1998 to 2007). The area under study is lying in the Archaean crystalline rock exposures, surrounded by hills with the Shevaroys (1033m) and Nagaramalai (619 m) on north, Jarugumalai (583 m) on the south, Kanjamalai (883 m) on the west, and Goudamalai (568 m) on the east.

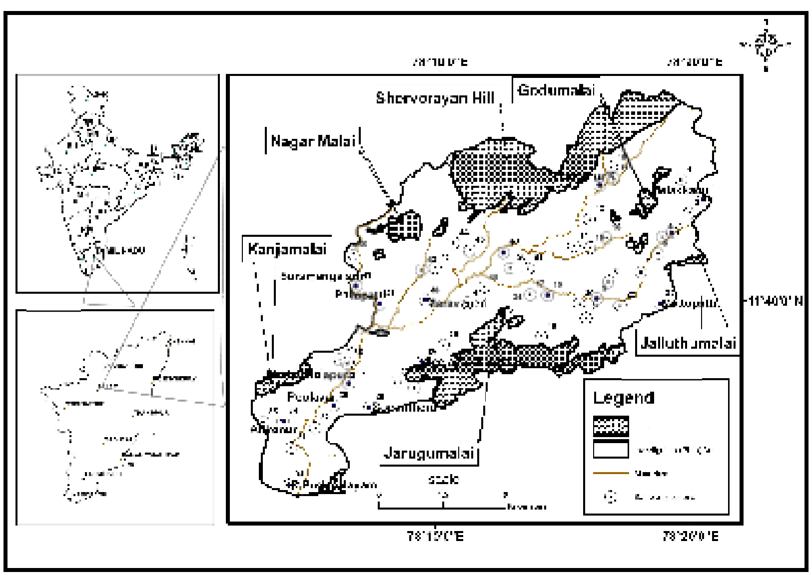

Fig.1: Study area of upper Thirumanimuthar sub-basin and sample locations.

MATERIALS AND METHODS

Geochemistry

Fifty one groundwater samples (bore wells) were collected during the pre-monsoon period (May) of the year 2007. Fig. 1 shows the locations of the groundwater samples. The samples were analysed by standard water analysis methods (Trivedy & Goel 1986, APHA 1995). The ionic constituents Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, CO_3^{2-}, HCO_3^{-}, SO_4^{2-} and the non-ionic constituents pH, electrical conductivity (EC), total dissolved solids (TDS) and total hardness (TH) were determined for these groundwaters. These data were subjected to multivariate analytical techniques such as factor and cluster analysis. Multivariate techniques can help to simplify and organize large data sets and to make useful generalizations that can lead to meaningful insight (Laaksoharju et al. 1999). Cluster and factor analyses are efficient ways of displaying complex relationships among many objects (Davis 1973). The two methods in cluster and factor analyses, i.e., Q- and R- mode analyses have been done for the data generated. R-mode analysis reveals the interrelations among the variables studied and the Q-mode analysis reveals the interrelation among the samples studied. The STATISTICA software has been used to carry out the analysis. The data have been standardized by using standard statistical procedures.

Statistical Analysis

Factor analysis (FA): The factors are constructed in a way that reduces the overall complexity of the data by taking advantage of inherent inter-dependencies. 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. The interpretation is based on rotated factors, rotated loadings and rotated Eigen values.

Hierarchical cluster analysis (HCA): Cluster analysis comprises a series of multivariate methods which are used to find true groups of data. In clustering, the objects are grouped such that similar objects fall into the same class (Danielsson et al. 1999). Hierarchical cluster analysis is the most widely applied techniques in the earth sciences and is used in this study. Hierarchical clustering joins the most similar observations, and then successively the next most similar observations. The levels of similarity at which observations are merged are used to construct a dendrogram. In this study, a standardized space Euclidian distance (Davis 1986) is

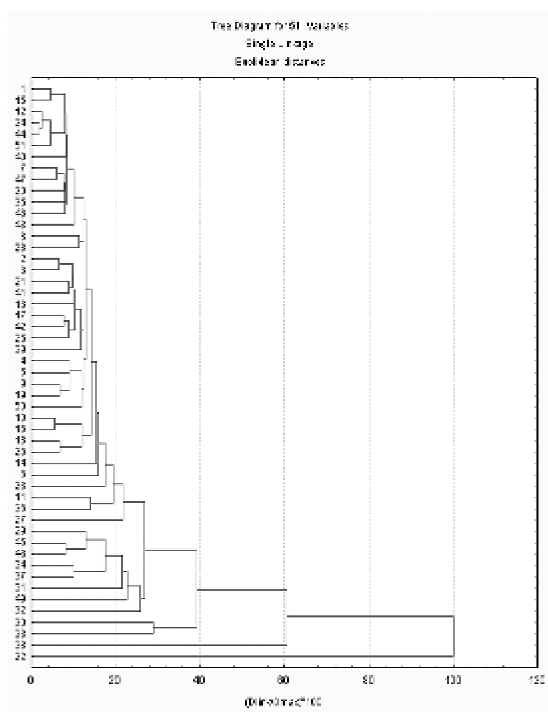

Fig. 2: Dendogram of the hierarchical cluster analysis using the Ward method.

used. A low distance shows that the two objects are similar or “close together”, whereas a large distance indicates dissimilarity.

RESULTS AND DISCUSSION

R-mode factor analysis: R-mode factor analysis for the cations and anions, TDS, EC, pH and TH have been considered for the present study. The analysis generated 8 factors which together account for 99.9% of variance. The varimax raw loadings, eigen values, percentage of variance and cumulative percentage of variance of all the 8 factors are given in Table 1.

The first eigen value is 6.00 which accounts for 49.97% of the total variance and this constitutes the first and main factor. The second and third eigen values are 2.77 and 1.22, which account for 23.05% and 10.17% respectively of the total variance. The rest of the eigen values each constitute less than 10% of the total variance. The first factor, which accounts for 49.97% of the total variance, is characterized by very high loadings of Na, Cl and TDS and moderate to high loadings of HCO_3 and EC. This factor reveals that the TDS and EC in the study area are mainly due to Na and Cl, though bicarbonate also plays a substantial role in determining EC and TDS. The second factor, which accounts for 23.05% of the total variance, is mainly associated with very high loading of EC and hardness and also with moderate loading of bicarbonate. This factor accounts for the temporary hardness of water. The loading of bicarbonate in this factor is lower than the first factor.

Factors 3 and 4 are characterized by dominance of only two variables each, such as Ca is with very high loading and TH is also with moderate loading of bicarbonate (factor 3), and very high loadings of Mg and moderate to high loadings of TH (factor 4). These factors reveal that TH in the study area is mainly due to Ca and Mg, which also play a substantial role in determining TH and together these factors account for 21.80 of the total variance.

The remaining factors (from 5 to 8) are characterized by the dominance of only one variable each, such as SO_4 (factor 5), HCO_3 (factor 6), K (factor 7), EC (factor 8) and together these factors account

Table 1: R-mode factor analysis with varimax normalized rotation.

Variable	Factor-1	Factor-2	Factor-3	Factor-4	Factor-5	Factor-6	Factor-7	Factor-8
Ca	0.252	0.005	0.932	0.115	0.201	0.114	-0.015	0.032
Mg	0.234	-0.118	0.127	0.920	0.205	0.155	0.026	0.032
Na	0.960	-0.057	-0.030	-0.029	0.098	0.249	-0.002	-0.047
K	0.122	-0.869	-0.058	0.175	0.102	0.093	0.420	0.018
HCO_3	0.445	0.136	0.199	0.241	0.037	0.827	0.020	0.024
CO_3	-0.090	-0.985	-0.021	-0.036	-0.043	-0.094	-0.094	-0.024
SO_4	0.233	0.010	0.267	0.255	0.898	0.032	0.018	0.024
Cl	0.861	0.036	0.355	0.346	0.092	0.035	0.034	-0.018
pH	-0.094	-0.985	-0.026	-0.037	-0.049	-0.093	-0.090	-0.023
EC	0.770	0.156	0.308	0.260	0.162	0.116	0.018	0.425
TDS	0.803	0.117	0.348	0.314	0.238	0.250	0.043	-0.004
TH	0.308	0.172	0.629	0.618	0.250	0.181	0.029	0.044
Eigen value	5.996	2.766	1.221	0.704	0.582	0.424	0.167	0.139
Percentage of variance	49.970	23.051	10.173	5.870	4.854	3.531	1.389	1.161
Cumulative Percentage	49.970	73.021	83.194	89.064	93.917	97.448	98.837	99.998

for 10.92% of the total variance. The single dominance of variables in each factor indicates non-mixing or partial mixing of different types of waters.

Q-Mode factor analysis: The rotated loadings, eigen values, percentage of variance and cumulative percentage of variance of the factors are given in Table 2. The Q-mode factor analysis of the 51 groundwater samples has generated four factors which together account for 99.83% of the variance. The first three factors, which constitute for 99.83% of the variance, are considered as representative of the factor model and have been taken for interpretation.

The first factor which accounts for 97.8% of the variance consists of high loadings of samples 1-21, 23-29, 31-32 and 34-37. The second factor, which accounted for 0.96% of the variance, consists of high loadings of samples 30 and 33. Factor 3, which accounts for 0.57% of the variance, consists of high loadings of sample 38. On the other hand, groundwater samples from one location i.e., 22 has high loadings in the fourth factor accounting for 0.43% of the variance. The distribution of wells are well explained by factors 2, 3 and 4, which do not conform to any kind of spatial pattern. However, the majority of the samples within factor 1 fall on either side of the main course of the river system. This strongly suggests that there is an exchange between the river water and adjacent groundwater. It is also discussed by Reghunath et al. (2002). However, the majority of the samples within factor 1 fall on rock interaction of the groundwater.

Table 2: Q-mode factor analysis with varimax normalized rotation.

S.No.	Factor-1	Factor-2	Factor-3	Factor-4	S.No.	Factor-1	Factor-2	Factor-3	Factor-4
1	0.546	0.524	0.598	0.258	27	0.590	0.566	0.508	0.268
2	0.536	0.598	0.528	0.277	28	0.415	0.555	0.669	0.270
3	0.535	0.594	0.532	0.279	29	0.504	0.618	0.536	0.276
4	0.473	0.511	0.669	0.252	30	0.393	0.736	0.472	0.283
5	0.550	0.513	0.594	0.280	31	0.518	0.609	0.530	0.278
6	0.495	0.657	0.494	0.279	32	0.553	0.505	0.603	0.268
7	0.441	0.569	0.642	0.254	33	0.458	0.716	0.448	0.279
8	0.536	0.536	0.592	0.274	34	0.377	0.669	0.574	0.284
9	0.500	0.522	0.634	0.272	35	0.328	0.526	0.734	0.270
10	0.474	0.550	0.623	0.274	36	0.382	0.720	0.506	0.281
11	0.466	0.675	0.495	0.284	37	0.418	0.634	0.588	0.277
12	0.563	0.456	0.632	0.274	38	0.527	0.649	0.472	0.274
13	0.492	0.529	0.637	0.269	39	0.492	0.513	0.648	0.272
14	0.368	0.626	0.633	0.263	40	0.422	0.445	0.740	0.272
15	0.513	0.548	0.604	0.262	41	0.542	0.548	0.570	0.279
16	0.525	0.458	0.660	0.272	42	0.577	0.511	0.577	0.267
17	0.547	0.521	0.596	0.271	43	0.525	0.444	0.679	0.253
18	0.503	0.510	0.644	0.267	44	0.547	0.491	0.616	0.279
19	0.529	0.531	0.606	0.262	45	0.495	0.653	0.501	0.278
20	0.369	0.591	0.663	0.273	46	0.427	0.477	0.722	0.259
21	0.512	0.551	0.600	0.273	47	0.437	0.571	0.636	0.278
22	0.334	0.486	0.389	0.707	48	0.494	0.638	0.522	0.277
23	0.406	0.674	0.551	0.275	49	0.604	0.531	0.530	0.265
24	0.538	0.486	0.632	0.272	50	0.449	0.477	0.710	0.250
25	0.519	0.544	0.598	0.274	51	0.473	0.513	0.660	0.279
26	0.654	0.481	0.518	0.266					
Eigen value						49.919	0.487	0.289	0.218
Percentage of variance						97.880	0.955	0.566	0.427
Cumulative Percentage						97.880	98.836	99.402	99.829

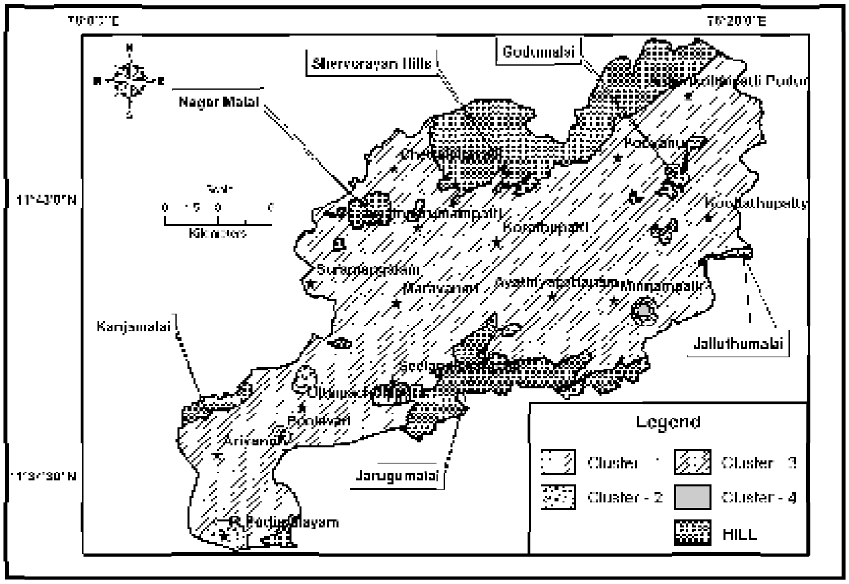

Fig. 3: Cluster classification spatial distribution map.

Hierarchical cluster analysis (HCA): 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 that similar objects fall into the same class (Danielsson et al. 1999). The HCA is a data classification technique. There are different clustering techniques, but the hierarchical clustering is most widely applied in earth sciences (Davis 1986) and often used in the classification of hydrogeochemical data (Steinhorst & Williams 1985, Schot & Van der Wal 1992, Ribeiro & Macedo 1995, Gu¨ler et al. 2002). The result of the hierarchical cluster analysis was given in the form of a dendrogram (Fig. 2). For this, the Euclidean distance was chosen as the distance measure, or similarity measurement between sampling sites. The sampling sites with the larger similarity are first grouped. Next, groups of samples are joined with a linkage rule, and the steps are repeated until all the observations have been classified. With this geochemical data set, Ward's method was more successful to form clusters that are more or less homogenous and geochemically distinct from other clusters, compared to other methods such as the weighted pair-group average. Ward's method is distinct from other linkage rules because it uses an analysis of variance approach to evaluate the distances between clusters (Stat Soft Inc. 2004). Other studies used Ward's method as linkage rule in their cluster analysis (Adar et al. 1992, Schot & Van der Wal 1992). Gu¨ler et al. (2002) also found that using the Euclidean distance as a distance measure and Ward's method as a linkage rule produced the most distinctive group.

Q-mode cluster analysis: The output of the Q-mode cluster analysis with four major clusters, is given as a dendrogram (Fig. 2). Clusters 1, 2, 3 and 4 correspond to the factors 1, 2, 3 and 4, respectively. The similarity of the Q-mode cluster analysis to the Q-mode factor analysis confirms the interpretations made using the Q-mode factor analysis. To understand the spatial distribution of various clusters class in the study area, the results were taken into GIS platform wherein spatial distribution map is prepared (Fig 3). The result of spatial distribution map is given in Table 3.

Table 3: Spatial distribution map results of cluster classification.

Cluster classification	Area in km^2	Area in %
Cluster - I	338.76	97.79
Cluster - II	6.23	1.80
Cluster - III	0.62	0.18
Cluster - IV	0.41	0.12

CONCLUSION

The non-mixing or partial mixing of different types of groundwaters as deduced by the R-mode factor analysis indicates slow movement of groundwater or the absence of interconnected underground fractures. The Q-mode factor and cluster analyses indicate that exchange between the river water and the groundwater plays a dominant role in the hydrochemical evolution of groundwater. Cluster classification map reveals that 97.79% of the study area comes under cluster I classification.

ACKNOWLEDGEMENT

The authors are thankful to DST-NRDMS Division, Government of India for the financial support to execute the work. Thanks are also to GSI, for using the geological map of the study area.

REFERENCES

Adar, E.M., Rosenthal, E., Issar, A.S. and Batelaan, O. 1992. Quantitative assessment of the flow pattern in the southern Arava valley (Israel) by environmental tracers and a mixing cell model. Journal of Hydrology, 136: 333-352.

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

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

Davis, J.C. 1973. Statistics and Data Analysis in Geology. Wiley, New York, 550 pp.

Davis, J.C. 1986. Statistics and Data Analysis in Geology. John Wiley & Sons Inc., New York.

De Ceballos, B.S.O., Koning, A. and De Olivera, J.F. 1998. Dam reservoir eutrophication: A simplified technique for a fast diagnosis of an environmental degradation. Water Res., 32(11): 3477-3483.

Gu¨ler, C., Thyne, G.D., McCray, J.E. and Turner, A.K. 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeology Journal, 10: 455-474.

Helena, B., Pardo, B., Vega, M., Barrado, E., Fernandez, J.M. and Fernandez, L. 2000. Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by rincipal component analysis. Water Research, 34(3): 807-816.

Laaksoharju, M., Skarman, C. and Skarman, E. 1999. Multivariate mixing and mass balance (M3) calculation, a new tool for decoding hydrogeochemical information. Appl Geochem., 14: 861-71.

Lambrakis, N., Antonakos, A. and Panagopoulos, G. 2004. The use of multicomponent statistical analysis in hydrogeological environmental research. Water Res., 38; 1862-1872.

Lau, S.S.S. and Lane, S.N. 2002. Biological and chemical factors influencing shallow lake eutrophication: A long-term study. Sci. Total Environ., 288: 167-181.

Miller, C.V., Denis, J.M., Ator, S.W. and Brakebill, G.W. 1997. Nutrients in streams during base flow in selected environmental settings of the Potomac River basin. J. Am. Water Resour. Assoc., 33: 1155-1171.

Momen, B., Zehr, G.P., Boylen, C.W. and Sutherland, J.W. 1999. Determinants of summer nitrate concentration in a set of Adirondack lakes, New York. Water Air Soil Pollut., 111: 19-28.

Panagopoulos, G., Lamprakis, N., Tsolis-Katagas, P. and Papoulis, D. 2004. Cation exchange processes and human activities in incon-fined aquifers. Environ. Geol., 46: 542-552.

Papatheodorou, G., Hotos, G., Geraga, M., Avramidou, D. and Vorinakis, T. 2002a. Heavy metal concentrations in sediments of Klisova lagoon (S.E. Mesolonghi–Aitolikon Lagoon complex) W Greece. Fresen. Environ. Bull., 11(11): 951-956.

Papatheodorou, G., Lyberis, E. and Ferentinos, G. 1999. Use of factor analysis to study the distribution of metalliferous bauxitic tailings in the seabed of the Gulf of Corinth. Greece. Natural Resour. Res., 8(4): 277-285.

Perona, E., Bonilla, I. and Mateo, P. 1999. Spatial and temporal changes in water quality in a Spanish river. Sci Total Environ., 241: 75-90.

Reghunath, R., Sreedhara, M.T.R. and Raghavan, B.R. 2002. The utility of multivariate statistical techniques in hydrogeochemical studies: An example from Karnataka, India. Water Research, 36(10): 2437-2442.

Reisenhofer, E., Picciotto, A. and Li, D. 1995. A factor analysis approach to the study of the eutrophication of a shallow, temperate lake (San Daniele, North Eastern Italy). Anal. Chim. Acta, 306: 99-106.

Ribeiro, L. and Macedo, M.E. 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.

Schot, P.P. and Van der Wal, J. 1992. Human impact on regional groundwater composition through intervention in natural flow patterns and changes in land use. Journal of Hydrology, 134: 297-313.

Seymour, S.K., Christanis, K., Bouzinos, A., Papazisimou, S., Papatheodorou, G., Moran, E. and D´en‘es, G. 2004. Tephrostratigraphy and tephrochronology in the Philippi peat basin, Macedonia, Northern Hellas (Greece). Quarter. Int., 121: 53-65.

Simeonov, V., Stratis, J.A., Samara, C., Zachariadis, G., Voutsa, D., Anthemidis, A., Sofoniou, M. and Kouimtzis, T. 2003. Assessment of the surface water quality in Northern Greece. Water Res., 37: 4119-4124.

Steinhorst, R.K. and Williams, R.E. 1985. Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis. Water Resources Research, 21: 1149-1156.

Stat Soft Inc. 2004. STATISTICA (Data Analysis Software System), Version 6.

Suk, H. and Lee, K. 1999. Characterization of a ground water hydrochemical system through multivariate analysis: Clustering into ground water zones. Ground Water, 37(3): 358-366.

Trivedy, R.K., Goel, P.K. 1986. Chemical and Biological Methods for Water Pollution Studies. Environmental Publications, Karad, 248pp.

Usunoff, E.J. and Guzma´n-Guzma´n, A. 1989. Multivariate analysis in hydrochemistry: An example of the use of factor and correspondence analyses. Ground Water, 27: 27-34.

Vega, M., Pardo, R., Barrado, E. and Deban, L. 1998. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. Water Research, 32: 3581-3592.

Vengosh, A., Keren, R. 1996. Chemical modifications of groundwater contaminated by recharge of treated sewage effluent. Contam. Hydrol., 23: 347-360.

Vincent Cloutier, Rene´ Lefebvre, Rene´ Therrien and Martine M. Savard 2008. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. Journal of Hydrology, 353: 294-313.

Voudouris, K., Panagopolous, A. and Koumanatakis, J. 2000. Multivariate statistical analysis in the assessment of hydrochemistry of the Northern Korinthia prefecture alluvial system Peloponnese, Greece. Natural Resources Research, 9(2): 135-146.

Yu, S., Shang, J., Zhao, J. and Guo, H. 2003. Factor analysis and dynamics of water quality of the Songhua River, Northeast China. Water Air Soil Pollut., 144: 159-169.