PERGAMON

Water Research 36 (2002) 2437–2442

WATER RESEARCH

www.elsevier.com/locate/watres

The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India

Rajesh Reghunath*, T.R. Sreedhara Murthy, B.R. Raghavan

Department of Marine Geology, Mangalore University, Mangalagangotri 574 199, India

Received 1 November 1999; received in revised form 1 June 2000; accepted 7 November 2001

Abstract

Quality of groundwater is controlled by many factors amongst which the interaction of river water with adjacent groundwater and mixing/non-mixing of different types of groundwater may be important. An attempt has been made to study these processes using multivariate statistical techniques such as factor and cluster analyses. The Nethravathi catchment (India) which is a tropical river basin draining the Precambrian crystalline province of peninsular India, has been selected for this study. Hydrogeochemical data for 56 groundwater samples were subjected to Q- and R- mode factor and cluster analysis. R-mode analysis reveals the inter-relations among the variables studied and the Q-mode analysis reveals the inter-relations among the samples studied. The R-mode factor analysis shows that Na and CI 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 groundwater. Both Q-mode factor and Q-mode cluster analyses shows that there is an exchange between the river water and the adjacent groundwater.

Keywords: Factor analysis; Cluster analysis; Hydrogeochemistry; Nethravathi river basin

1. Introduction

The chemistry of water 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, saline water intrusion in coastal areas, human activities on the ground, etc. Apart from these factors, the interaction between the river water and the adjacent groundwater and the mixing\ non-mixing of different types of groundwater may also play important roles in determining the quality of the 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 trace the interaction between the groundwater and adjacent river water by sampling only the groundwater. The already available methods such as NETPATH algorithm developed by the US Geological Survey, M3 calculations [1] mainly trace the rock–water interaction and not the groundwater–river water interactions. This study also illustrates the usefulness of statistical analysis to improve the understanding of groundwater systems.

2. Study area

The Nethravathi river basin, in southern India has been selected for this study. The Nethravathi river basin is an 8th order tropical river basin which spreads over 4256.80 km^2 in five districts of Karnataka state, India (Fig. 1). The present study area falls in the Precambrian crystalline province of peninsular

*Corresponding author. Present address: Hydrologist, AHADS, Agali 678 581, India.

0043-1354/02/$ - see front matter
PII: S0043-1354(01)00490-0

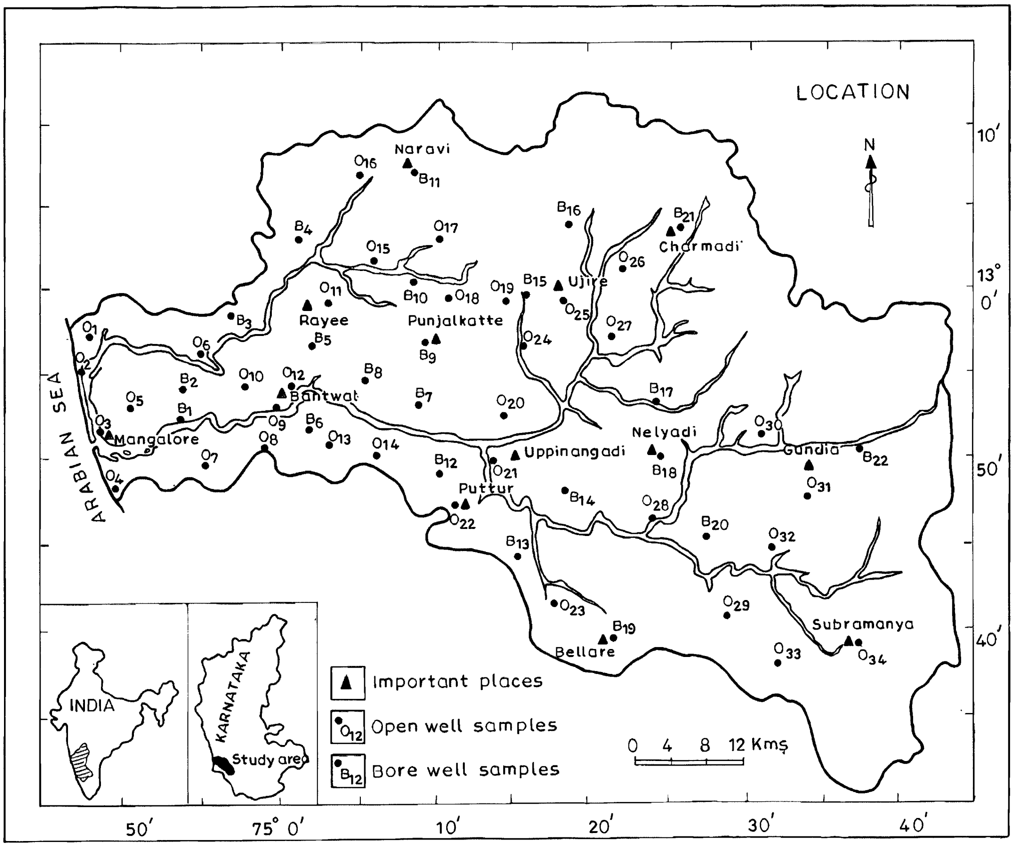

Fig. 1. Study area with the locations of groundwater samples.

India, which covers nearly two-thirds of the country. The Nethravathi river basin is underlain by various rock formations, which include the peninsular gneissic complex, charnockites, dolerite dykes, laterites, sand and gravel deposits of fluvial origin [2,3]. The gneissic rocks cover the major portion of the basin, with charnockites occuring locally as small linear patches. Physiographically, the Nethravathi river basin is broadly divisible into three zones namely the coastal tract, the midlands and the hilly tracts of Western Ghats [4]. The coastal tract in the west is a highly dissected landmass. The midlands are a rolling terrain of mounds and hillocks.

3. Materials and methods

Following the methods prescribed by Palmquist [5], 56 groundwater samples (from 34 open wells and 22 bore wells) were collected during the pre-monsoon period (May) of the year 1998. Fig. 1 shows the locations of the groundwater samples (sample numbers O1–O34 represent open well samples and sample numbers B1–B22 represent bore well samples). The samples were analysed using standard water analysis methods [6,7]. The ionic constituents Ca^{2+}, Mg^{2+}, Na^{+}, K^{+}, Cl^{-}, CO_3^{2-}, HCO_3^{-}, SO_4^{2-} and NO_3^{-} and the non-ionic constituents pH, electrical conductivity (EC), total dissolved solids (TDS) and hardness 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 organise large data sets and to make useful generalisations that can lead to meaningful insight [1]. Cluster and factor analyses are efficient ways of displaying complex relationships among many objects [8]. 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 interaction among the variables studied and the Q-mode analysis reveals the inter-relation among the samples studied. The computer package Statistical Package for Social Sciences (SPSS) has been used to carry out the analysis. The data have been standardised by using standard statistical procedures.

4. Statistical analyses

4.1. Factor analysis

The usual procedures of interpretation of chemical quality of groundwater with the help of plots of different ions and pairs of ions do not define simultaneously the similarities or otherwise between all ions or samples [9]. Factor analysis offers a powerful means of detecting such similarities among the variables or samples. The purpose of factor analysis is to interpret the structure within the variance–covariance matrix of a multivariate data collection. The technique which it uses is extraction of the eigen values and eigen vectors from the matrix of correlations or covariances [8]. Thus, factor analysis is a multivariate technique designed to analyse the inter-relationships within a set of variables or objects. 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.

4.2. R-mode factor analysis

R-mode factor analysis of different chemical constituents of the groundwater of Nethravathi river basin has been carried out. All cations and anions, TDS, EC, pH and hardness have been considered for the present analysis. The analysis generated 12 factors which together account for 99.9% of variance. The rotated loadings, eigen values, percentage of variance and cumulative percentage of variance of all 12 factors are given in Table 1.

The first eigen value is 5.83 which accounts for 44.9% of the total variance and this constitutes the first and main factor. The second and third eigen values are 2.33 and 1.67 and these account for 18% and 12.9%, 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 44.9% of the total variance) is characterised by very high loadings of Na, Cl and EC and moderate to high loadings of bicarbonate and TDS. This factor reveals that the EC and TDS in the study area are mainly due to Na and CI, though bicarbonate also plays a substantial role in determining EC and TDS. The second factor (which accounts for 18% of the total variance) is mainly associated with very high loading of Ca and hardness and also with moderate loading of bicarbonate. This factor accounts for the temporary hardness of the water. The loading of bicarbonate in this factor is lower than it is in the first factor.

Factors 3–8 are characterised by the dominance of only one variable each, such as Mg (factor 3), K (factor 4) NO_3 (factor 5), CO_3 (factor 6), pH (factor 7) and SO_4 (factor 8) and together these six factors account for 34.7% of the total variance. The single dominance of variables in each factor indicates non-mixing or partial mixing of different types of water. The 7th factor, with high loading only on pH, possibly implies biogenic or organic controls on the pH value of water or the major

Table 1
Varimax rotated R-mode factor loading matrix

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9	Factor 10	Factor 11	Factor 12
Ca	0.073	0.892	0.093	0.180	0.078	0.317	0.189	0.052	0.014	0.003	0.096	0.000
Mg	0.071	0.184	0.975	0.042	0.025	0.011	0.066	0.041	0.010	0.010	0.001	0.000
Na	0.980	0.041	0.071	0.014	0.014	0.022	0.096	0.120	0.014	0.065	0.013	0.030
K	0.149	0.324	0.063	0.879	0.228	0.069	0.005	0.183	0.052	0.015	0.000	0.000
HCO_3	0.606	0.532	0.379	0.163	0.074	0.072	0.193	0.117	0.061	0.334	0.001	0.002
CO_3	0.024	0.193	0.015	0.055	0.018	0.974	0.117	0.043	0.005	0.005	0.001	0.006
CI	0.969	0.019	0.086	0.037.	0.077	0.081	0.126	0.083	0.044	0.081	0.025	0.075
NO_3	0.166	0.106	0.023	0.187	0.947	0.016	0.138	0.075	0.045	0.005	0.005	0.000
SO_4	0.521	0.162	0.092	0.276	0.128	0.078	0.006	0.769	0.024	0.013	0.006	0.000
TDS	0.749	0.188	0.006	0.218	0.238	0.029	0.112	0.062	0.523	0.018	0.010	0.001
Hardness	0.121	0.733	0.404	0.370	0.187	0.042	0.153	0.148	0.089	0.006	0.232	0.0002
EC	0.911	0.240	0.060	0.163	0.135	0.005	0.124	0.188	0.055	0.032	0.065	0.054
pH	0.332	0.281	0.109	0.011	0.184	0.175	0.855	0.003	0.032	0.015	0.005	0.000
Eigen value	5.834	2.337	1.679	1.155	0.618	0.462	0.300	0.281	0.174	0.082	0.054	0.007
Percentage of variance	44.9	18	12.9	8.9	4.8	3.6	2.3	2.2	1.3	0.6	0.4	0.1
Cummulative percentage	44.9	62.9	75.8	84.7	89.4	93.0	95.3	97.5	98.8	99.4	99.8	99.9

contribution towards pH is not from the ions analysed in this study.

The remaining factors (from 9 to 12) are characterised by low to very low loading of all variables and hence can be considered as irrelevant for describing the factor model of the groundwater chemistry of Nethravathi river basin. The first eight factors which explain 97.5% of the total variance of the data, could be considered as representative of the factor model.

4.3. Q-mode factor analysis

Q-mode factor analysis of the water samples of the Nethravathi river basin has been carried out. All the 56 samples were considered for this analysis. The analysis has generated six factors which together account for 99.9% of the variance. The first three factors (which constitute for 99.1% of the variance) are considered as representative of the factor model and have been taken for interpretation. The rotated loadings, eigen values, percentage of variance and cumulative percentage of variance of the first three factors are given in Table 2.

The first factor which accounted for 89.9% of the variance consist of high loadings of samples O1, O10–O12, O15–O18, O20, O23, O24, O26–O28, O32, O33, B2, B6–B10, B15, B17–B21. The second factor which accounted for 5.3% of the variance consists of high loadings of samples O3, O4, O6, O13, O14, O21, O22, O29, O34, B1, B3–B5, B12 and B13. Factor 3 which accounted for 3.9% of the variance consist of high loadings of samples O2, O5, O7, O8, O9, O19, O25, O30, O31, B11, B14, B16 and B22.

The distribution of wells explained by factors 2 and 3 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 (except in areas adjacent to the river mouth and extreme upper reaches of the river basin). This strongly suggests that there is an exchange between the river water and adjacent groundwater (effluent or influent in nature). The lowest value of majority of the variables (either in bore wells or in open wells) also fall in the first factor. The high variance of this factor (89.9% out of 99.1%) suggests that this exchange between river water and groundwater plays a dominant role in the hydrochemical evolution of groundwater in the Nethravathi river basin.

4.4. Cluster analysis

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 [10]. The hierarchical method of cluster analysis, which is used in this study, has the advantage of not demanding any

Table 2
Varimax rotated Q-mode factor loading matrix

Sample No.	Factor 1	Factor 2	Factor 3
01	0.737	0.540	0.402
02	0.282	0.110	0.903
03	0.541	0.665	0.511
04	0.551	0.673	0.490
05	0.452	0.546	0.690
06	0.103	0.751	0.650
07	0.544	0.176	0.812
08	0.125	0.569	0.810
09	0.428	0.562	0.702
010	0.859	0.384	0.278
011	0.719	0.573	0.382
012	0.661	0.529	0.524
013	0.437	0.755	0.486
014	0.244	0.789	0.558
015	0.767	0.502	0.382
016	0.646	0.583	0.485
017	0.713	0.457	0.513
018	0.662	0.531	0.524
019	0.611	0.314	0.711
020	0.631	0.502	0.580
021	0.600	0.635	0.484
022	0.390	0.892	0.211
023	0.612	0.499	0.602
024	0.672	0.414	0.602
025	0.438	0.304	0.833
026	0.716	0.303	0.621
027	0.651	0.542	0.529
028	0.889	0.319	0.326
029	0.398	0.848	0.345
030	0.525	0.434	0.723
031	0.575	0.367	0.721
032	0.670	0.452	0.580
033	0.787	0.470	0.396
034	0.614	0.738	0.274
B1	0.592	0.732	0.334
B2	0.815	0.350	0.458
B3	0.575	0.696	0.427
B4	0.531	0.787	0.301
B5	0.632	0.678	0.342
B6	0.732	0.528	0.428
B7	0.697	0.643	0.329
B8	0.856	0.350	0.379
B9	0.749	0.557	0.330
B10	0.770	0.365	0.487
B11	0.590	0.371	0.710
B12	0.387	0.904	0.169
B13	0.255	0.943	0.206
B14	0.590	0.453	0.653
B15	0.811	0.254	0.520
B16	0.565	0.493	0.653
B17	0.792	0.563	0.229
B18	0.673	0.583	0.448
B19	0.636	0.461	0.616
B20	0.756	0.405	0.507
B21	0.767	0.221	0.599
B22	0.505	0.389	0.760
Eigen value	50.34	2.94	2.18
Percentage of variance	89.9	5.3	3.9
Cummulative precentage	89.9	95.2	99.1

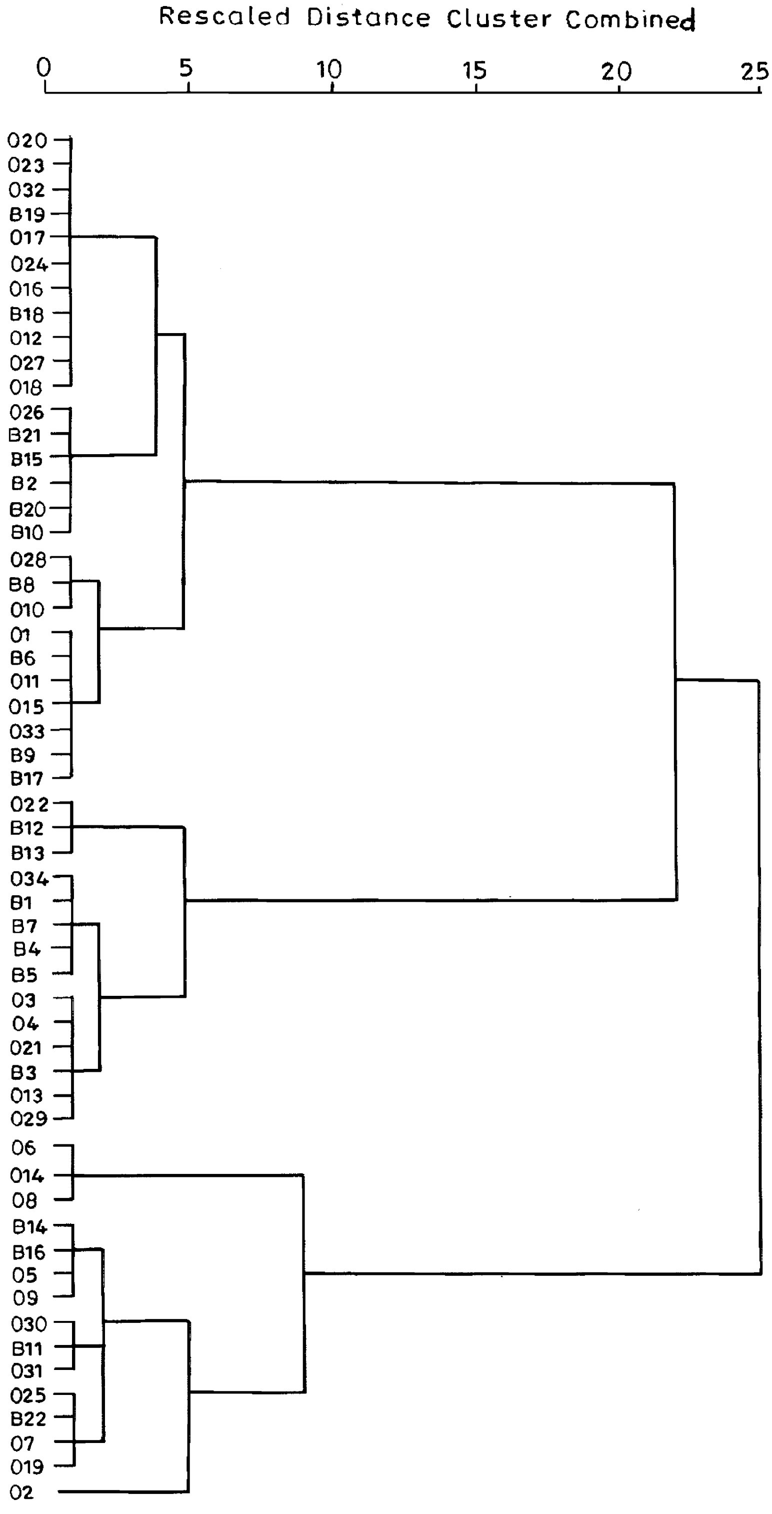

Fig. 2. Dendrogram of the Q-mode cluster analysis (The axis shown at the top indicates the relative similarity of different cluster groups. Lesser the distance, greater the similarity between objects).

prior knowledge of the number of clusters, which the non-hierarchical method does. A review by Sharma [11] suggests Ward's clustering procedure to be the best, because it yields a larger proportion of correct classified observations than do most other methods. 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 [12].

4.5. Q-mode cluster analysis

The output of the Q-mode cluster analysis is given as a dendrogram (Fig. 2). There are three major clusters as shown in Fig. 2. Clusters 1, 2 and 3 correspond to the factors 1, 2 and 3, respectively, of the Q-mode factor analysis except for three stations (sample numbers B7, O6 and O14). The similarily of the Q-mode cluster analysis to the Q-mode factor analysis confirms the interpretations made using the Q-mode factor analysis.

5. Conclusions

The non-mixing or partial mixing of different types of groundwater 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 Nethravathi river basin groundwaters. This study also illustrates the utility of multivariate statistical analyses in hydrogeochemical studies.

Acknowledgements

The first author is grateful to Mangalore University for awarding a university research fellowship. The authors thank Mr. Vinod T. Bhat and Mr. Shyama Prasad for assistance during the fieldwork and Ms. Rajasree Kammath for the help during the statistical analyses.

References

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

[2] Balasubrahmanyan MN. Geochronology and geochemistry of Archaean tonalitic gneisses and granites of South Kanara district, Karnataka state, India. In: Windley BN, Naqvi SM, editors. Archaean geochemistry. Amsterdam: Elsevier, 1978. p. 59–78.

[3] Subrahmanya KR, Gangadhara Bhat H. Revised geological map of Dakshina Kannada-utilising IRS data and aerial Photographs. In: Muralikrishna IV, editor. Remote sensing and GIS for environmental planning. New Delhi: Tata Mc Graw-Hill, 1995. p. 160–5.

[4] Chandrasekharappa KN. Morphometric analysis of Nethravathi river basin, Karnataka, India, Ph.D. thesis (unpublished) Mangalore University, 1989. 134pp.

[5] Palmquist WN. Sampling groundwater quality investigations, Department of Mines and Geology, Government of Karnataka. Groundwater stud. 1973;125:14.

[6] Trivedy RK, Goel PK. Chemical and biological methods for water pollution studies, Karad:Environmental Publications, 1986. 248pp.

[7] A.P.H.A. Standard methods of analysis of water, waste water, American Public Health Association, USA, 14th ed., Washington DC, 1995. 1457 pp.

[8] Davis JC. Statistics, data analysis in geology. New York:Wiley, 1973. 550pp.

[9] Dalton MG, Upchurch SB. Interpretation of hydrochemical facies by factor analysis. Groundwater 1978;16:228–33.

[10] Danielsson A, Cato I, Carman R, Rahm L. Spatial clustering of metals in the sediments of the Skagerrak/Kattegat. Appl Geochem 1999;14:689–706.

[11] Sharma S. Applied multivariate techniques. New York: Wiley, 1996.

[12] Fovell R, Fovell MY. Climate zones of the conterminous United States defined using cluster analysis. J Climate 1993;6:2103–35.