

Clustering of groundwaters by Q-mode factor analysis according to their hydrogeochemical origin: A case study of the Cariri Valley (Northern Brazil) wells

Carlos José Freire Machado¹, Maria Marlúcia Freitas Santiago², Horst Frischkorn^{3*} and Josué Mendes Filho⁴

¹Universidade Federal do Pará/Campus de Santarém, Física, Santarém-Pa/Brazil

²Universidade Federal do Ceará, Depto. de Física, Fortaleza-Ce/Brazil

³Universidade Federal do Ceará, Depto. Engenharia Hidráulica e Ambiental, Fortaleza-Ce/Brazil

⁴Universidade Federal do Ceará, Depto. de Física, FortalezaCe/Brazil

Abstract

Factor analysis was applied to 56 groundwater samples collected from wells located in the Araripe Sedimentary Basin, in the north-east of Brazil. The parameters are a set of 9 physicochemical, chemical, and isotope data, constituted by electrical conductivity (EC), ionic concentrations of Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, SO₄²⁻, alkalinity and $\delta^{18}\text{O}/_{00}$. In R-mode factor analysis, the first 3 factors explain 62% of the variance, their loadings allowing the interpretation of hydrogeochemical processes that take place in the area. Q-mode factor analysis on the 56 water samples decreases space dimensionality to 6, explaining 93% of the total database information. With the aid of a scalar and angular measurement method, objects were clustered, resulting in 11 groups classified according to their inherent characteristics, related to their hydrogeological origin.

Keywords: hydrogeochemistry, R-mode factor analysis, Q-mode factor analysis, Araripe sedimentary basin, Cariri valley

Introduction

The ever-increasing demand for potable water requires knowledge of the quality of stored waters, as well as of the natural and anthropogenic processes that influence it. Waters stored in the same aquifer system can differ in their chemical composition due to internal and external processes, and suitable methodologies are needed for their identification.

A great manifold of parameters is used in water research for assessing water quality, pollution, evaporation, flow dynamics and chemical evolution through the water cycle. Thus, great amounts of data are generated. In order to gain insight into the relationships between the parameters associated with a given set of objects, multivariate techniques have been applied to reveal hidden affinities present in the database, and undetectable by other means. Mathematically, these methods reduce space dimensionality by a suitable choice of new dimensions constructed as linear combinations of the original ones, simplifying the representation of the data set and facilitating its interpretation.

Many multivariate analysis techniques have been applied in hydrological studies: R-mode analysis in groundwater quality studies (Grande et al., 1996; Liu et al., 2003; Panagopoulos et al., 2004; Garcia-Rodriguez et al., 2007); R-mode, Q-mode and cluster analysis to assess surface/groundwater interaction and groundwater mixing (Reghunath et al., 2002); R-mode and cluster analysis to study groundwater quality in the Blue Nile basin (Hussein 2004); principal component analysis (PCA), cluster, and discriminant analysis to evaluate spatial and

temporal variations in river waters (Wunderlin et al., 2001; Singh et al., 2004); PCA and R-mode factor analysis to understand origin and variation of each solute in natural waters (Anazawa et al., 2005). Geochemical data were used to test the influence of different factor-analysis techniques on the results extracted (Reimann et al., 2002).

Material and methods

A set of 56 groundwater samples was analysed for 9 physical and chemical parameters comprising major ion concentrations (Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, SO₄²⁻), alkalinity (alka), electrical conductivity (EC) and the isotope oxygen-18 ($\delta^{18}\text{O}/_{00}$). The samples were taken from wells located in the Cariri valley, part of the Araripe sedimentary basin, Brazil, embedded in Precambrian basement rock. This basin is divided between the Federal States of Ceará, Pernambuco (Pe) and Piauí (Pi). The greatest part is in Ceará and comprises the Araripe plateau and the Cariri valley, containing the most important groundwater storage of the State. Figure 1 shows the region under study, enclosing the towns of Crato, Juazeiro do Norte, and Barbalha and the areas of the formations Exu, Arajara, Santana, and Rio da Batateira. Water samples were collected from the Rio da Batateira aquifer, from wells that provide water for industrial, rural and urban use.

Factor analysis

Factor analysis is a multivariate statistical method which, through a linear dependence model constructed in an abstract space called factor score space, searches for correlations among measured variables that characterise a set of objects/samples. Its main feature is to decrease space dimensionality through the construction of a new dimensional base that preserves the essential information contained in the original database. Linear

* To whom all correspondence should be addressed.

+55 85 3366 9775; fax: +55 85 3366 9627;

e-mail: cariri@ufc.br

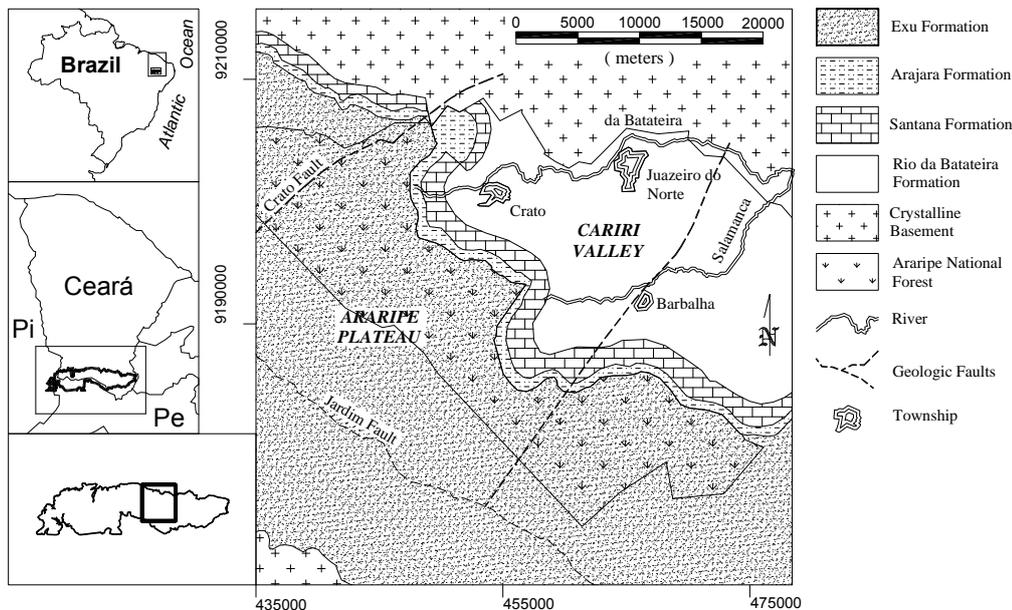


Figure 1
Study area location and map with outcropping areas of Araripe sedimentary basin formations

dependencies of variables are measured in that new space, where new variables are defined by the column vectors of a so-called factor-loading matrix (A) in the space spanned by the column vectors of the factor score matrix (F).

R-mode factor analysis searches for interrelationships among variables. The mathematical model is:

$$Y_{N \times p} = F_{N \times k} A'_{k \times p} + E_{N \times p}$$

where:

Y is the data matrix in deviate form

$$y_{ij} = x_{ij} - \langle x_j \rangle \quad (\text{with } x_{ij} \text{ representing parameter } j \text{ of object/sample } i \text{ and } \langle x_j \rangle \text{ the mean of variable } j) \text{ or in standardised form, } y_{ij} = \frac{x_{ij} - \langle x_j \rangle}{s_j} \quad (s_j \text{ being the standard deviation of variable } j)$$

A' is A transposed and E the residual matrix.

The maximum likelihood estimation method was used to compute estimates for A by a numerical iterative procedure (Jöreskog 1967; 1977; Davis, 1986).

In R-mode factor analysis, to define the best dimensionality (k) of space, we have calculated chi-square (χ_k^2) and the number of degrees of freedom (d_k) for every factor space dimensionality. A measure of the relative importance in increasing the number of dimensions by one is defined by $\Delta_k = (\chi_k^2 - \chi_{k+1}^2) / (d_k - d_{k+1})$ as the difference ratio between chi-square and degrees of freedom.

Q-mode factor analysis is a multivariate technique intended to classify objects according to interrelations among them, so that each object (row) in the data matrix is understood as a combination of hypothetical or real objects with specific parameter values. The technique consists of measuring the resemblance among objects (index of proportional similarity) normalising data matrix rows (objects) so that measured variables can be interpreted as proportions, $|W_{ij}| = x_{ij} / \sqrt{\sum_{j=1}^p x_{ij}^2}$. Imbrie and Purdy (1962) defined the similarity coefficients as $\cos \theta_{nm} = \mathbf{w}_n \cdot \mathbf{w}_m$ of the angle between any two data matrix row vectors (objects n and m), where $\mathbf{w}_n = [w_{n1} \ w_{n2} \ \dots \ w_{np}]$ is a row vector of matrix W . Then, the similarity matrix can be written as $\mathbf{H}_{N \times N} = \mathbf{W}\mathbf{W}'$.

The model is expressed as the product of a factor-loading matrix ($A_{N \times k}$) and a factor score matrix ($F_{p \times k}$), $\mathbf{W}_{N \times p} \approx \mathbf{A}_{N \times k} \mathbf{F}'_{k \times p}$, and the similarity matrix can be written as

$\mathbf{H} = \mathbf{W}\mathbf{W}' = \mathbf{A}\mathbf{F}'\mathbf{F}\mathbf{A}'$. This matrix can be factorised (Reyment et al., 1996) to find F and A .

To achieve simplicity (with the elements of factor-loading vectors approaching 0 or 1) varimax orthogonal rotation, designed by Kaiser (1958) so as to maximise the variance of the factors, was applied to the calculated factor-loading matrices.

Results and discussions

R-mode

The parameters computed for our data set are listed in Table 1. Considerable information is gained when dimensionality increases from 1 to 2 ($\Delta_k = 9.74$), from 2 to 3 ($\Delta_k = 2.59$), but not from 3 to 4 ($\Delta_k = 0.74$). So the best choice for dimensionality is 3, with 74% of accumulated information.

The varimax rotated factor-loading matrix is shown in Table 2 (where only factors with modulus greater than 0.24 are represented). The first factor explains 26% of total variance, the second, 20% and the third, 16%, total accumulated variance being 62%. This is the percentage of variance explained (in the entire database) without overestimating the amount of information available, according to the chi-square analysis. Space dimensionality decreased from the original 9 variables to only 3, so that, with the aid of multivariate statistical analysis, 3 main hydrogeochemical processes can explain the complexity of Cariri valley waters, as presented below.

k	(χ_k^2)	d_k	Δ_k
1	106.00	27	9.74
2	28.20	19	2.59
3	10.10	12	0.74
4	5.68	6	0.77
5	1.84	1	—

TABLE 2
Factor-loading matrix and variance
(only factors with modulus greater than 0.24 are represented)

Variables	Factor-loading matrix		
	1	2	3
CE	0.54	0.61	0.56
Ca ⁺⁺	0.79	0.40	
Mg ⁺⁺	0.75		0.43
Na ⁺	0.30	0.79	
K ⁺			0.57
Cl ⁻			0.41
SO ₄ ⁻	0.51	0.38	
Alka	0.69	0.36	-0.63
O ¹⁸		-0.59	
Variance explained by factor (%)			
	26.09	20.17	15.68
Cumulative variance (%)			
	26.09	46.26	61.94

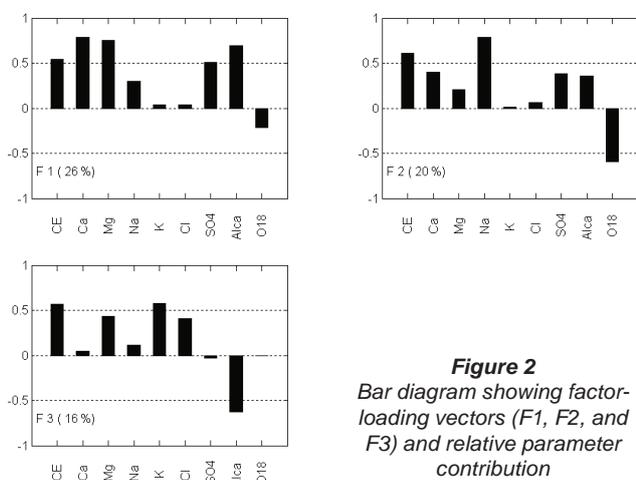


Figure 2
Bar diagram showing factor-loading vectors (F1, F2, and F3) and relative parameter contribution

Figure 2 is a bar diagram illustrating the relative importance of the variables in the factor-loading vectors from Table 2. All 3 factors have high EC loadings. The 1st factor, explaining 26% of the entire sample set variance, shows high correlation between Ca²⁺, Mg²⁺, SO₄²⁻, alkalinity and EC. As limestone and gypsum are common minerals in the Santana formation (Ponte and Appi, 1990), this factor proves that hydrogeochemical reactions relating precipitation/dissolution processes with calcite, dolomite, and gypsum minerals are important in water quality evolution in this area.

The 2nd factor, corresponding to 20% of total variance, is related to δ¹⁸O₀₀, Na⁺, Ca²⁺, SO₄²⁻ and alkalinity. High correlation with Na⁺ and, to a lesser extent, with Ca²⁺, SO₄²⁻, and alkalinity can be associated with ion exchange by clay minerals, abundant in the Rio da Batateira formation.

The 3rd factor, responsible for 16% of the total variance, shows high correlations with EC, Mg²⁺, K⁺, Cl⁻ and inverse correlation with alkalinity, and so it could represent contamination of waters by domestic sewage.

Q-mode

As in the R-mode case, Q-mode factor analysis also needs to define the dimensionality of factor score space. Table 3 defines space dimensionality and shows the information carried by each factor and total accumulated information as space dimensionality increases. When factors number 6, 93% of the information is accumulated; information is more uniformly distributed among factors from the second one on. The information for Factor 7, in 7 dimensions, and for Factors 7 and 8, in 8 dimensions, has low importance and can be discarded.

Results from varimax rotated factor-loading matrix calculation are given in Table 4 (next page) together with object (well) identification.

Application of the selection criteria (described above at the end of Q-mode factor analysis description) to the 56 elements in factor space resulted in 11 groups (Table 5 – next page). In order to interpret groups' characteristics, parameter means for each group were calculated (Table 6 – next page). As dimensionality is greater than 3, it is impossible to visualise the results from this procedure graphically in our three-dimensional visual space. Instead of grouping objects by visual inspection, they were analysed with respect to their angular and scalar distance between each member of the set, represented by vectors in a six-dimensional space, and from the group's centroid (defined by the mean vector, with unitary modulus, calculated considering all the elements in the group). If the angular separation and scalar distance between a given vector (object) and the group centroid in this factor score space is less than or equal to a respective predefined cut-off value, then the analysed element becomes an element of this group.

In our analysis, a cut off angle (θ_c) of 45° and a cut-off scalar distance (d_c) equal to the equivalent distance between unitary vectors separated by the cut-off angle, i.e. $d_c = 2[1 - \cos(\theta_c)]$ were chosen.

Figure 3 shows bi-dimensional plots of elements' factor score space positions, marked by geometric symbols according to groups. All graphs have Factor 1 as abscissa. To avoid overloading the graphs, only group centroids are shown. Ordinates, representing the 2nd dimension, are Factors 2 to 6, respectively. Values approaching ±1 imply increasing importance.

Table 6 shows that Group 1 (star) waters are only slightly saline and have a δ¹⁸O mean value of -3.1‰, very close to the rainwater value (≈ -3.2‰; Santiago et al., 1997). These waters represent recent recharge derived directly from rainfall. The fact that this group is the most numerous is not surprising, because its member wells exploit the uppermost unconfined aquifer in the Cariri valley. Factor 1 is the most important one to discriminate this group.

TABLE 3
Factor information and accumulated information by factors

	Factors								Accumulated information (%)	Number of factors
	1	2	3	4	5	6	7	8		
Information by factor (%)	23.9	18.0	17.9	15.1	11.5	–	–	–	86.4	5
	31.1	14.4	12.6	12.6	11.9	10.6	–	–	93.2	6
	26.5	17.3	14.1	13.6	12.3	9.2	3.8	–	96.8	7
	25.1	17.7	13.6	13.5	12.3	9.1	3.8	3.7	98.9	8

Ident	Factor-loading matrix					
	1	2	3	4	5	6
PBb01	-0.59			0.27	0.29	-0.64
PBb02	0.76					0.49
PBb03	0.68	-0.26	-0.27		0.53	
PBb04	0.67		0.38	0.55		
PBb05	0.59		-0.26	0.25	0.68	
PBb09	-0.55				-0.70	0.35
PBb10		0.74	-0.44		-0.46	
PBb11	-0.61	0.43			-0.53	
PBb12	0.57			0.72		
PBb13	0.50	-0.39			0.49	0.53
PBbF1	-0.60	0.40		-0.42	-0.32	0.38
PBbF2	0.82	0.51				
PBbF3	-0.31				0.74	
PBbF4	0.76		0.27	0.49		
PBbF5	-0.37	0.69		-0.43		0.37
PCt01			-0.92			
PCt03	-0.70		-0.63			
PCt04	-0.75	-0.47		-0.29	0.27	
PCt05	-0.88				-0.25	
PCt06	-0.78				-0.34	-0.34
PCt07	-0.79				0.33	-0.41
PCt08				0.58	0.68	0.33
PCt09	-0.92					
PCt10	-0.55	0.28	-0.39	-0.31	-0.51	-0.26
PCt11		0.30	-0.87			-0.26
PCt12			-0.96			
PCt13	-0.34		-0.64		0.48	-0.31
PCt14	0.85	-0.30	-0.35			
PCt15		-0.62	0.46	-0.51	0.31	
PCtF1	0.42				0.89	
PCtF2	-0.76	0.29		-0.50		
PCtF3	-0.56			-0.75		0.26
PCtF4		0.41	-0.72	-0.37		-0.25
PCtF5						0.94
PCtF6		-0.39			0.84	
PCtF7	-0.52			-0.52		-0.50
PJN01	-0.84					-0.34
PJN02	-0.91				-0.36	
PJN03						0.92
PJN04	0.32	0.27	0.49			0.65
PJN05	-0.38		0.59			0.57
PJN06	0.40	0.33		0.69	-0.34	0.36
PJN07	0.27			0.79		
PJN08	0.84		0.38	0.25		
PJN09	-0.90					-0.28
PJN10		0.32		0.87		0.26
PJN11	-0.45			0.84		
PJN14		-0.93				
PJN15	0.34	-0.82				0.29
PJN16	0.34	-0.88				
PJN17		-0.77	0.41		-0.31	
PJN18	0.75		0.30	0.40	-0.29	-0.28
PJNF1		0.91				
PJNF2		0.51				
PJNF4	0.55	0.44	0.49	0.29	-0.26	
PJNF5	0.84		0.37			
Explained information (%)						
	31.1	14.4	12.6	12.6	11.9	10.6
Accumulated explained information (%)						
	31.1	45.6	58.2	70.8	82.6	93.2

Group	N	Identification						
1	13	PBb09	PBb11	PBbF1	PCt03	PCt05	PCt06	PCt07
		PCt09	PCt10	PCtF2	PJN01	PJN02	PJN09	
2	9	PJNF5	PBb04	PJN08	PJN18	PBb12	PBbF4	
		PJNF4	PJN07	PJN06				
3	7	PBb13	PCtF1	PBb05	PBbF2	PBb03	PBb02	
4	5	PCt01	PCt11	PCt13	PCtF4	PCt12		
5	5	PJN17	PJN14	PJN16	PJN15	PCt15		
6	4	PJN04	PCtF5	PJN03	PJN05			
7	4	PBb10	PJNF2	PJNF1	PBbF5			
8	3	PCtF7	PCt04	PCtF3				
9	3	PJN11	PCt08	PJN10				
10	2	PCtF6	PBbF3					
11	1	PBb01						

Group	N	CE μS/ cm	Ca ⁺⁺	Mg ⁺⁺	Na ⁺	K ⁺	Cl ⁻	SO ₄ ⁻	Alka (CaCO ₃)	¹⁸ O δ‰ ₀₀
1	13	151.3	11.8	6.2	5.9	4.9	16.4	4.6	41.4	-3.1
2	9	443.1	32.3	16.8	44.3	5.0	28.5	41.2	148.5	-3.9
3	7	408.9	34.4	23.0	13.2	5.2	40.3	22.4	113.8	-3.0
4	5	228.2	17.5	15.4	7.0	3.1	22.1	7.2	68.8	-2.9
5	5	553.6	27.6	28.2	17.7	19.0	83.2	3.5	25.8	-3.5
6	4	272.3	28.1	9.2	11.8	6.6	26.2	2.2	100.1	-3.2
7	4	210.3	21.8	10.0	11.6	3.0	5.3	3.4	104.9	-3.1
8	3	182.3	14.2	7.3	11.7	10.0	33.9	8.4	25.5	-2.7
9	3	213.3	28.4	10.7	6.1	7.2	16.6	12.0	114.0	-3.3
10	2	271.0	20.6	15.6	11.3	11.1	24.0	16.2	70.0	-2.8
11	1	226.0	16.0	9.7	8.0	5.9	28.4	17.1	60.0	-3.2

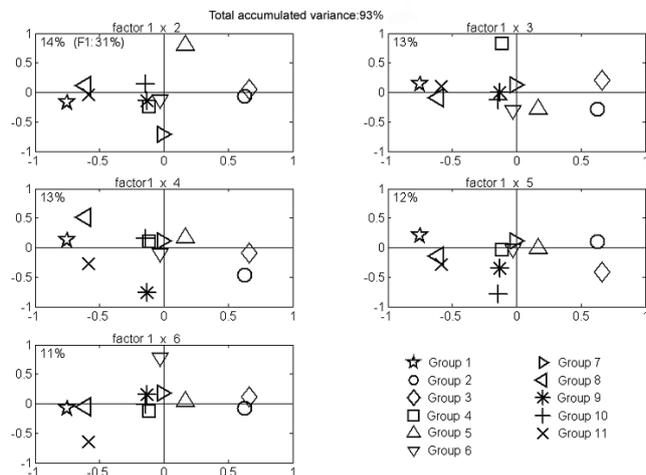


Figure 3
Two-dimensional factor-loading plots for group centroids:
factors 1x2, 1x3, 1x4, 1x5 and 1x6

Group 2 water samples (circle) have high salinity and the lowest values of $\delta^{18}\text{O}$ (-3.9‰). High concentration values of Ca^{2+} , SO_4^{2-} and alkalinity imply that gypsum and limestone dissolution/precipitation processes are involved. These minerals are characteristic of the Santana formation lithology. Thus, this group was interpreted as recharge waters from the top of

the Araripe plateau that percolated the Santana formation, and the low $\delta^{18}\text{O}$ could be due to altitude effect on rainfall and/or to the presence of palaeo-waters because of the long transit time through that aquitard.

Water samples in Group 3 (diamond) show high EC combined with very high Ca^{2+} and high Mg^{2+} and SO_4^{2-} concentrations as well as high alkalinity. Major ions' mean values are similar to those of Group 2, indicating the same geochemical environment. $\delta^{18}\text{O}$, however, is slightly higher (-3.0‰), pointing to an origin from rainfall in lower altitude and/or slightly enriched by evaporation during runoff. We interpret these waters as recharge that leached Araripe plateau cliff matter.

Group 4 (square) waters have as principal characteristics high $\delta^{18}\text{O}$ values (-2.9‰) and low Ca^{2+} , Na^+ , K^+ concentrations and EC, implying fast infiltration to the aquifer. The elevated $\delta^{18}\text{O}$ indicates slightly evaporated water. Factor 3 is important in discriminating this group.

Group 5 (triangle) waters are characterised by very high EC and Cl^- , high K^+ but low Ca^{2+} and SO_4^{2-} concentrations. As Cariri valley natural waters have low Cl^- concentration, these waters, from urban areas, are associated with chlorine pollution through residential wastewater, which is a major source of Cl^- . $\delta^{18}\text{O} = -3.5\text{‰}$ shows that these waters are mixed with palaeo-waters (uprising due to a reduction of hydraulic heads in the superior aquifer, caused by excessive pumping in well-fields for public supply). Factor 2 is of high importance in this group's discrimination.

The water samples in Group 6 (triangle) have mean parameter values near the universal mean. However, SO_4^{2-} concentrations are the smallest of all groups. $\delta^{18}\text{O} = -3.2\text{‰}$ indicates recent, fast recharge without evaporation. Factor 6 better discriminates this water type.

Group 7 waters (arrows to the right) show very low concentration of K^+ and the lowest one for Cl^- . The mean value of $\delta^{18}\text{O} = -3.1\text{‰}$ indicates rainfall-derived recent recharge waters. Like Group 5, Factor 2 discriminates this group, but with negative values near -1 . In this sense, it is the opposite of Group 5.

Water samples in Group 8 (arrows to the right) have very low alkalinity, low Ca^{2+} , Mg^{2+} concentrations, and EC. The high $\delta^{18}\text{O}$ value (-2.7‰) reveals recent recharge waters that suffer evaporation before infiltration. Factors 1 and 4 (with positive correlation) best discriminate this group.

Group 9 waters (asterisk) have high alkalinity and close to rainfall $\delta^{18}\text{O}$ (-3.3‰). Factor 4 (with negative correlation) best discriminates this group.

Group 10, with 2 elements, and Group 11, with only 1, could not be interpreted hydrogeologically, but one can see that Group 9 is near Group 10, and Group 11 is near Groups 1 and 8. If a larger cut-off angle had been adopted and the 'discrimination power' reduced that way, these groups would have been integrated into their respective groups.

Conclusions

Multivariate statistical methods of factor analysis are shown to be an important tool for characterising hydrogeochemical processes and clustering groundwaters according to their shared hydrochemical characteristics. The 3 principal factors identified by R-mode factor analysis correspond to 3 principal processes taking place in the study area: precipitation/dissolution processes of calcium carbonate and gypsum, cation exchange processes occurring in clay layers, and processes related to anthropogenic contamination with chlorine. Q-modal analysis grouped all 56 samples collected in the study area into 11 groups, detecting similarities.

The relatively high number of groups found shows the wide variety of these groundwaters. In spite of it the methodology applied was efficient enough to permit association of factors and groups with hydrogeological environmental features of the research area.

Acknowledgments

The authors thank COGERH (Companhia de Gestão dos Recursos Hídricos do Estado do Ceará) and FUNCAP (Fundação de Amparo à Pesquisa do Estado do Ceará) for logistical and financial support, CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Ministério da Educação—Brazil) for a fellowship for one of the authors (Machado) during postgraduate studies.

References

- ANAZAWA K and OHMORI H (2005) The hydrochemistry of surface waters in andesitic volcanic area, Norikura volcano, central Japan. *Chemosphere* **59** 605-615.
- DAVIS JC (1986) *Statistical and Data Analysis in Geology*. John Wiley & Sons, New York. 656 pp.
- GARCIA-RODRIGUEZ F, BATE GC, SMAILES P, ADAMS JB and METZELTIN D (2007) Multivariate analysis of the dominant and sub-dominant epipellic diatoms and water quality data from South African rivers. *Water SA* **44** (5) 653-658. <http://www.wrc.org.za/downloads/watersa/2007/Oct%2007/2083.pdf>
- GRANDE JA, GONZÁLEZ A, BELTRÁN R and SÁNCHEZ-RODAS D (1996) Application of factor analysis to the study of contamination in the aquifer system of Ayamonte-Huelva (Spain). *Ground Water* **34** 155-161.
- HUSSEIN MT (2004) Hydrochemical evaluation of groundwater in the Blue Nile Basin, eastern Sudan, using conventional and multivariate techniques. *Hydrogeol. J.* **12** 144-158.
- IMBRIE J and PURDY E (1962) Classification of modern Bahamian carbonate sediments. *Mem. Am. Assoc. Petrol. Geol.* **7** 253-272.
- JÖRESKOG KG (1967) Some contributions to maximum-likelihood factor analysis. *Psychometrika* **32** 443-482.
- JÖRESKOG KG (1977) Factor analysis by least squares and maximum likelihood. In: Enslein K, Ralston A and Wilf HS (eds.) *Statistical Methods for Digital Computers*. John Wiley & Sons, New York. 454 pp.
- KAISER HF (1958) The varimax criterion for analytic rotation in factor analysis. *Psychometrika* **23** 187-200.
- LIU, CHEN-WUING, LIN, KAO-HUNG, KUO and YI-MING (2003) Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci. Total Environ.* **313** 77-89.
- PANAGOPOULOS G, LAMBRAKIS N, TSOLIS-KATAGAS P and PAPOULIS D (2004) Cation exchange processes and human activities in unconfined aquifers. *Environ. Geol.* **46** 542-552.
- PONTE FC and APPI CJ (1990) Proposta de revisão da Coluna Estratigráfica da Bacia do Araripe. *Proc. XXXVI Brazilian Congress of Geology* (XXXVI Congresso Brasileiro de Geologia). Associação Brasileira de Geologia de Engenharia e Ambiental-ABGE; Natal/Brazil, Oct 1990. **1** 211-226.
- REGHUNATH R, SREEDHARA MURTHY TR and RAGHAVAN BR (2002) The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, India. *Water Res.* **36** 2437-2442.
- REIMANN C, FILZMOSER P and GARRETT RG (2002) Factor analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* **17** 185-206.
- REYMENT R and JÖRESKOG KG (1996) *Applied Factor Analysis in the Natural Sciences* (2nd edn.) Cambridge University Press, Cambridge. 371 pp.
- SANTIAGO MMF, SILVA, CMSV, MENDES FILHO J and FRISCHKORN H (1997) Characterization of groundwater in the Cariri

(Ceará, Brazil) by environmental isotopes and electric conductivity. *Radiocarbon* **39** (1) 49-59.

SINGH KP, MALIK A, MOHAN D and SINHA S (2004) Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) – A case study. *Water Res.* **38** 3980-3992.

SUBBARAO C, SUBBARAO NV and CHANDU SN (1996) Characterization of groundwater contamination using factor analysis. *Environ. Geol.* **28** 175-180.

WUNDERLIN DA, DIÁZ MP, AMÉ MV, PESCE SF, HUED AC and BISTONI MA (2001) Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. A case study: Suquia River Basin (Córdoba-Argentina) *Water Res.* **35** (12) 2881-2894.
