

Factor Analysis in the Study of Geochemical Processes in the Apodi Sedimentary Basin, Northeast of Brazil

Michel Lopes Granjeiro¹; Maria Marlúcia Freitas Santiago¹; Carla Maria Salgado Vidal Silva¹;
Bruno Alves de Mesquita¹; Horst Frischkorn²; Josué Mendes Filho¹

marlucia@fisica.ufc.br; cariri@ufc.br

Recebido: 04/04/12 - revisado: 25/10/12 - aceito: 08/11/13

ABSTRACT

Wells of the Jandaira aquifer (in the Northeast of Brazil) were sampled for analyses of Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl^- , SO_4^{2-} , HCO_3^- , CO_3^{2-} , NH_4^+ , NO_2^- , NO_3^- , Fe , pH , hardness, and electric conductivity (EC). 87 wells were sampled during the dry season of 2009 and during the rainy season of 2010. R-mode factor analysis was applied, using WINSTAT software and the centroid method and Varimax rotation, for the identification of interrelations between the parameters studied. Variables that contribute significantly to the principal factor during both dry and rainy season are Cl^- , EC, hardness, Na^+ and Mg^{2+} ; K^+ plays a role during the dry season only and CO_3^{2-} during the rainy season. It is seen that, in general, recharge to the aquifer reduces the load of the two principal factors in favor of the others. R-mode factor analysis emphasizes the seasonal changes that occur, even when they are too small to be detected in the individual results of chemical analyses on the large compound of samples.

Keywords: hydro-geochemistry, R-mode factor analysis, Jandaira aquifer.

INTRODUCTION

Statistical techniques are used to help describe complex systems; the main feature of this type of analysis and which makes it so valuable in the scientific environment is the ability to reduce data, which makes their interpretation much easier (MINGOTI, 2005).

Factor analysis was used, among other applications, to the groundwater domain, in order to investigate: recharge areas in the Floridan aquifer system, located in Florida, United States (LAWRENCE and UPCHURCH, 1982); the contamination of the Ayamonte-Huelva aquifer system in the province of Huelva in Spain (GRANDE et al., 1996); the hydro-chemical evolution of groundwater in the province of Karnataka, in India (REGHUNATH et al., 2002); the quality of groundwater in a contaminated area in Taiwan (LIU et al., 2003); the concentration of heavy metals in Mirim Lagoon, on the border of Brazil and Uruguay (SANTOS et al., 2003); the level of natural radioactivity along the Gediz River Basin in Turkey

(BAKAÇ, 2004); the influence of a landfill and domestic effluents on superficial waters of River Tarumã and its tributaries, located in Manaus, Brazil (SANTOS et al., 2006); the quality of groundwater of Deltaic Plain Sands aquifer of Warry, located in Nigeria (OLOBANIYI and OWOYEMI, 2006); the effects of high tides and anthropic actions along the coastal area of Cuddalore, Tamil Nadu, in India (SENTHILKUMAR et al., 2008); the quality of groundwater in the area including the Crato-Juazeiro Graben, in the Araripe Sedimentary Basin, located in Southern Ceará, Northeast Brazil (MACHADO, 2005); the quality of groundwater stored in Dunes, Paleodunes, the Barreiras Formation and Migmatite Complex in areas of Fortaleza's metropolitan area in Northeast Brazil (FERNANDES, 2007).

This paper was developed with data from the National Program for Water Resources Development - Proagua Nacional entitled: "Assessing water resources and proposing a shared management model for aquifers of the Apodi Plateau (Chapada do Apodi) between the states of Rio Grande do Norte and Ceará" and its goal is to describe geochemical processes which occur in Potiguar Basin's groundwater, stored in the Jandaira limestone aquifer, using R-mode factor analysis.

¹ Departamento de Física - Universidade Federal do Ceará

² Departamento de Engenharia Hidráulica e Saneamento
Universidade Federal do Ceará

Table 1 - Basic statistics for the 15 parameters analyzed (hard.:hardness, EC: electric conductivity (μScm^{-1}), concentrations are in mgL^{-1} , SDev.: Standard deviation, Sk: coefficient of assimetry, CV: coefficient of variation)

	September 2009					January 2010				
	Mean	Median	SDev.	Sk	CV(%)	Mean	Median	SDev.	Sk	CV((%)
EC	2599.4	1557.0	3452.1	0.91	132.8	2250.9	1534.0	1929.5	1.10	85.7
pH	6.7	6.8	0.3	-1.00	4.5	6.9	6.8	0.4	0.75	5.8
Hard.	705.5	490.0	685.7	0.94	97.2	633.7	500.0	410.4	0.98	64.9
Ca ²⁺	148.6	128.0	110.7	0.56	74.5	153.8	136.0	74.6	0.72	48.5
Mg ²⁺	80.9	43.7	111.3	1.00	72.7	61.2	41.3	73.6	0.81	120.3
Na ⁺	228.5	99.3	469.7	0.83	205.6	172.1	100.4	203.8	1.10	118.4
K ⁺	11.2	6.8	14.8	0.89	132.1	11.7	6.8	15.7	0.94	134.2
Cl ⁻	571.7	231.4	1167.0	0.87	204.1	458.4	253.1	602.6	1.00	131.5
SO ₄ ²⁻	55.2	41.5	41.9	0.98	75.9	87.5	55.6	101.7	0.94	116.2
CO ₃ ²⁻	16.9	14.3	5.8	1.30	34.3	56.7	46.8	35.3	0.84	62.3
HCO ₃ ⁻	450.8	466.5	122.0	-0.39	27.0	423.9	443.9	163.4	-0.37	38.5
NH ₄ ⁺	2.4	1.5	4.4	0.61	183.3	1.6	1.6	0.6	0.00	37.5
NO ₂ ⁻	6.7	1.1	20.4	0.82	304.5	4.0	0.7	16.4	0.60	410.0
NO ₃ ⁻	29.4	24.5	37.1	0.40	126.2	26.2	27.0	13.3	-0.18	50.8
Fe	0.3	0.1	0.8	0.75	266.7	0.6	0.1	1.6	0.94	266.7

of Rio Grande do Norte, and Jaguaruana, Quixeré and Limoeiro do Norte, in the state of Ceará, in two periods: 87 samples were collected during the dry period, in September 2009, and again during the rainy period, in January 2010.

We used standard methods (APHA 1992) and 15 parameters were analyzed in each collected sample: ionic concentrations of Ca²⁺, Mg²⁺, Na⁺, K⁺, Cl⁻, SO₄²⁻, HCO₃⁻, CO₃²⁻, NH₄⁺, NO₂⁻, NO₃⁻ and Fe and pH, hardness and electric conductivity(EC).

R-mode factor analysis was used to identify existing relationships between the studied variables (Mingoti, 2005) using WINSTAT software, with data being included into a Microsoft Office Excel spreadsheet.

FACTOR ANALYSIS

This analysis explains the information from a multi-variable matrix with many data in terms of the compound of few variables only, the so-called factors, which are vectors, orthogonal to each other within an abstract space, and are interpreted as common features that are not observed among original variables. Few factors should provide approximately the same information provided by the original set of data, reducing the system's complexity. High factor loading means that the variable makes a significant contribution to the formation of that

factor. Factors are obtained after some kind of rotation has been performed to the original matrix of data, with the first factor always explaining the highest percentage value of total variance and the higher numbered ones explain progressively smaller portions of the variance.

RESULTS AND DISCUSSION

Table 1 shows a summary of statistics for the fifteen parameters analysed. R-mode factor analysis was performed on both sets of data using the centroid method and Varimax rotation. Table 3 shows eigenvalues of fifteen factors obtained in two simulations and the variance percentage explained by each factor. Tables 4 and 5 show respectively the results of this analysis for data collected in September 2009 and January 2010. In these two tables only those factors that contribute more than 1 % to the accumulated variance are listed. They show that the aquifer recharge decreased the sum of factor loadings of the two major factors, F1 and F2, and increased that of remaining factors.

Figures 3a and 3b allow for comparing the contributions of each variable to the formation of factors F1 and F2, respectively. Figure 3a shows that the variables which contributed significantly to the formation of factor F1 in the dry period and in the rainy period were: Cl⁻, EC, hardness, Na⁺ and Mg²⁺;

Table 2 - t-test and Mann-Whitney U-test

	t-Test		Mann-Whitney U	
	t-value	p-value	U	p-value
Hard.	0.811	0.419	3841	0.910
pH	-3.083	0.002	3219	0.048
CE	0.782	0.435	3825	0.874
Ca ²⁺	-0.370	0.712	3205	0.047
Mg ²⁺	1.349	0.179	4349	0.124
Na ⁺	0.961	0.338	3873	0.986
K ⁺	-0.217	0.828	2841	0.002
Cl ⁻	0.787	0.433	3738	0.677
SO ₄ ²⁻	-2.845	0.005	2882	0.003
CO ₃ ²⁻	-4.501	0.000	19	0.000
HCO ₃ ⁻	1.254	0.211	4194	0.281
NH ₄ ⁺	1.369	0.173	3632	0.554
NO ₂ ⁻	0.440	0.661	4293	0.222
NO ₃ ⁻	1.120	0.264	3360	0.848
Fe	-1.111	0.269	1044	0.335

Table 3 - Factors(F), eigenvalues, variance, and accumulated variance

F	September / 2009			January / 2010			
	Eigen values	Variance (%)	Variance (%acc.)	F	Eigen values	Variance (%)	Variance (%acc.)
1	6.524	43.495	43.495	1	6.284	41.896	41.896
2	1.713	11.422	54.918	2	1.379	9.193	51.090
3	1.267	8.451	63.370	3	1.055	7.036	58.127
4	0.569	3.793	67.163	4	0.773	5.159	63.286
5	0.539	3.598	70.762	5	0.551	3.678	66.964
6	0.426	2.846	73.608	6	0.381	2.545	69.510
7	0.155	1.039	74.648	7	0.282	1.880	71.391
8	0.108	0.725	75.374	8	0.258	1.726	73.117
9	0.087	0.581	75.955	9	0.118	0.788	73.905
10	0.081	0.540	76.495	10	0.092	0.617	74.523
11	0.044	0.294	76.790	11	0.030	0.204	74.728
12	0.018	0.122	76.913	12	0.028	0.187	74.916
13	0.011	0.075	76.988	13	0.013	0.089	75.005
14	0.006	0.043	77.032	14	0.012	0.082	75.087
15	0.0003	0.002	77.034	15	0.003	0.020	75.107

K⁺ ions contributed significantly only in the dry period and CO₃²⁻ only in the rainy period. Such results point to a change in water produced by recharge water in the aquifer. Figure 3b shows that the only variables that contributed significantly to the formation of factor F2 were bicarbonate, in both periods, and pH, in the dry period only.

Table 6 compares factors F3 to F8 in both sets of samples. One can observe that, upon recharge, ions SO₄²⁻ and NH₄⁺ start contributing to

factors that describe a greater share of data variance (sulfate migrated from factor F6 to F3 and ammonium from factor F7 to F6). There was also a precipitation of calcium (the ion Ca²⁺ moved from F1 to F4, iron went from F3 to F8 while nitrate remained in factor F5). Geochemical processes involving calcium carbonate, iron carbonate, calcium sulfate and recharge water explain those changes. In the Jandaíra limestone aquifer, calcium carbonate, calcium sulfate and iron carbonate dissolution contributes to

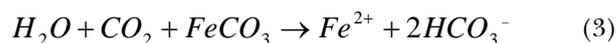
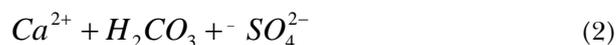
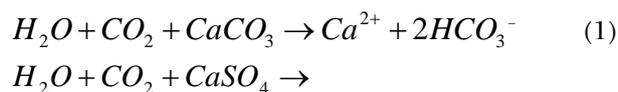
Table 4 - Factor-loading matrix for September/2009 samples (hard: hardness; for units see Table 1)

Variable	Factor						
	1	2	3	4	5	6	7
Cl ⁻	0.995	-0.047	-0.013	0.016	-0.001	0.003	0.021
EC	0.994	-0.017	0.000	0.017	0.028	0.083	0.015
Hard	0.982	-0.046	-0.018	-0.028	0.093	0.063	-0.036
Na ⁺	0.980	0.027	0.018	0.032	-0.061	0.045	0.033
Mg ²⁺	0.928	-0.028	-0.065	-0.026	0.131	0.224	0.019
Ca ²⁺	0.882	-0.073	0.056	-0.031	0.014	-0.178	-0.109
K ⁺	0.790	-0.235	0.017	-0.060	-0.044	0.022	0.370
HCO ₃ ⁻	-0.047	0.864	-0.022	-0.101	0.007	0.156	-0.061
pH	-0.142	0.806	0.122	-0.038	-0.180	-0.091	0.014
Fe	-0.040	0.109	0.755	0.215	0.155	0.071	0.049
CO ₃ ²⁻	-0.053	-0.217	0.253	0.715	-0.006	0.107	-0.052
NO ₃ ⁻	0.113	-0.244	0.227	-0.003	0.672	0.079	-0.006
SO ₄ ²⁻	0.429	0.197	0.155	0.200	0.110	0.598	-0.027
NH ₄ ⁺	0.051	-0.069	0.096	-0.078	0.001	-0.034	0.579
NO ₂ ⁻	0.002	-0.088	0.156	0.029	0.155	0.020	0.016

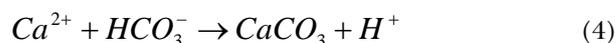
Table 5 - Factor-loading matrix for January/2010 samples (hard: hardness; for units see Table 1)

	Factor							
	1	2	3	4	5	6	7	8
Cl ⁻	0.972	0.160	0.013	0.152	-0.024	-0.026	-0.022	0.010
Na ⁺	0.969	-0.095	0.006	-0.023	0.001	0.063	0.005	-0.053
EC	0.967	-0.148	0.085	0.169	0.021	0.004	-0.022	-0.036
hard	0.852	-0.081	0.325	0.357	0.052	-0.097	-0.056	-0.002
Mg ²⁺	0.829	-0.054	0.511	-0.026	-0.031	-0.084	0.029	0.029
CO ₃ ²⁻	0.601	-0.600	0.332	0.057	0.053	-0.048	-0.091	-0.076
HCO ₃ ⁻	-0.341	0.789	0.024	-0.076	-0.012	-0.149	-0.026	-0.123
SO ₄ ²⁻	0.560	-0.062	0.737	-0.085	0.087	-0.017	0.054	-0.035
Ca ²⁺	0.516	-0.093	-0.088	0.706	0.084	-0.052	-0.104	-0.094
NO ₃ ⁻	0.040	-0.038	0.093	0.120	0.684	-0.013	0.137	-0.042
NH ₄ ⁺	-0.100	-0.237	-0.067	-0.103	-0.013	0.611	0.005	0.179
pH	-0.076	0.016	0.066	-0.164	0.154	0.003	0.652	0.121
Fe	-0.095	-0.170	-0.043	-0.168	-0.060	0.170	0.149	0.611
K ⁺	0.176	-0.100	-0.082	-0.001	-0.130	0.194	0.066	0.038
NO ₂ ⁻	-0.098	-0.138	-0.040	-0.058	0.238	0.224	-0.095	0.079

the presence of Ca²⁺, SO₄²⁻, Fe²⁺ and HCO₃⁻ in the water. The dissolution of minerals results from re-charge with waters containing CO₂:



On the other hand, carbonate precipitation occurs when the calcium hardness, pH or alkalinity are increased:



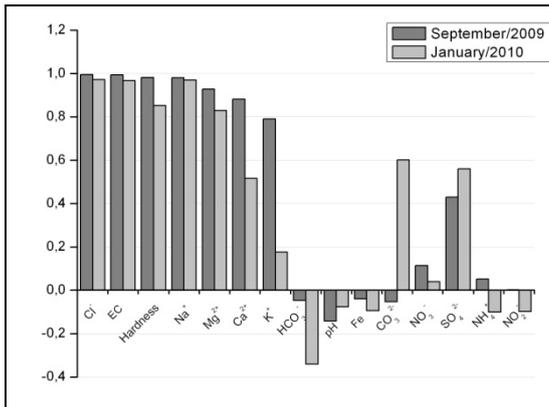


Figure 3a – Bar diagram with F1 factor-loadings and relative parameter contribution.

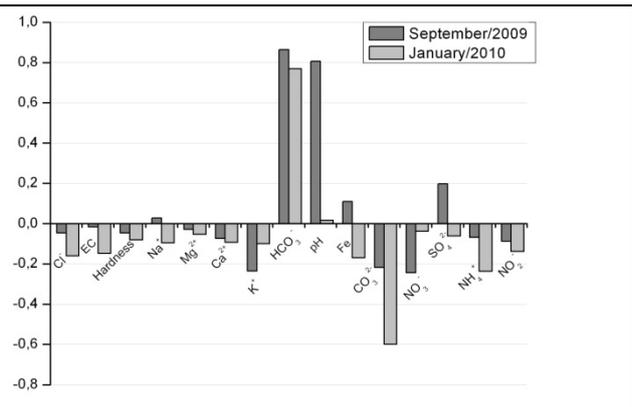
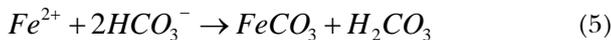


Figure 3b - Bar diagram with F2 factor-loadings and relative parameter contribution.

According to equation (4), Ca^{2+} concentrations decrease, due to the precipitation of carbonate, and H^+ concentrations increase, causing pH to drop, which explains why in the analysis for January data the pH had a high loading in factor F7 only. The decrease in concentrations by precipitation is dominant over the dilution by recharge water; otherwise Mg^{2+} concentrations (from dolomite dissolution) would also be changed in factor contributions, which is not the case.

In the presence of abundant HCO_3^- iron carbonate may precipitate as shown by the following equation (5).



Elevated concentrations of Cl^- and Na^+ result from the deposition of marine aerosol and the high solubility of NaCl.

The increase in NO_3^- and NH_4^+ concentration suggests the recent arrival of water with high ammonium concentrations, certainly originating from the proximity of wells and septic tanks, both located for 'convenience' close to the houses.

In order to detect possible seasonal effects on the water chemistry two statistical tests were applied. The parametric T test shows that out of the fifteen parameters analyzed (Table 2), only three, pH, SO_4^{2-} and CO_3^{2-} , showed a significant difference (between samples collected during September 2009 and January 2010) at a significance level of 5%. Applying the non-parametric Mann Whitney U- test, also Ca^{2+} and K^+ enter this list at the same significance level of 5%.

Table 6 - Variables that contribute most to form the factors.

Factor	Variable	
	(September/2009)	(January/2010)
F3	Fe	SO_4^{2-}
F4	CO_3^{2-}	Ca^{2+}
F5	NO_3^-	NO_3^-
F6	SO_4^{2-}	NH_4^+
F7	NH_4^+	pH
F8	NO_2^-	Fe

CONCLUSIONS

The survey area is extensive with varying stratigraphy which leads, in general, to high standard deviations and coefficients of variation. In addition, some of the chemical parameters suffer exogenous influences from localized contamination, such as the nitrogenous compounds NH_4^+ , NO_2^- e NO_3^- , therefore producing elevated values of these statistical parameters.

The development of large-scale farming in the area suggests that the elevated standard deviation of chloride originates from the use of KCl in banana plantations, where potassium is absorbed by plants whereas chloride remains.

Geochemical processes change the ionic concentrations of water stored in Jandaíra aquifer due to the arrival of early recharge water at the onset of the rainy season. Elevated hardness is affected by the dissolution of aquifer rocks and the decrease usually occurs through dilution with less mineralized water and through the precipitation of minerals.

The effect of recharge is dominant over the three parameters, pH, SO_4^{2-} , and CO_3^{2-} found with the parametric test and less strong on Ca^{2+} and K^+ also found with the nonparametric test. The highest and associated parameters Cl, EC, hardness, Na^+ and Mg^{2+} that occurred in the two samplings are explained by the dissolution of limestone rock and the semiarid climate that favors the increase in salinity (EC) and elevated concentrations of Cl and Na^+ , very common in the region.

Easily detectable with factor analysis, the entry of contaminated water into the aquifer during recharge was identified by the presence of ammonium in the factors. Factors determined by R-mode statistics highlighted the changes in ionic concentrations, even when they were too small to be identified in individual results of hydro-chemical analysis of a large set of samples. Thus, R-mode statistical analysis has proven to be a very important tool in the study of hydro-geochemical processes.

ACKNOWLEDGEMENTS

The authors are indebted to the National Water Agency-ANA, the National Research Council-CNPq, and to UFC for supporting data collection.

REFERENCES

- A.P.H.A.- American Public Health Association. 1992. *Standard Methods for the Examination of Water and Wastewater*. 16ed. New York: APHA.
- BAKAÇ, M. 2004. *R-mode Factor analysis applied to the distribution of radionuclides in Turkey*. Radiation Physics and Chemistry, 71 pp.781-782.
- FERNANDES M.A.B. 2007. *Análise Estatística R-modal e Modelo Geoquímico Inverso no Estudo da Qualidade da Água Subterrânea na Região Metropolitana de Fortaleza- CE*. Doctoral Thesis, UFC, Fortaleza. 115 pp.
- GRANDE, J.A.; GONZÁLE, Z. A.; BELTRÁN, R.; SÁNCHEZ-RODAS, D. 1996. *Application of Factor Analysis to the study of contamination in the Aquifer System of Ayamonte-Huelva (Spain)*. Groundwater, 34, pp. 155-161.
- LAWRENCE, F.W.; UPCHURCH, S.B. 1982. *Identification of Recharge Areas using Geochemical Factor Analysis*. Groundwater, 20 pp. 680-687.
- LIU, CHEN-WUING, LIN, KAO-HUNG, KUO; YI-MING. 2003. *Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan*. Sci. Total Environment, 313, pp. 77-89.
- MACHADO, C.J.F.M. 2005. *Modelagem Geoquímica e de Fluxos do Sistema Aquífero Subterrâneo na Bacia Sedimentar do Araripe*. Doctoral Thesis, UFC, Fortaleza. 116 pp.
- MATIAS, G.D.V.; SILVA, L.M.R.; KHAN, A.S. 2003. *Perfil dos produtores de frutas do município de Limoeiro do Norte – CE frente ao novo paradigma de desenvolvimento do setor*. Revista Ciência Agronômica, 34(1) 7pp.
- MINGOTI SA .2005. *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Belo Horizonte, UFMG, 295 pp.
- OLONANIYI, S.B.; OWOYEMI, F.B. 2006. *Characterization by factor analysis of the chemical facies of groundwater in the Deltaic Plain Sands aquifer of Warri, western Niger delta, Nigeria*. African Journal of Science and Technology (AJST), Science and Engineering Series, 7 (1) pp. 73-81.
- REGHUNATH, R.; MURTHY, T.R.S.; RAGHAVAN, B.R. 2002. *The utility of multivariate statistical techniques in hydrogeochemical studies: an example from Karnataka, Índia*. Water Research, 36, pp. 2437-2442.
- SANTOS, I.N.; HORBE, A.M.C.; SILVA, M.S.R.; MIRANDA, S.Á.F. 2006. *Influência de um aterro sanitário e de efluentes domésticos nas águas superficiais do rio Tarumã e afluentes, AM*. Revista Acta Amazonica, 36 (2) pp. 229-236.
- SANTOS, I.R.; BAISCH, P.; LIMA, G.T.N.P.; SILVA-FILHO, E.V. 2003. *Metais pesados em sedimentos superficiais da lagoa Mirim, fronteira Brasil-Uruguai*. Geochimica Brasiliensis, 17 (1) pp. 37-47.
- SENTHILKUMAR, G.; RAMATHAN, A.L.; NAINWAL, H.C.; CHIDAMBARAM, H. S. 2008. *Evaluation of the hydro geochemistry of groundwater using factor analysis in the Cuddalore coastal region, Tamil Nadu, Índia*. Indian Journal of Marine Sciences, 37(2) pp. 181-185.

Análise Fatorial No Estudo De Processos Geoquímicos Na Bacia Sedimentar Do Apodi, Nordeste do Brasil

RESUMO

Os poços do aquífero de Jandaira (no Nordeste do Brasil) foram amostrados para análises de Ca^{2+} , Mg^{2+} , Na^+ , K^+ , Cl , SO_4^{2-} , HCO_3^- , CO_3^{2-} , NH_4^+ , NO_2^- , NO_3^- , Fe , pH , dureza e condutividade elétrica (CE). Oitenta e sete poços foram amostrados durante a estação seca de 2009 e durante a estação de chuvas de 2010. A análise do fator de moda R (R-mode factor analysis) foi aplicada usando software WINSTAT e o método centróide e rotação Varimax, para a identificação de interrelações entre os parâmetros estudados. Variáveis que contribuem significativamente ao fator principal tanto durante a estação seca, quanto na de chuvas são Cl , CE , dureza, Na^+ e Mg^{2+} ; K^+ desempenha um papel apenas durante a estação seca e CO_3^{2-} durante a estação de chuvas. Observa-se que, em geral, a recarga do aquífero reduz a carga dos dois fatores principais em favor dos outros. A análise de fator de moda R enfatiza as mudanças sazonais que ocorrem, mesmo quando são pequenos demais para serem detectados nos resultados individuais das análises químicas sobre o grande composto de amostras.

Palavras-chave: hidro-geoquímica, análise de fator de moda R, aquífero de Jandaira.