

Multivariate Statistics Applied to the Identification of Compositional Control Parameters for Groundwater

Estatística Multivariada Aplicada à Identificação de Parâmetros de Controle Composicional das Águas Subterrâneas

Karen Vendramini de Araújo¹ , George Satander Sá Freire¹ , Itabaraci Nazareno Cavalcante¹ , Rafael Mota de Oliveira¹ , Diolande Ferreira Gomes¹ , Maria da Conceição Rabelo Gomes²  & Filipe da Silva Peixoto³ 

¹Universidade Federal do Ceará, Departamento de Geologia, Fortaleza, CE, Brasil

²Universidade do Estado do Pará, Centro de Ciências Naturais e Tecnologia, Belém, PA, Brasil

³Universidade do Estado do Rio Grande do Norte, Departamento de Geografia, Natal, RN, Brasil

E-mails: karenva83@gmail.com; satanderfreire@gmail.com; itabaracicavalcante@gmail.com; rafaelmotageo@gmail.com; dfreire68@gmail.com; conceicaoabelo@yahoo.com.br; fpeixoto10ufc@gmail.com

Corresponding author: Karen Vendramini de Araújo; karenva83@gmail.com

Abstract

The objective of the present study was to identify the most influent parameters in the composition of groundwater in the municipality of Icapuí, Ceará - Brazil, seeking correlations with the composition of the percolating aquifer formations that can be associated with the sources of these components. For this purpose, multivariate statistical techniques were applied by means of a Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). The PCA allowed a reduction of physicochemical parameters and determined the two components responsible for approximately 86% of total variance in the data for both sampling periods (rainy and dry). The first component is represented by variables that indicate natural rock weathering processes, and the second comprises seasonality and pollution indicators. Samples were also correlated through HCA according to compositional similarities, which were associated with possible natural or human sources.

Keywords: Factorial analysis; Principal component analysis; Hierarchical cluster

Resumo

Este estudo teve como objetivo identificar os parâmetros de maior influência na composição das águas subterrâneas no município de Icapuí, Ceará - Brasil, buscando correlacionar com a composição das formações aquíferas percoladas associando às possíveis fontes destes constituintes. Para tanto foram aplicadas técnicas estatísticas multivariadas de Análise de Componentes Principais (ACP) e Análise de Agrupamento Hierárquico (AAH). A ACP promoveu a redução dos parâmetros físico-químicos e definiu duas componentes responsáveis por aproximadamente 86% da variância total dos dados para ambos os períodos de coleta (chuvoso e estiagem). A primeira componente é representada pelas variáveis indicadoras de processos de intemperismo natural das rochas e a segunda é formada por indicadores de sazonalidade e poluição. Por meio da AAH, as amostras foram correlacionadas de acordo com as similaridades composicionais, as quais foram associadas às possíveis fontes naturais ou antrópicas.

Palavras-chave: Análise fatorial; Análise de componentes principais; Agrupamento hierárquico

1 Introduction

Groundwater is a considerably important resource, particularly in regions such as northeastern Brazil, where there is an unbalance between water supply and demand. Rainfall variability is a natural driving factor in these types of regions, which when temporally poorly distributed, and in combination with high evaporation rates, decreases the supply of water especially in superficial reservoirs.

Groundwater is the only source of water supply in nearly 40% of Brazilian municipalities, particularly small ones where it guarantees safe water with low treatment costs (Villar 2016). In the semiarid region of the country, aquifers have great strategic importance, particularly in coastal areas, such as in Icapuí, state of Ceará. This municipality has groundwater as its only water supply, which makes it extremely important for the water security of local communities and to guarantee economic activities.

Maia (2018) monitored the concentrations of nitrogen and microbiological compounds in groundwater in the municipality of Icapuí and correlated them with the cases of waterborne diseases. The results made it possible to attribute the presence of bacteria and nitrogen compounds to the increase in cases of diarrheal diseases, as well as the reappearance of cases of viral hepatitis.

Characterization and monitoring of the chemical composition of groundwater are indispensable when determining compatible and adequate uses with the qualitative conditions of this resource (Heibati et al. 2017; Molinari et al. 2019; Besser et al. 2019). Consistent space-time monitoring involves determining many water characteristics in several places and periods, which generates a large amount of correlated information (Bertossi et al. 2013).

Multivariate statistical analysis is an important tool when manipulating and interpreting data with many variables, since it allows to reduce the information contained in the original variables into a smaller set of statistical variables with minimal loss of information (Hair et al. 1998).

Different multivariate statistical techniques, such as the Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA), have been increasingly applied to studies regarding groundwater (Ado et al. 2019; Everst & Özcan 2019; Keita & Zhonghua 2017; Liu et al. 2017).

Gomes and Cavalcante (2017) used the techniques of Principal Component Analysis and Hierarchical Cluster Analysis to identify the similarity of the determinant variables of groundwater quality in the municipality of Fortaleza. They defined three components responsible for about 90% of the total variability of the data, the first component is indicative of pollution, the second of alkalinity and the third of salinity.

Thus, the aforementioned techniques were used in the present investigation with the aim of identifying the variables that most contribute to the physicochemical characteristics of groundwater, seeking to detect temporal and spatial variations caused by either human or natural factors associated with seasonality. Therefore, contributing to the understanding of a relationship between the composition of percolated rocks and groundwater in the municipality of Icapuí, Ceará – Brazil.

2 Study Area

The municipality of Icapuí is located in the eastern extremity of the state of Ceará (Figure 1), covers about 421 km², and has approximately 20,000 inhabitants (IBGE 2021). Water supply is carried out through tubular wells administered by the publicly owned Autonomous Water and Sewage Service (SAAE) and through private wells.

The climate is Tropical Warm Semi-arid, with an average temperature of 26 to 28° C, and maximums between 30 and 31° C. According to the historical series from 1988 to 2021, the average total annual rainfall is 894 mm, with the highest precipitation between January and June, averaging 840 mm, and only 54 mm of average rainfall between July and December (FUNCEME 2021). Thus, the first half of the year is defined as the rainy season and the second as the dry season.

According to Mente (2008), the region is inserted within the hydrogeological context of the Coastal Province, more specifically in the Potiguar sub-province, represented by the Dunas, Barreiras, Jandaíra, and Açú aquifers.

The Dunas Aquifer corresponds to wind deposits that occur across the whole coastline, with widths between 6 and 20 meters and is composed, according to Maia (2018), of well selected sands of fine to mean grain size.

The Barreiras Aquifer is found outcropping or subjacent to the Dunas Aquifer and has the greatest water potential in the area, which is widely used. It is formed by silt-clay fine-grained sandstones and red sandstones of mean to coarse sand sizes, with conglomerate levels (Sousa 2002).

The Jandaíra Aquifer is located at depths between 20 and 108 meters, with concordant upper contact with the Barreiras Aquifer and lower contact with the Açú Aquifer. It is mainly composed of limestones, marls, siltites, and dolomites (Morais et al. 2005; Vasconcelos, Teixeira & Alves Neto 2010).

The Açú Aquifer occurs under confinement conditions, caused by both the clay layers of the upper portion of the Açú Formation and by the base of the Jandaíra Aquifer, and can be detected at depths greater than 200 meters. It is composed predominantly of coarse to conglomeratic

sandstones at its base, moving up to medium sandstones in its intermediate portion, and with finer sandstones at the top, showing a continuous vertical increase in the amount of clay (Morais et al. 2005).

3 Material and Methods

Data from 50 samples of groundwater collected from 25 wells in the municipality of Icapuí were used for physicochemical analyses (Figure 1). These samples were collected over the course of two campaigns, carried out in the months of February (rainy season) and August (dry season) of 2019, aiming to identify the influence of seasonality in the physicochemical characteristics of the water. The parameters analyzed included pH, electrical conductivity (EC), total dissolved solids (TDS), total alkalinity (TA), total hardness (TH), turbidity, major elements (bicarbonate, calcium, magnesium, chloride, sulfate, sodium, and silica), and minor elements (potassium, nitrate, fluoride, and bromide). X-ray fluorescence data was also obtained

from four samples of the main aquifers in the area (Dunas, Barreiras, Jandaíra, and Açú) to identify the main oxides present and establish their proportions in each sample. This step was conducted seeking to associate possible origins to the constituents found in the groundwater. Sampling points were selected based on access to the sampling location and use of the water obtained, considering that 80% of the wells are intended for public supply.

The multivariate statistical analyses carried out were a Factorial Analysis, by means of PCA, and an HCA, using the SPSS Statistics software (version 17.0). The PCA regards the explanation of covariance structure through few linear combinations of the original variables being studied, with the objective of reducing the original dimension, facilitating the interpretation of the analyses conducted (Ferreira 1996). In general, the explanation for all the variability of a system determined by the variables can be done through its principal components (Buttafuoco et al. 2017; Calazans et al. 2018).

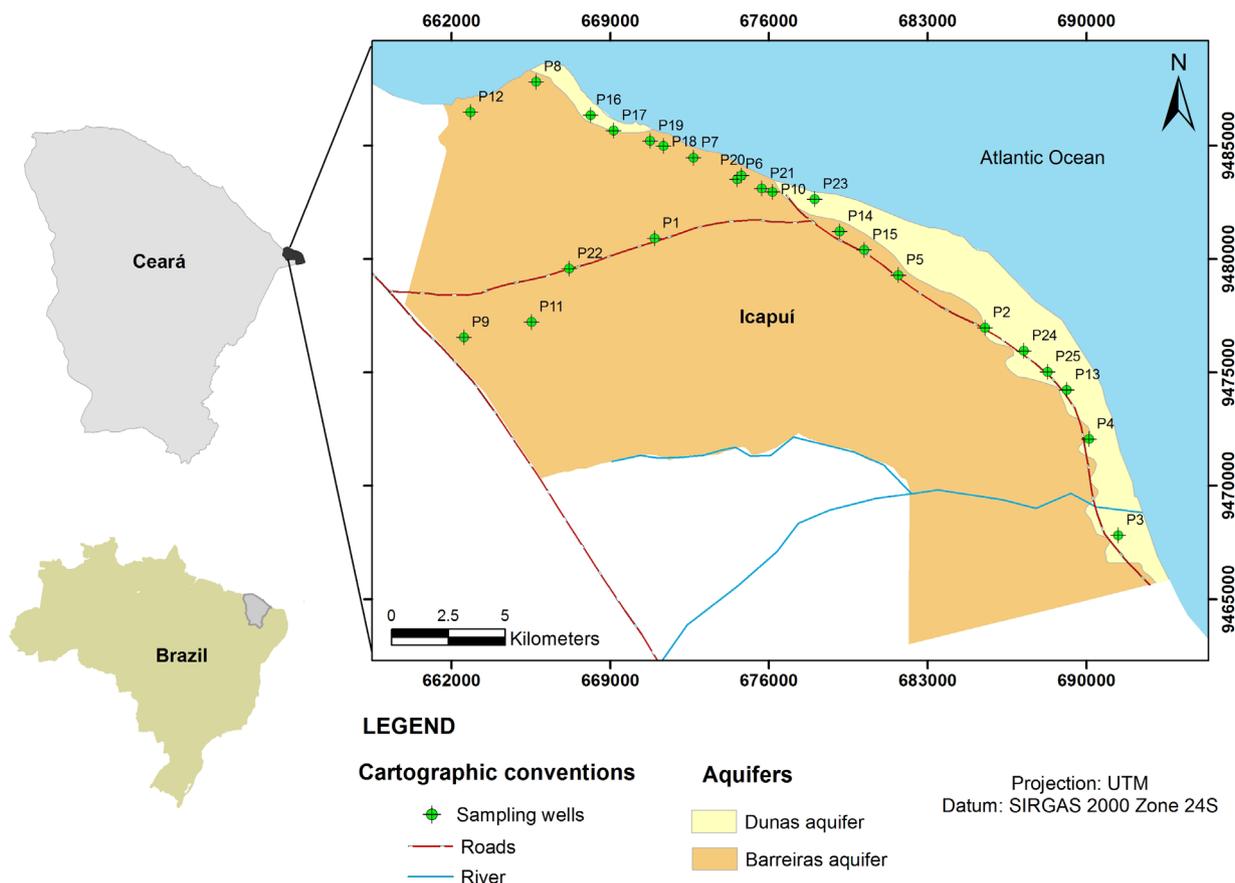


Figure 1 Distribution of the wells sampled and aquifers in the study area.



The factorial analysis was initiated by transforming the original data matrix into a correlation matrix. The KMO rate (Kaiser-Meyer-Olkin Measure of Sampling Adequacy) was used to verify the adequacy of the data in the factorial analysis, which considers an analysis satisfactory with KMO > 0.5, in addition to Bartlett's Test of Sphericity, considered significant at $p < 0.01$, to test the null hypothesis that the variables analyzed were not correlated (Ferreira 1996; Hair et al. 1998). The values of the correlation matrix were later extracted through the PCA method. Then, the rotation of the factorial load matrix, generated in the extraction, was carried out using the varimax rotation method, which minimizes the number of variables with high loads in different factors, allowing the association of a variable with a single factor (Gomes, Anjos & Daltro 2020; Keita & Zhonghua 2017).

Finally, the samples were clustered based on the similarities of the characteristics analyzed, considering the attributes explained by each component in the factorial analysis. To do so, the HCA technique was used, which interlinks the samples according to their associations, producing a dendrogram where samples that are considered similar due to the variables chosen are clustered together (Dabgerwal & Tripathi 2016; Tiri, Lahbari & Boudoukha 2017). The HCA was conducted through the Ward method, with the similarity measure obtained by means of the squared Euclidean distance. The groups generated were submitted to an analysis of variance (One-Way ANOVA) and presented significance levels below 5%, indicating that they formed a relatively stable set of groups (Gomes & Cavalcante 2017; Gomes, Anjos & Daltro 2020).

4 Results and Discussion

4.1 Principal Component Analysis (PCA)

The PCAs were carried out based on simulations initially using 17 physicochemical variables. In the final simulation, 12 variables presented significant results in the definition of the most adequate model to explain the variation in data in both sampling periods (rainy and dry).

Regarding the model for rainy season sampling, the variables that were found to be significant were EC, chloride, calcium, magnesium, TH, sulfate, potassium, bromide, TDS, pH, bicarbonate, and TA. In turn, the representative variables for the dry period were EC, chloride, calcium, magnesium, TH, sulfate, bromide, TDS, pH, bicarbonate, TA, and nitrate (N-NO₃⁻).

The correlation between the variables selected for the model referring to the rainy period can be observed in Table 1. Nearly 77% of correlation coefficients were above 0.5, indicating a significant correlation between variables, as indicated by Gomes and Cavalcante (2017) and Gomes, Anjos and Daltro (2020). A similar behavior was observed for the dry period (Table 2), with nearly 70% of coefficients presenting values higher than 0.5.

The correlation found among pH, bicarbonate, and alkalinity was expected, given that pH is essentially a function of dissolved carbonic gas and alkalinity of the water. Moreover, alkalinity is a direct consequence of the presence or absence of carbonates, bicarbonates, and hydroxides. The strong correlation (> 0.8) among EC, chloride, calcium, magnesium, TH, bromide, and TDS

Table 1 Correlation matrix of the variables from samples collected in the rainy period (February/2019).

Variable	pH	EC	HCO ₃ ⁻	TA	Cl ⁻	Ca ²⁺	Mg ²⁺	TH	SO ₄ ²⁻	K ⁺	Br ⁻	TDS
pH	1.000											
EC	0.373	1.000										
HCO ₃ ⁻	0.810	0.652	1.000									
TA	0.825	0.633	0.999	1.000								
Cl ⁻	0.285	0.988	0.543	0.524	1.000							
Ca ²⁺	0.348	0.845	0.608	0.589	0.842	1.000						
Mg ²⁺	0.326	0.949	0.584	0.563	0.947	0.810	1.000					
TH	0.354	0.940	0.627	0.606	0.937	0.956	0.946	1.000				
SO ₄ ²⁻	0.192	0.687	0.338	0.321	0.685	0.403	0.733	0.589	1.000			
K ⁺	0.084	0.707	0.290	0.266	0.702	0.418	0.713	0.587	0.635	1.000		
Br ⁻	0.348	0.919	0.656	0.640	0.902	0.810	0.805	0.849	0.557	0.522	1.000	
TDS	0.429	0.995	0.696	0.678	0.977	0.840	0.947	0.936	0.715	0.689	0.919	1.000

Note: Highlighted in bold are the coefficients were above 0.5, indicating a significant correlation between variables.

indicated a greater influence of these ions in the variations of conductivity in the water analyzed, as well as in the concentration of TDS. The correlation with hardness can be explained by the presence of calcium and magnesium, since these ions define water hardness.

The PCAs applied showed KMO rates of 0.785 and 0.680 for the samplings conducted in the rainy and dry periods, respectively, indicating that the description of data variability was satisfactory.

The simulations conducted with the 12 variables collected during the rainy and dry periods explained 86.2% and 85.8% of total variance, respectively, and presented two components. Component 1 for the rainy and dry periods explained 70.9% and 68.5% of data variability, and were represented by the variables EC, chloride, calcium, magnesium, TH, sulphate, potassium, bromide, and TDS. These characterize components associated with water mineralization processes that result from water/rock interactions throughout percolation in different substrates. The component 2 of each period corresponded to 15.3% and 17.3% of the variance in the data sets collected in both seasons (rainy and dry, respectively) and encompass the variables pH, bicarbonate, alkalinity, and nitrate. These were defined as components associated with alterations in the acidity, alkalinity, and neutrality conditions of the waters analyzed. The presence of nitrate as a significant variable in samples from the dry period could be associated with a lower volume of rainfall that interferes in the dilution of this parameter, promoted by rainfall recharge.

Table 3 shows the factorial loads and variance explained by the components after applying the orthogonal rotation through the varimax method to minimize the number

of variables that present high loads in each factor, as used in other studies (Ado et al. 2019; Gomes & Cavalcante 2017; Gomes, Anjos & Daltro 2020; Keita & Zhonghua 2017).

4.2 Hierarchical Cluster Analysis (HCA)

The HCA allowed the compartmentalization of samples into groups according to chemical similarities. The number of clusters was defined by the first big difference between coefficients that were re-dimensioned in the cluster. The cut-off point to define the homogeneous clusters was at 5, allowing a better measure of similarity for the formation of groups in both periods. In total, three clusters were formed for the first component and two clusters for the second component for both seasons.

The dendrograms of the samples collected during the rainy period show that in the first component (EC, Cl⁻, Ca²⁺, Mg²⁺, TH, SO₄²⁻, K⁺, Br⁻ and TDS), out of the 25 samples, 8 (32%) compose cluster 1, 14 (56%) form cluster 2, and 3 (12%) are in cluster 3 (Figure 2A). In the second component (pH, HCO₃⁻ and TA) the samples are divided into two clusters, in which the first comprises 11 samples (44%) and the second, 14 (56%) (Figure 2B).

The dendrograms of the samples collected during the dry period show that in the first component (EC, Cl⁻, Ca²⁺, Mg²⁺, TH, SO₄²⁻, Br⁻ and TDS), the 25 samples were distributed among the clusters as follows: 68% (17 samples) in cluster 1, 20% (5 samples) in cluster 2 and 12% (3 samples) in cluster 3 (Figure 3A). In the second component (pH, HCO₃⁻, TA and N-NO₃⁻) the two clusters formed comprised: 13 samples (52%) in cluster 1 and 12 samples (48%) in cluster 2 (Figure 3B).

Table 2 Correlation matrix of the variables from samples collected in the dry period (August/2019).

Variable	pH	EC	HCO ₃ ⁻	TA	Cl ⁻	Ca ²⁺	Mg ²⁺	TH	SO ₄ ²⁻	N-NO ₃ ⁻	Br ⁻	TDS
pH	1.000											
EC	0.421	1.000										
HCO ₃ ⁻	0.831	0.654	1.000									
TA	0.848	0.631	0.998	1.000								
Cl ⁻	0.348	0.991	0.561	0.538	1.000							
Ca ²⁺	0.390	0.862	0.613	0.591	0.853	1.000						
Mg ²⁺	0.321	0.961	0.529	0.505	0.968	0.794	1.000					
TH	0.379	0.956	0.608	0.584	0.953	0.959	0.933	1.000				
SO ₄ ²⁻	0.284	0.657	0.507	0.478	0.606	0.365	0.694	0.538	1.000			
N-NO ₃ ⁻	-0.537	-0.073	-0.391	-0.411	-0.039	0.161	-0.016	0.088	-0.283	1.000		
Br ⁻	0.424	0.945	0.648	0.625	0.932	0.880	0.854	0.916	0.503	-0.012	1.000	
TDS	0.479	0.995	0.713	0.691	0.977	0.858	0.955	0.950	0.687	-0.113	0.932	1.000

Note: Highlighted in bold are the coefficients were above 0.5, indicating a significant correlation between variables.

Table 3 Factorial loads and variance explained in the principal component analyses, after rotation through the varimax method.

Variable	Rainy period		Dry period	
	C1	C2	C1	C2
pH	0.044	0.915	0.242	0.883
Electrical conductivity	0.920	0.371	0.962	0.237
Bicarbonate	0.345	0.914	0.511	0.800
Total alkalinity	0.318	0.926	0.482	0.819
Chloride	0.949	0.262	0.968	0.146
Calcium	0.751	0.431	0.908	0.118
Magnesium	0.923	0.293	0.949	0.125
Total hardness	0.875	0.384	0.977	0.128
Sulphate	0.764	0.051	0.576	0.342
Bromide	0.804	0.423	0.923	0.216
Total dissolved solids	0.901	0.424	0.945	0.309
Potassium	0.798	-0.046	-	-
Nitrate (N-NO ₃ -)	-	-	0.179	-0.791
Variance explained %	70.913	15.326	68.506	17.334
Variance accumulated %	86.240		85.841	

Legend: C1 – Component 1; C2 – Component 2.

Note: Highlighted in bold are the factorial loads of the variables that represent each component.

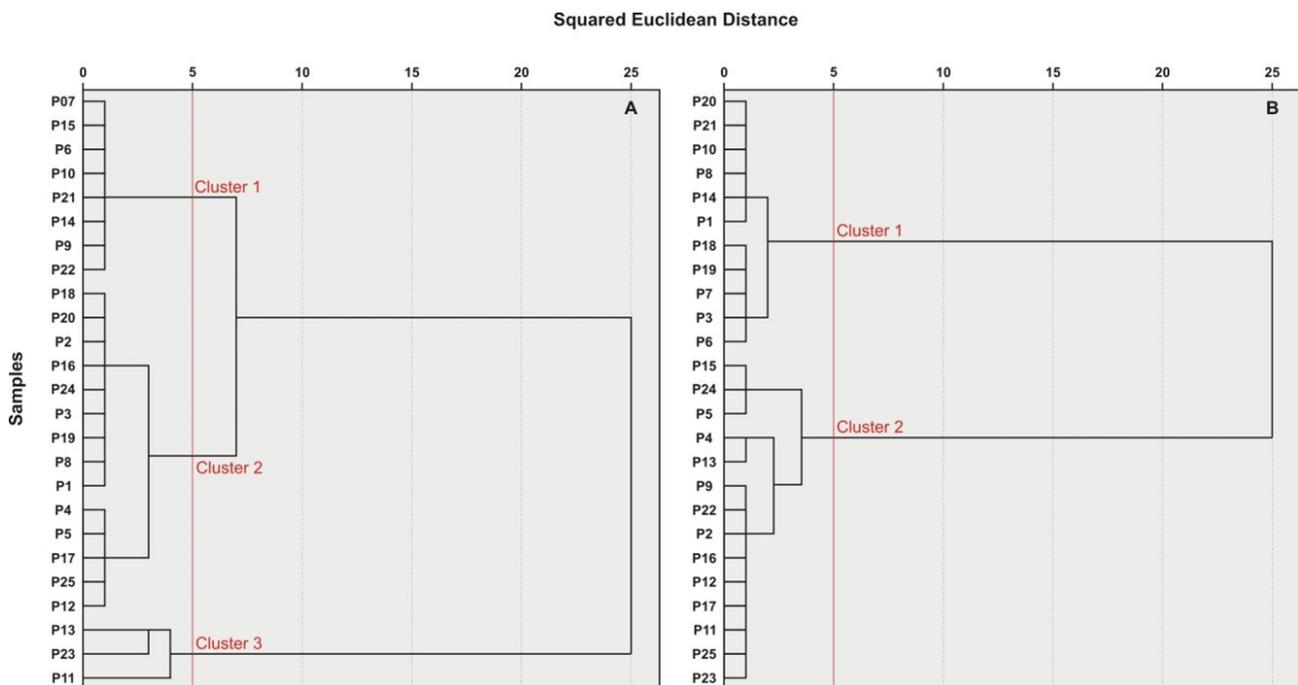


Figure 2 Dendrograms resulting from the hierarchical cluster analysis, for the rainy period, of the variables explained by: **A.** Component 1; **B.** Component 2.



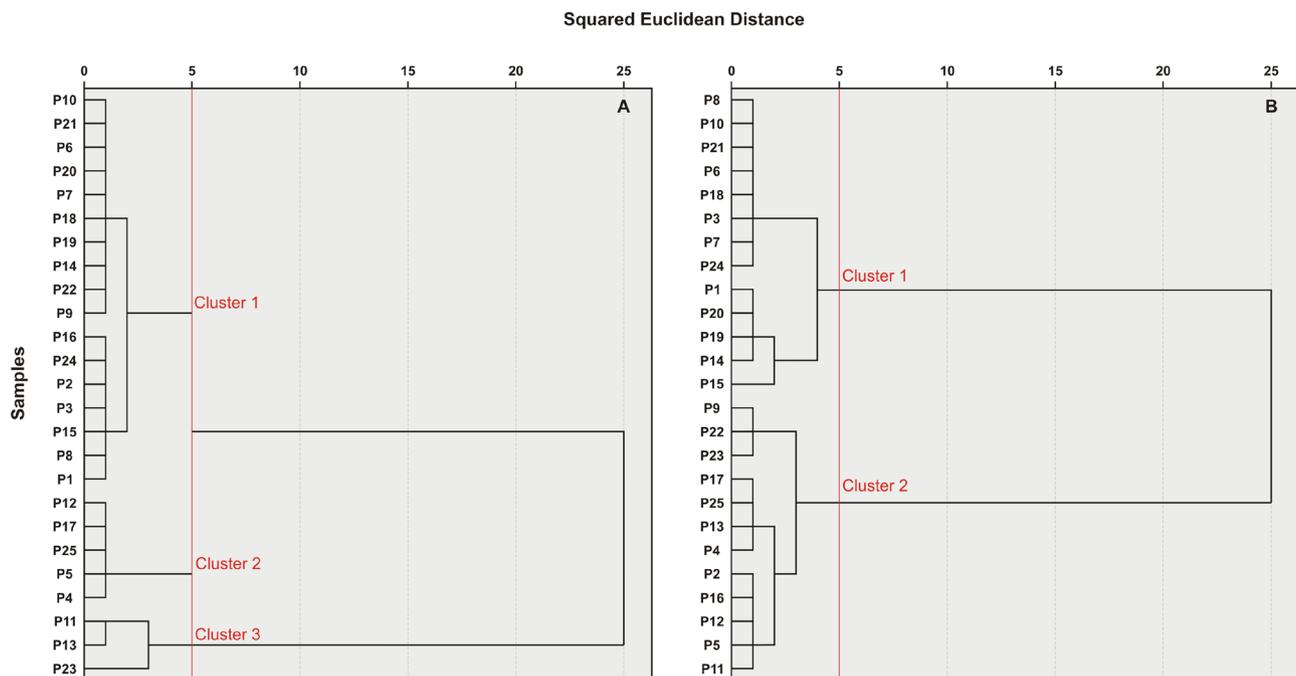


Figure 3 Dendrograms resulting from the hierarchical cluster analysis, for the dry period, of the variables explained by: **A.** Component 1; **B.** Component 2.

The distribution of samples in the clusters of each component was based on the variations in concentration of the elements present, mainly due to the aquifers involved, influence of seasonality, and human factors. Table 4 shows the mean concentrations of variables for the samples in each cluster.

The wells that compose cluster 1 of the first component regarding the rainy period have depths that vary between 10 and 685 meters and obtain water from the Dunas, Barreiras, and Açú aquifers, the latter of which can only be reached by wells that are at least 200 meters deep (western portion of the area). The waters from these wells are characterized by low ionic concentrations associated with the sandy constitution of the aquifers, predominantly composed of SiO_2 (Figure 4), which has low solubility in normal temperature conditions but higher solubility in geothermal groundwater conditions (Gomes, Furtado & Souza 2018). The low salt concentrations also indicate a lower marine influence in these wells. Cluster 2 encompasses wells with intermediate depths, varying between 36 and 130 meters. These waters originate from the Barreiras and Jandaíra aquifers, the latter of which does not outcrop and is detected 100 m below the surface in the western portion of the study area and from 30 m in depth onwards in the eastern coastal zone. The waters in this cluster present ions with higher mean concentrations in relation to cluster 1, thus indicating a greater contribution

from seawater, as well as from the limestones in the Jandaíra aquifer due to dissolution processes involving calcite. The three wells that form cluster 3 present variables depths. The two more shallow ones are located along the coastal zone (10 and 32 meters), catching water from the Dunas and Barreiras aquifers, respectively, while the deepest one is located in the western portion of the study area (594 meters) and catches water from the Açú aquifer. Its waters have the highest mean concentrations in the component and may be related to a marine influence, which the Dunas and Barreiras unconfined aquifers are more susceptible, as well as to the mineralization of the waters resulting from the interaction with the substrate over the percolation associated with the Açú aquifer. According to Celligoi (1999), higher concentrations of sulphate, bromide, and chloride can be attributed to both marine aerosols and to the influence of seawater over continental water, due to the location of the wells near the coastline. High potassium and magnesium values may be associated with percolation through the Açú aquifer, given the presence of these elements in the rocks that compose its framework, identified by means of X-ray fluorescence analyses (Figure 4). These analyses also enabled to attribute the higher concentrations of calcium in the water to the limestones of the Jandaíra aquifer, considering the predominance of this element (nearly 93%) in the composition of the rocks analyzed, associated with high mobility (Reimann & Caritat 1998).

Table 4 Mean values of the parameters of samples collected during the rainy and dry periods, distributed in the hierarchical clusters.

Component	Cluster	Rainy period											
		pH	EC	HCO ₃ ⁻	TA	Cl ⁻	Ca ²⁺	Mg ²⁺	TH	SO ₄ ²⁻	K ⁺	Br	TDS
		(µS/cm)					(mg/L)						
1	1	6.5	518.6	79.19	68.9	80.5	7.1	4.6	37.2	26.1	9.1	0.6	331.6
	2	6.5	1 472.9	139.6	114.4	312.8	39.4	25.8	206.0	49.9	20.7	1.2	841.8
	3	7.2	3 612.0	259.6	212.7	933.0	146.7	85.6	723.3	174.8	25.9	2.9	2 104.7
2	1	5.7	845.7	25.1	20.6	184.3	13.0	12.2	83.3	40.2	15.7	0.7	465.4
	2	7.3	1 878.8	220.7	183.2	413.9	64.6	37.3	316.9	70.7	19.1	1.6	1 116.6

Component	Cluster	Dry period											
		pH	EC	HCO ₃ ⁻	TA	Cl ⁻	Ca ²⁺	Mg ²⁺	TH	SO ₄ ²⁻	N-NO ₃	Br	TDS
		(µS/cm)					(mg/L)						
1	1	5.9	898.9	68.6	58.6	178.1	18.3	14.1	104.2	39.8	7.9	0.7	508.9
	2	7.0	2 046.2	242.5	198.8	427.9	70.7	29.9	301.6	89.1	6.3	2.5	1 175.8
	3	7.0	4 379.3	273.6	224.2	1 159.7	172.0	96.0	830.0	131.1	6.7	4.7	2 394.3
2	1	5.5	887.8	25.0	20.5	188.3	15.6	15.3	102.5	42.4	9.2	0.7	474.6
	2	7.2	2 259.1	239.5	199.6	516.5	81.5	39.9	369.8	80.3	5.5	2.5	1 295.3

Legend EC – Electrical conductivity; TA – Total alkalinity; TH – Total hardness; TDS – Total dissolved solids.

The wells that represent cluster 1 in the second component of the rainy period are shallow to reasonably deep, with waters that originate from the Dunas, Barreiras and Jandaíra aquifers. This cluster is characterized by acidic water that may be associated with the composition of the Dunas and Barreiras aquifers, which are rich in silica and aluminium (Figure 4), and with the sampling period. According to Gomes and Cavalcante (2017), the introduction of recent waters in the aquifer (meteoric water), with higher concentrations of CO₂, promotes an increase in groundwater acidity. Cluster 2 is composed of wells with depths that vary between 10 and 685 meters (western sector) that obtain water from all four aquifers: Dunas, Barreiras, Jandaíra, and Açú. The waters in this cluster vary from neutral to alkaline. According to Hounslow (1995), higher values of pH are usually found in waters in which there is a predominance of Na⁺ and Ca²⁺ ions or in bicarbonate-rich waters. The more alkaline property of these waters in the cluster may be an indication of a greater contribution from the Jandaíra aquifer in its origin.

Cluster 1 of the first component for the dry period is formed by the same wells as cluster 1 and another nine wells from cluster 2 of the first component for the rainy period. Thus, these are also characterized by waters with low salinity, though an increase in the values of this cluster can be observed in comparison to the samples collected during the rainy period. This demonstrates the influence of seasonality in the variation of ionic concentrations in

groundwater, especially in regions where there are shallow aquifers, in which a decrease in underground recharge during periods with less rainfall contributes to an increase of salt concentrations in the waters. The 5 wells that compose this cluster 2 are also part of the second cluster of the first component for the rainy period sampling, presenting the same characteristics regarding intermediate salinity of the waters. Cluster 3 was also formed by the same wells that compose the third cluster of the first component in the rainy period sampling. This cluster is represented by waters with high salt concentrations, particularly chloride. The presence of chloride in groundwater can be attributed to the dissolution of saline deposits, saline intrusions, effluent discharges from chemical industries and domestic sewage (Celligoi 1999). The higher chloride values in the waters of this cluster are possibly associated with marine influence and influence from domestic sewage, given the absence of basic sanitation conditions in the area.

Cluster 1 of the second component in the dry period is composed, again, by the same wells of cluster 1 and another two wells from cluster 2 of the second component in the rainy period, represented by shallow wells with acidic waters. Nitrate was a significant parameter in the samples collected during the dry period, probably because of the lower dilution process that takes place in this condition in comparison to the rainy period. The absence of basic sanitation in the area, with consequent use of septic tanks, contributes to an increase in nitrate concentrations in the

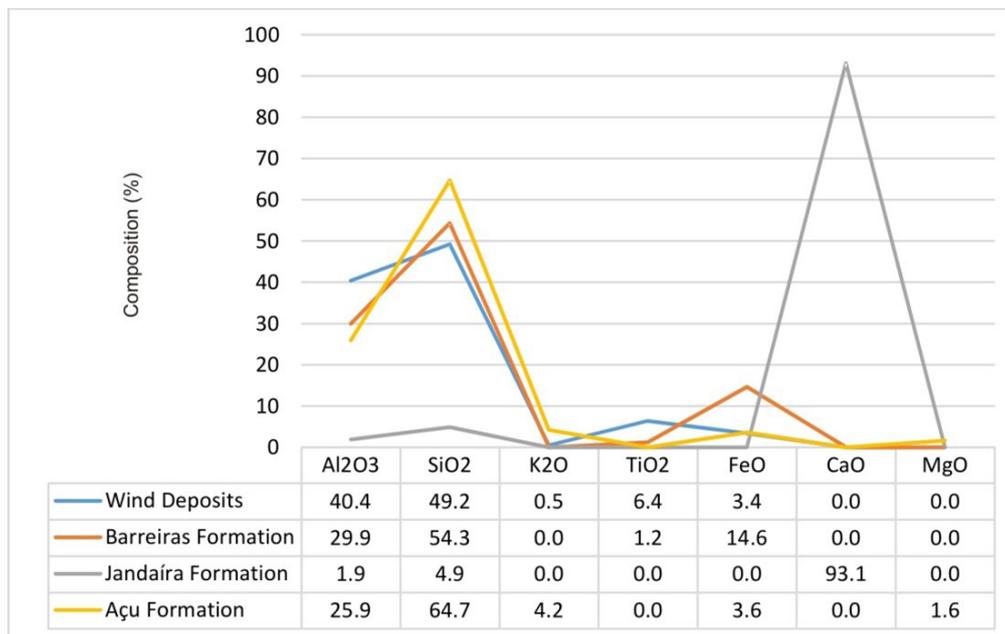


Figure 4 Main oxides present in the aquifers of the study area.

water (Sadler et al. 2016; Maia 2018). In the cluster 1 the nitrate presents values varying between 3.6 and 21.4 mg/L for N-NO₃⁻, with six of the thirteen samples (46%) showing values above the recommendations by the Ministry of Health for drinking water in Brazil, which is a maximum value of 10 mg/L for N-NO₃⁻ and approximately equivalent to the orientation of the World Health Organization (WHO) that establishes limits at 50 mg/L for NO₃⁻ and 11.3 mg/L for N-NO₃⁻. This value was defined as a protective measure against childhood methemoglobinemia (WHO 2011), though other negative health outcomes, including cancer, are associated with ingesting water with high levels of nitrate (Sadler et al. 2016; Maia 2018). Cluster 2 of the second component in the dry period is composed by wells that also part of the cluster 2 of the second component in the rainy period. This cluster encompasses wells deeper than cluster 1, reaching 685 meters, and that obtain water from the Dunas, Barreiras and Açu aquifers, which have pH varying between neutral and alkaline. Nitrate values were lower here than in cluster 1, probably due to the greater depths of the wells. This makes the water less susceptible to superficial contaminations, as also observed by Maia (2018).

5 Conclusions

The factorial analysis by means of PCA allowed the selection of the parameters that hold the most significance regarding the composition of groundwater in the municipality of Icapuí, Ceará. Compositional variations were represented by two components, one associated

with mineralization processes, controlled by electrical conductivity, chloride, calcium, magnesium, total hardness, sulphate, potassium, bromide and total dissolved solids, and another related to seasonality and water pollution, controlled by pH, bicarbonate, total alkalinity and nitrate.

The HCA promoted the clustering of water samples with similar chemical characteristics for each factor. This enabled establishing possible correlations regarding potential origins and elements present. These, in turn, were attributed mainly to natural sources generated through geochemical processes caused by the interaction of groundwater with percolated aquifers.

The multivariate statistical analysis showed an important contribution as a tool in the development of studies that include monitoring physicochemical parameters in groundwater. This is due to the reduction of parameters for analysis without compromising the quality of the information, making the process more efficient from a logistical and economic perspective, especially in regions that have groundwater as their main source of water supply, requiring constant monitoring of compositional alterations.

6 Acknowledgments

The authors would like to thank the Environmental Geochemistry and Electron Microscopy Laboratories of the Geology Department of the Federal University of Ceará, for their support in carrying out the physicochemical and X-ray fluorescence analyses.

7 References

- Ado, M., Adeyeye, O., Xiao, C. & Liang, X. 2019, 'Hydrogeochemistry of Groundwater from Kazaure Area, NW Nigeria using Multivariate Statistics', *E3S Web of Conferences*, vol. 98, no. 07001, pp. 6, DOI:10.1051/e3sconf/20199807001.
- Bertossi, A.P.A., Menezes, J.P.C., Cecílio, R.A., Garcia, G.O. & Neves, M.A. 2013, 'Seleção e agrupamento de indicadores da qualidade de águas utilizando Estatística Multivariada', *Semina: Ciências Agrárias, Londrina*, vol. 34, no. 5, pp. 2025-36.
- Besser, H., Redhaoui, B., Bedouia, S., Ayadia, Y., Khelifib, F. & Hamedb, Y. 2019, 'Geochemical, isotopic and statistical monitoring of groundwater quality: Assessment of the potential environmental impacts of the highly polluted CI water in Southwestern Tunisia', *Journal of African Earth Sciences*, vol. 153, pp. 144-55, DOI:10.1016/j.jafrearsci.2019.03.001.
- Buttafuoco, G., Guagliardi, I., Tarvainen, I. & Jarva, J. 2017, 'A multivariate approach to study the geochemistry of urban topsoil in the city of Tampere, Finland', *Journal of Geochemical Exploration*, vol. 181, pp. 191-204, DOI:10.1016/j.gexplo.2017.07.017.
- Calazans, G.M., Pinto, C.C., Costa, E.P., Perini, A.F. & Oliveira, S.C. 2018, 'Using multivariate techniques as a strategy to guide optimization projects for the surface water quality network monitoring in the Velhas river basin, Brazil', *Environmental Monitoring and Assessment*, vol. 190, no. 12, e726, DOI:10.1007/s10661-018-7099-z.
- Celligoi, A. 1999, 'Considerações sobre análises químicas de águas subterrâneas', *Geografia (Londrina)*, vol. 8, no. 1, pp. 91-7, DOI:10.5433/2447-1747.1999v8n1p91.
- Dabgerwal, D.K. & Tripathi, S.K. 2016, 'Assessment of surface water quality using hierarchical cluster analysis', *International Journal of Environment*, vol. 5, no. 1., pp. 32-44, DOI:10.3126/ije.v5i1.14563.
- Everest, T. & Özcan, H. 2019, 'Applying multivariate statistics for identification of groundwater resources and qualities in NW Turkey', *Environmental Monitoring and Assessment*, vol. 191, no. 47, DOI:10.1007/s10661-018-7165-6.
- Ferreira, D.F. 1996, *Análise multivariada*, Lavras, Minas Gerais. Fundação Cearense de Meteorologia e Recursos Hídricos – FUNCEME 2021, viewed 17 August 2021, <<http://www.funceme.br/index.php/areas/rede-de-monitoramento/postos-pluviometricos>>.
- Gomes, L.S., Furtado, A.C.R. & Souza, M.C. 2018, 'A Sílica e suas Particularidades', *Revista Virtual de Química*, vol. 10, no. 4, pp. 1018-38, DOI:10.21577/1984-6835.20180072
- Gomes, M.C.R. & Cavalcante, I.N. 2017, 'Aplicação da análise estatística multivariada no estudo da qualidade da água subterrânea', *Revista Águas Subterrâneas*, vol. 31, no. 1, pp. 134-49, DOI:10.14295/ras.v31i1.28617.
- Gomes, M.C.R., Anjos, J.A.S.A. & Daltro, R.R. 2020, 'Multivariate statistical analysis applied to the evaluation of groundwater quality in the central-southern portion of the state of Bahia – Brazil', *Revista Ambiente & Água*, vol. 15, no. 1, e2408, DOI:10.4136/ambi-agua.2408.
- Hair Jr, J.F., Anderson, R.E., Tathan, R.L. & Black, W.C. 1998, *Multivariate data analysis*, Prentice Hall, New Jersey, 928 p.
- Heibati, M., Stedmon, C.A., Stenroth, K., Rauch, S., Toljander, J., Säve-Söderbergh, M. & Murphy, K.R. 2017, 'Assessment of drinking water quality at the tap using fluorescence spectroscopy', *Water Research*, vol. 125, pp. 1-10, DOI:10.1016/j.watres.2017.08.020.
- Hounslow, A.W. 1995, *Water quality data: analysis and interpretation*, Lewis Publishers New York, Boca Raton.
- Instituto Brasileiro de Geografia e Estatística – IBGE 2021, viewed 18 June 2021, <<http://www.ibge.gov.br>>.
- Keita, S. & Zhonghua, T. 2017, 'The assessment of processes controlling the spatial distribution of hydrogeochemical groundwater types in Mali using multivariate statistics', *Journal of African Earth Sciences*, vol. 134, pp. 573-89, DOI:10.1016/j.jafrearsci.2017.07.023.
- Liu, P., Hoth, N., Drebenstedt, C., Sun, Y. & Xu, Z. 2017, 'Hydrogeochemical paths of multi-layer groundwater system in coal mining regions - Using multivariate statistics and geochemical modeling approaches', *Science of the Total Environment*, vol. 601-602, pp. 1-14, DOI:10.1016/j.scitotenv.2017.05.146.
- Maia, S.R.R. 2018, 'Estudo hidroquímico-ambiental do aquífero costeiro do município de Icapuí – CE', PhD thesis, Universidade Federal do Ceará.
- Mente, A. 2008, 'A Água Subterrânea no Brasil', in F.A.C. Feitosa, J. Manoel Filho, E.C. Feitosa, J.G.A. Demetrio, JGA (org.), *Hidrogeologia: Conceitos e Aplicações*, 3rd. ed. revisada e ampliada, CPRM /LABHID, Rio de Janeiro, pp. 31-48.
- Molinari, A., Guadagnini, L., Marcaccio, M. & Guadagnini, A. 2019, 'Geostatistical multimodel approach for the assessment of the spatial distribution of natural background concentrations in large-scale groundwater bodies', *Water Research*, vol. 149, pp. 522-32, DOI:10.1016/j.watres.2018.09.049.
- Morais, F., Melo, J.G., Medeiros, J.I., Srivastava, N.K., Diniz Filho, J.B., Lopes, V.L., Oliveira, J.A. & Vasconcelos, M.B. 2005, *Comportamento das bacias sedimentares da região semi-árida do Nordeste brasileiro. Avaliação do aquífero Açua na borda sul da bacia Potiguar – Trecho: Upanema-Afonso Bezerra*, CPRM/FINEP, Recife.
- Reimann, C. & Caritat, P. 1998, *Chemical Elements in the Environment: Factsheets for the Geochemist and Environmental Scientist*, Springer-Verlag, Berlin, Heidelberg.
- Sadler, R., Maetam, B., Edokpolo, B., Connell, D., Yu, J., Stewart, D., Park, M.J., Gray, D. & Laksono, B. 2016, 'Health risk assessment for exposure to nitrate in drinking water from village wells in Semarang, Indonesia', *Environmental Pollution*, vol. 216, pp. 738-45, DOI:10.1016/j.envpol.2016.06.041.
- Sousa, D.C. 2002, 'Litoestratigrafia e deformação Cenozóica na região de Icapuí, Ceará, e implicações para a estruturação de campos de petróleo na borda ocidental da Bacia Potiguar (NE do Brasil)', PhD thesis, Universidade Federal do Rio Grande do Norte.
- Tiri, A., Lahbari, N. & Boudoukha, A. 2017, 'Assessment of the quality of water by hierarchical cluster and variance analyses of the Koudiat Medouar Watershed, East Algeria', *Applied Water Science*, vol. 7, pp. 4197-206, DOI:10.1007/s13201-014-0261-z.

Vasconcelos, S.M.S., Teixeira, Z.A. & Alves Neto, J. 2010, 'Caracterização do aquífero Jandaíra, porção situada no estado do Ceará, Brasil', *Revista de Geologia*, vol. 23, no. 1, pp. 50-60.

Villar, P.C. 2016, 'As águas subterrâneas e o direito à água em um contexto de crise', *Revista Ambiente & Sociedade*, vol. 19, pp. 83-102.

World Health Organization - WHO 2011, *Guidelines for drinking-water Quality*, Geneva.

Author contributions

Karen Vendramini de Araújo: conceptualization; formal analysis; methodology; validation; writing-original draft; writing – review and editing; visualization. **George Satander Sá Freire:** supervision; validation. **Itabaraci Nazareno Cavalcante:** supervision; writing – review and editing. **Rafael Mota de Oliveira:** writing-original draft; writing – review and editing. **Diolande Ferreira Gomes:** writing – review and editing. **Maria da Conceição Rabelo Gomes:** methodology; validation; writing – review and editing. **Filipe da Silva Peixoto:** writing – review and editing.

Conflict of interest

The authors declare no conflict of interest.

Data availability statement

All data included in this study are publicly available in the literature.

Funding information

Not applicable.

Editor-in-chief

Dr. Claudine Dereczynski

Associate Editor

Dr. Gerson Cardoso da Silva Jr.

How to cite:

Araújo, K.V., Freire, G.S.S., Cavalcante, I.N., Oliveira, R.M., Gomes, D.F., Gomes, M.C.R. & Peixoto, F.S. 2024, 'Multivariate Statistics Applied to the Identification of Compositional Control Parameters for Groundwater', *Anuário do Instituto de Geociências*, 47:49797. https://doi.org/10.11137/1982-3908_2024_47_49797

