Eclética Química

Original article

revista.iq.unesp.br

| Vol. 48 | n. 4 | 2023 |

Multivariate statistical analysis of physicochemical parameters of groundwater quality using PCA and HCA techniques

Antonio José Ferreira Gadelha[1+], Clarice Oliveira da Rocha[1], José Germano Veras Neto[2], Mirelly Alexandre Gomes[2]

1. Federal Institute of Paraíba, Campina Grande Campus, Campina Grande, Brazil.
2. State University of Paraíba, Campina Grande Campus, Campina Grande, Brazil.

+Corresponding author: Antonio José Ferreira Gadelha, **Phone:** +55 (83) 2102-6200, **Email address:** antonio.gadelha@ifpb.edu.br

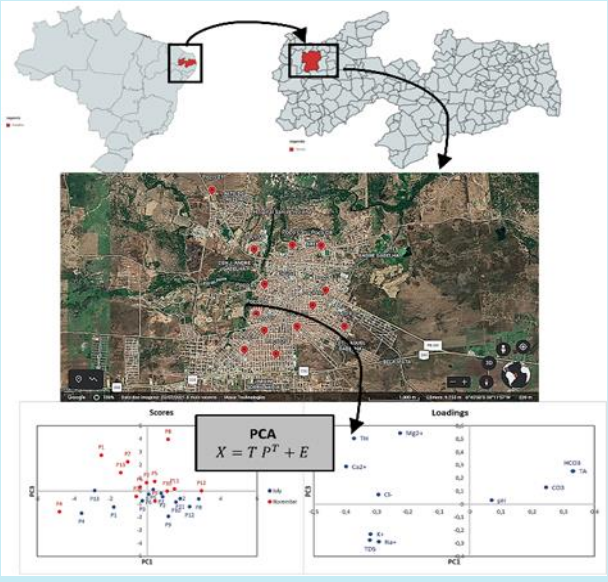
ARTICLE INFO

Article history:
Received: January 10, 2023
Accepted: August 13, 2023
Published: October 03, 2023

Section Editors: Boutros Sarrouh

Keywords:
1. chemometrics
2. hydric resources
3. Brazilian semiarid
4. exploratory analysis

ABSTRACT: Multivariate analysis techniques are powerful tools in the study of groundwater quality, providing an expanded view of quality parameters. This work presents a multivariate analysis of groundwater quality in the city of Sousa, Paraíba state, through the techniques of principal component analysis (PCA) and hierarchical cluster analysis (HCA). Samples from 13 tubular wells were collected in different districts of the city of Sousa, during the rainy and dry seasons. For these samples, 11 parameters were analyzed: hydrogenic potential (pH), total dissolved solids, total alkalinity, carbonates, bicarbonates, total hardness, magnesium, calcium, sodium, potassium, and chlorides. PC1, PC2, PC3 and PC4 explain 87.48% of the total variance of the data. The PCA shows that there was a change in patterns between the analyzed periods. The correlation matrix corroborates the PCA data, showing the relationships between the physical-chemical variables evaluated. The HCA confirmed the correlations between the samples, making it possible to assess the degree of similarity between the composition of the wells and between the parameters evaluated.



Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

1. Introduction

Water scarcity and poor distribution are the greatest obstacles for the socioeconomic development of the Brazilian semiarid region. According to Rossiter *et al.* (2020), in the northeastern semiarid, water does not have a uniform distribution, neither in time nor in space, and depends on climate vulnerability. The relationship between precipitation and evaporation presents a negative balance, resulting in long periods of drought throughout the year.

In regions with a semiarid climate, well drilling is the main activity to access water, an indispensable resource for life, industry, and agriculture. Deterioration in both quantity and quality of underground water poses a potential threat to urban and rural communities that depend on this resource.

Underground water quality largely depends on hydrochemical processes carried out through regional hydrogeological and anthropogenic activities under saturated and unsaturated soil conditions.

Some human activities that cause inadequate disposal of domestic sewage can significantly compromise the quality of underground water in shallow aquifers, which are quite vulnerable to contamination. Underground waters in urban areas are more susceptible to quality deterioration due to inappropriate use and occupation of the soil, the flow of various effluents discarded in the soil, and seasonality, which affects the recharge of these springs.

The analysis, monitoring and assessment of underground water quality parameters can be used to detect possible contamination events and as an indication of significant changes in physical and chemical properties of water, thereby avoiding financial and social losses. However, the assessment of a single parameter in isolation is not enough to describe the water quality, requiring a larger set of parameters and a multivariate analysis.

In this sense, multivariate statistical analysis techniques can be valuable tools to understand the characteristics and behavior of the aquifer that affect the quality control of underground water.

According to Taşan *et al.* (2022), the use of multivariate statistical methods together with geographic information systems contributes to the efficient management of water resources, planning and decision-making.

Carvalho *et al.* (2015) reported that, in recent years, the use of multivariate statistical methods, such as the principal component analysis (PCA) and hierarchical cluster analysis (HCA), have been frequently applied in numerous studies as a useful chemometric tool to extract a greater number of information obtained through the analysis of physicochemical and microbiological parameters, and metallic elements in samples of surface water, underground water, rain and minerals.

According to Gomes and Cavalcante (2017), by using the PCA technique, it is possible to select the characteristics with the highest participation in each component and define which physical-chemical parameters of the water should be monitored, thus reducing costs with analyses of factors of lesser importance in water quality.

In short, "PCA is an unsupervised pattern recognition method capable of transforming a set of experimental data into informative graphs about the similarity between samples and their respective variables" (Valderrama *et al.*, 2016, p. 245). According to Lyra *et al.* (2010), in this technique, a data matrix called matrix X is decomposed into a product of two other matrices, one called matrix of scores (T) and the other called matrix of loadings (P), such as Eq. 1:

$$X = TP^T + E \quad (1)$$

where P^T represents the transposed loadings matrix and E a residue matrix. The matrix of scores (T) presents information about the samples (rows of X, while the matrix of loadings (P) provides information about the variables (columns of X).

Importantly, several works have been developed in recent years with the purpose of applying exploratory data analysis tools to evaluate the hydrogeochemical characteristics of different water sources, including: Carvalho *et al.* (2015), Pan *et al.* (2019), Nnorom *et al.* (2019), Chaves *et al.* (2020) and Lopes *et al.* (2022).

Carvalho *et al.* (2015) investigated underground water samples from 17 locations distributed in the urban area of Belém, Pará state, Brazil. For all samples, seven physicochemical parameters and nine trace elements were evaluated. The PCA revealed the separation of two distinct groups of samples (A and B), due to the differences presented by the variables total dissolved solids (TDS), electrical conductivity and turbidity among the studied neighborhoods. The combination of principal components (PC) explained 85.7% of the total variance of the data, and PC1 (30.3%) and PC2 (22.4%) were the ones that most contributed to the discrimination of the samples.

Nnorom *et al.* (2019) examined the physicochemical and trace element contents of ground and surface water sources in the shale bedrock terrain of Southeastern Nigeria. A total of 124 water samples were collected from rural areas and analyzed for 21 elements. Different multivariate statistical approaches applied to assess the

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

origins of elements in water bodies identified six source types that accounted for 70.88% of the total variance. Anthropogenic activities were considered to contribute much of Cu, Pb, Cd, Cr, Li and P, while Al, As, Co, Fe, Se, Ni, Y and V were likely from crustal materials, minerals and ores, and natural environments. Cluster analysis was adopted to classify 124 sample points into two water pollution groups, reflecting influences from soil materials and anthropogenic sources.

Pan *et al.* (2019) conducted a study on the groundwater quality of the Condie Aquifer in Saskatchewan, Canada, where the Regina landfill was constructed without an engineered liner. An integrated statistical approach using PCA, correlation analysis, ion plots and multiple linear regression was used to evaluate groundwater contamination at the Regina landfill. Correlation analysis and ion plots pointed to gypsum and halite dissolution as the major factors affecting groundwater chemistry. PCA yielded three principal components, responsible for 80.7% of the total variance. A group analysis of the wells suggested possible groundwater contamination from the landfill operation. A two-step multiple linear regression was used to develop a model for predicting total hardness.

Chaves *et al.* (2020) studied the groundwater of 20 locations in Parauapebas, Pará state, Brazil, investigating nine physicochemical parameters in each sample. PCA and HCA revealed significant differences between the samples, being possible to observe the formation of two distinct groups (A and B). The most significant physicochemical parameters for separating the two groups were temperature, pH, electrical conductivity, color, chloride content, and TDS.

Lopes *et al.* (2022) studied the effects of the transfer of the São Francisco River on the performance of the water treatment plant of Gravatá, Paraíba state, Brazil. Using factor analysis combined with PCA (FA/PCA), the authors identified changes in the apparent color and turbidity of raw water, thus requiring interventions in the coagulation/flocculation/decantation processes. Moreover, by monitoring the volume of the Epitácio Pessoa reservoir, from January 2016 to December 2017, PCA and HCA exhibited the distinction of different phases in the water quality of the reservoir.

In this context, the present study aims to perform a multivariate analysis of the physicochemical quality parameters of underground water from tubular wells in the municipality of Sousa, Paraíba state, Brazil, to maximize the amount of information extracted and consequently, better interpret the correlations and similarities between samples and variables.

2. Experimental

The study was conducted in the urban area of the municipality of Sousa, Paraíba state, located in the semiarid zone of Brazilian northeastern region.

Samples from 13 tubular wells were collected from different districts of Sousa, as illustrated in Fig. 1 and described in Table 1, at the end of the rainy season (July) and the dry season (November). Samples were collected at the same time, starting at 7:00 am, and were packed in 500 mL plastic bottles, properly cleaned and sterilized, kept in thermal boxes during transport, cooled to 10 °C. The analyzes were carried out at the Chemistry Laboratory of the Campus Sousa, of the Instituto Federal de Educação, Ciência e Tecnologia da Paraíba.

Some wells drilled by the municipal government, over time, were out of operation, either because their water reserves were empty, or because of problems with the pumping system. As a result, the number of wells in operation during the execution of the study was reduced, limiting the number of wells sampled to 13. The points chosen by the municipal water department for drilling the wells included public areas such as squares, sidewalks, parks, etc. As the city is located on a crystalline subsoil, heterogeneity in both flow and depth in drilled wells is common.

For these samples, 11 physicochemical quality parameters were investigated: hydrogen potential (pH), TDS, total alkalinity (TA), carbonates (CO_3^{2-}), bicarbonates (HCO_3^-), total hardness (TH), magnesium (Mg), calcium (Ca), sodium (Na), potassium (K) and chlorides (Cl). Theses parameters are associated to the geochemical characteristics of the region's subsoil. The methodology used to determine the physicochemical parameters was conducted in accordance with the *Standard Methods for the Examination of Water and Wastewater*, 23 ed. (Baird *et al.*, 2017).

The parameters TA, TH, carbonates (CO_3^{2-}), bicarbonates (HCO_3^-), magnesium (Mg^{2+}), calcium (Ca^{2+}), sodium (Na^+), potassium (K^+) and chlorides (Cl^-) were determined by titration, in triplicate. The pH was measured using a pH meter (model pHB-500, Ion), and the TDS content was obtained using a multiparameter equipment (model Logen, LS).

For the multivariate statistical analysis, PCA, correlation matrix and HCA techniques were used, with the aid of The Unscrambler software, version 9.7, CAMO.

Initially, a data matrix was constructed with 26 lines (referring to the 13 wells sampled at two different times) and 11 columns (referring to the 11 parameters

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

evaluated). The first 13 lines refer to the data of the quality parameters at the end of the rainy season (July), and the last 13 lines refer to parameter data in the dry season (November). As a resource for preprocessing of the original data matrices, autoscaling was used. The autoscaling technique was chosen because the values of the evaluated parameters show distributed values in different ranges, and it is necessary that the same weight be assigned to each variable under study.

PCA was used to explore the association between parameters that influence groundwater quality, reducing the number of variables and verifying which variables or sets of variables explain most of the total variability, showing the relationships between them. As a result, two-dimensional plots of scores and loadings are obtained, allowing a better visualization of the distribution of experimental data and the relationships between variables and between samples (Souza and Poppi 2012).

Regarding HCA, its use consists of analyzing a set of data through hierarchically defined groups, verifying the similarity between variables or samples, as a complementary strategy to PCA (Mingoti, 2007). In this study, HCA was applied to the autoscaled data, and the measure of similarity chosen was the squared-Euclidean distance and as a hierarchical clustering criterion, the Single-Linkage method was used, which considers the total sum of deviations of each object in relation to the group mean. From the HCA calculations, the dendrograms of the samples and variables were generated, allowing to identify the degree of similarity between the groups.

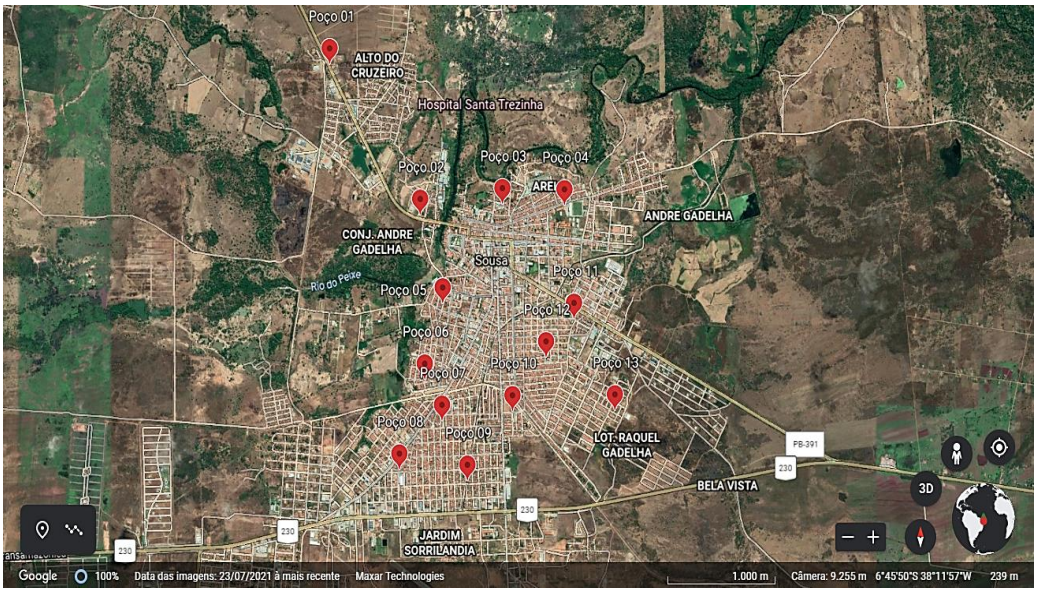


Figure 1. Geolocation of the sampled wells.
Source: Adapted from Google Earth, 2021.

Table 1. Technical information of the sampled wells.

Well	Drilling date	Depth (m)	Flow (l/h)	Location	Localization
01	06/09/2015	51	1,500	Alto do Cruzeiro	–6.74605, –38.24398
02	05/28/2015	50	600	Várzea da Cruz (PSF)	–6.75668, –38.23644
03	05/24/2015	51	2,000	Guanabara (PSF)	–6.75585, –38.2295
04	05/25/2015	51	1,800	Estádio Marizão	–6.75594, –38.22425
05	07/20/2015	41	5,538	Alto Capanema	–6.7632, –38.23455
06	08/08/2015	49	10,000	Jardim Santana	–6.76883, –38.23607
07	05/27/2015	50	600	Estação (PSF)	–6.77187, –38.23461
08	07/24/2015	40	3,130	Jardim Bela Vista (Varejão)	–6.77547, –38.23821
09	06/23/2015	51	1,142	Jardim Sorrilândia II	–6.77631, –38.23246
10	08/18/2015	50	2,300	Casas Populares (CSU)	–6.77116, –38.22865
11	07/20/2015	51	1,400	Condomínio Doca Gadelha	–6.76434, –38.22344
12	06/19/2015	51	700	Conjunto Dr. Zezé	–6.76718, –38.22581
13	06/06/2015	51	5,200	Alto do DNOCS	–6.7711, –38.21999

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

3. Results and discussion

The average values of the results corresponding to the 11 physicochemical parameters determined in the 13 examined wells, at the end of the rainy season, are described in Table 2, while data for the dry season are shown in Table 3.

Table 2. Results of physicochemical analysis at the end of the rainy season (July).

	Parameters										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	TA	CO_3^{2-}	HCO_3^-	TH	Mg^{2+}	Ca^{2+}	Cl^-	Na^+	K^+	pH	TDS
Well 01	129.60	0.52	157.05	171.00	15.75	47.29	493.44	494.70	1.81	7.83	1027.0
Well 02	355.20	1.68	429.92	124.20	21.00	16.83	172.58	212.10	1.77	7.90	490.0
Well 03	276.00	1.80	333.06	144.00	12.68	40.88	397.69	489.60	1.08	8.04	987.0
Well 04	412.80	2.69	498.14	230.40	23.18	60.12	480.23	1781.80	4.10	8.04	2713.0
Well 05	590.40	1.31	717.62	156.60	15.75	40.88	459.22	504.90	2.20	7.57	993.0
Well 06	396.00	1.01	481.06	163.80	15.75	44.09	166.58	159.60	2.60	7.63	453.0
Well 07	508.80	1.42	617.84	136.80	10.94	40.88	282.13	316.20	2.00	7.67	688.0
Well 08	616.80	4.41	743.54	39.60	5.25	8.02	315.15	453.90	1.19	8.08	788.0
Well 09	511.20	4.58	614.34	34.20	2.62	10.42	472.73	1080.70	1.49	8.18	1723.0
Well 10	528.00	1.23	641.66	77.40	7.44	20.84	184.59	337.90	1.96	7.59	613.0
Well 11	626.40	1.92	760.30	73.80	9.19	16.03	342.16	515.10	1.70	7.71	921.0
Well 12	494.40	6.08	590.80	34.20	3.50	8.82	316.65	617.10	1.07	8.32	1038.0
Well 13	376.80	0.86	457.95	243.00	20.12	71.34	997.98	766.80	2.50	7.58	1497.0
Minimum	129.60	0.52	157.05	34.20	2.62	8.02	166.58	159.60	1.07	7.57	453.0
Maximum	626.40	6.08	760.30	243.00	23.18	71.34	997.98	1781.80	4.10	8.32	2713.0
Average	447.88	2.27	541.79	125.31	12.55	32.80	390.86	594.65	1.96	7.86	1071.6
Standard deviation	143.01	1.70	173.09	70.01	6.75	20.69	216.88	429.15	0.81	0.25	612.2

Units: TA and TH (mg $CaCO_3$ L^{-1}); CO_3^{2-}, HCO_3^-, Mg^{2+}, Ca^{2+}, Cl^-, Na^+ and K^+ (mg L^{-1}); pH (unit); TDS (ppm).

Table 3. Results of physicochemical analysis during the dry season (November).

	Parameters										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	TA	CO_3^{2-}	HCO_3^-	TH	Mg^{2+}	Ca^{2+}	Cl^-	Na^+	K^+	pH	TDS
Well 01	355.20	2.54	428.18	392.40	50.74	81.76	580.78	749.70	1.03	8.08	1073.0
Well 02	388.80	3.18	467.86	212.40	29.74	40.08	220.61	327.60	2.20	8.14	578.2
Well 03	259.20	2.43	311.28	165.60	10.50	54.51	556.02	392.70	1.03	8.20	1000.0
Well 04	259.20	2.27	311.60	273.60	17.50	89.78	675.32	1615.70	3.20	8.17	3162.0
Well 05	648.00	2.87	784.73	181.80	9.19	64.13	589.78	423.30	2.00	7.87	1035.0
Well 06	460.80	0.94	560.27	208.80	3.06	87.37	240.87	136.50	2.50	7.53	461.7
Well 07	494.40	2.09	598.92	322.20	33.68	81.76	366.93	270.30	2.00	7.85	763.6
Well 08	717.60	7.89	859.43	327.60	55.55	44.09	378.18	408.00	0.99	8.27	878.1
Well 09	516.00	6.50	616.31	117.00	7.44	38.48	578.53	949.40	1.50	8.33	1847.0
Well 10	561.60	2.79	679.49	135.00	5.69	49.70	312.90	365.80	1.92	7.92	718.9
Well 11	679.20	3.69	821.12	126.00	0.87	54.51	429.96	515.10	1.57	7.96	1025.0
Well 12	636.00	10.25	755.07	75.60	1.75	30.46	357.92	576.30	0.91	8.44	1070.0
Well 13	566.40	2.87	685.16	255.60	24.06	69.74	1040.00	674.50	1.79	7.93	1367.0
Minimum	259.20	0.94	311.28	75.60	0.87	30.46	220.61	136.50	0.91	7.53	461.7
Maximum	717.60	10.25	859.43	392.40	55.55	89.78	1040.00	1615.70	3.20	8.44	3162.0
Average	503.26	3.87	606.11	214.89	19.21	60.49	486.75	569.61	1.74	8.05	1152.3
Standard deviation	152.49	2.67	183.22	94.72	18.40	20.09	220.19	380.91	0.67	0.24	698.4

Units: TA and TH (mg $CaCO_3$ L^{-1}); CO_3^{2-}, HCO_3^-, Mg^{2+}, Ca^{2+}, Cl^-, Na^+ and K^+ (mg L^{-1}); pH (unit); TDS (ppm).

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

Notably, similar values in magnitude were obtained for the physicochemical parameters of groundwater quality by Jiang *et al.* (2015), in Mongolia, Taşan *et al.* (2022), in Turkey, and Chaves *et al.* (2020), in Pará state, Brazil.

The results of the multivariate analysis are hereafter presented. As PCA causes a change in the vector space of the data set, each object (each of the 13 sampled wells) that was represented in a space with 11 variables (11 physicochemical parameters) is represented by 11 principal components. Figure 2 illustrates the variances explained by the first 7 accumulated principal components referring to the data collected in the months of July and November. Since the first principal components are responsible for most of the data variance, it is possible to concentrate the analysis focusing on a smaller number of variables, without a significant loss of information being observed. In this case, the first four principal components explain 87.48% of the total variance, which were chosen because they have eigenvalues greater than 1.0.

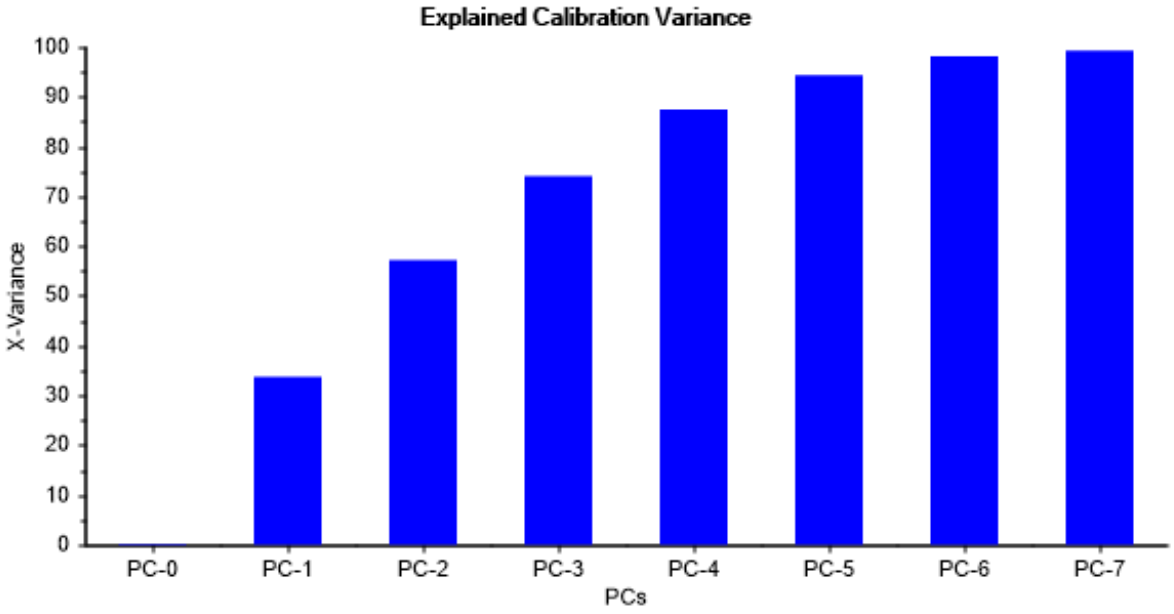


Figure 2. Cumulative variance explained by the principal components.

Figures 3 and 4 show the score and loading plots, respectively, for PC1 and PC3, for the months of July and November. According to Lyra *et al.* (2010), the scores are projections of the original objects in the space of the principal components, that is, they are the new coordinates of the objects in the new variables. In this case, the score plot indicates the relationships between the sampled wells. Furthermore, the authors state that the loadings, or weights, geometrically represent the cosines of the angles that the principal components make with the original variables. In this case, in the loading plot, we can observe the relationships between the variables, that is, between the physicochemical parameters evaluated.

Some important patterns can be noticed in the score plot in Fig. 3. For example, the wells 1, 4 and 13 are located on the negative side of PC1, isolated from the other wells, hence indicating that they differ significantly from the others in terms of their chemical composition. This can be verified in Tables 2 and 3, which reveal the high concentration of salts in these three samples.

Another arrangement of scores that draws considerable attention refers to the wells 2, 3, 5, 6 and 7, which are concentrated in the same region, indicating that these samples have some similarity in their composition. This fact can be associated with the proximity of these wells to the bed of the Rio do Peixe that passes through the urban area of the municipality. This proximity may be promoting a more uniform recharge in these wells in comparison to the others.

It is also visible the pattern changes between the sampled wells in relation to the period of the year in which the analysis was conducted. Figure 3 clearly illustrates that the data from the wells sampled in July, at the end of the rainy season, are concentrated on the negative side of PC3. The data from the wells sampled in November (drought period) are on the positive side of PC3. This finding indicates a substantial change in the composition of groundwater in the studied area in relation to the time of collection in the year. Comparing with the loadings plot in Fig. 4, it can be inferred that the

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

data from the analysis for the month of July are concentrated on the negative side of PC3, as well as the variables Na^+, K^+ and TDS. This can be explained because, during the rainy season, there is a greater percolation of water in the soil, causing a greater dissolution of sodium and potassium salts, increasing the concentration of these ions in the water. With respect to the data for the month of November, the scores are concentrated on the positive side of PC3, as well as the parameters related to alkalinity and hardness of the water, which are parameters more related to the rock in which the groundwater is deposited. With the lack of recharge during the dry season, the ions responsible for alkalinity and hardness tend to be concentrated during this period.

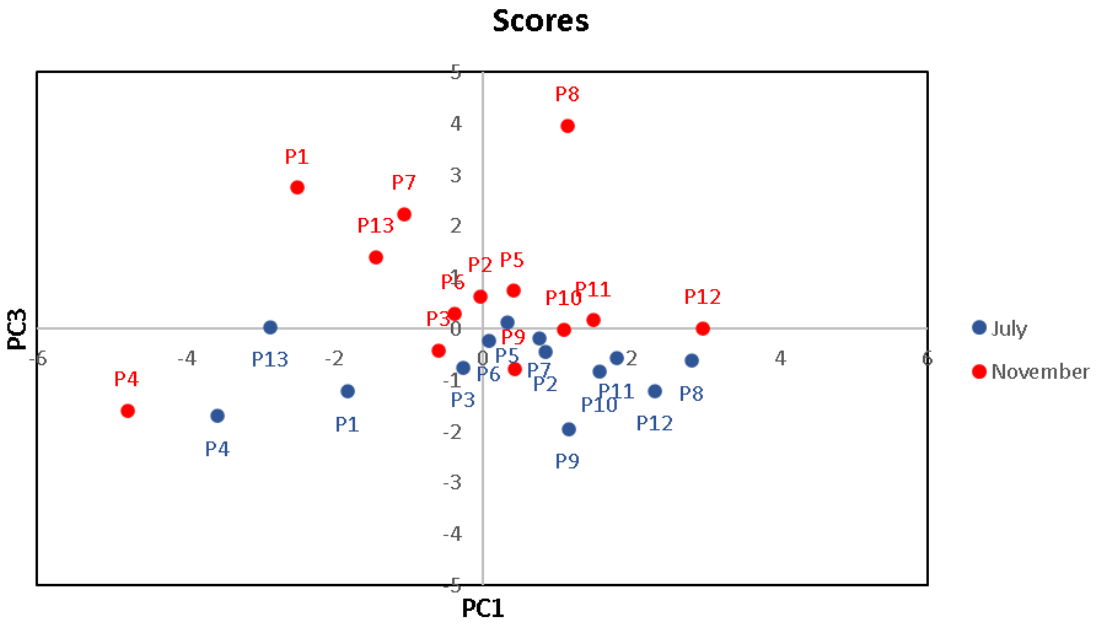


Figure 3. Plot of PCA scores of the wells sampled in July and November.

In the PCA loadings plot (Fig. 4), correlations between variables are easily detected. It is the case of TA, concentrations of carbonate ions (CO_3^{2-}), bicarbonates (HCO_3^-) and pH, indicating that the hydrogenic potential of the water of the wells throughout the year is mainly determined by the content of carbonate and bicarbonate ions, which, due to the basic character, are responsible for the alkalinity, maintaining the pH above 7.0, as recorded in Tables 2 and 3. It is worthwhile noting the close relationship between total alkalinity and bicarbonate ion concentration (HCO_3^-), confirming that most of the alkalinity of these wells is owing to the presence of the HCO_3^- ion.

Another pattern observed is the relationship of similitude between sodium ion and TDS content, indicating that this ion is the main responsible for the high TDS values.

Additionally, the group of variables total hardness, calcium and magnesium are also very close in the loadings plot, thus implying that they are correlated. Indeed, the definition of total hardness is intrinsically associated with the concentrations of divalent metal cations (Harris, 2017).

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

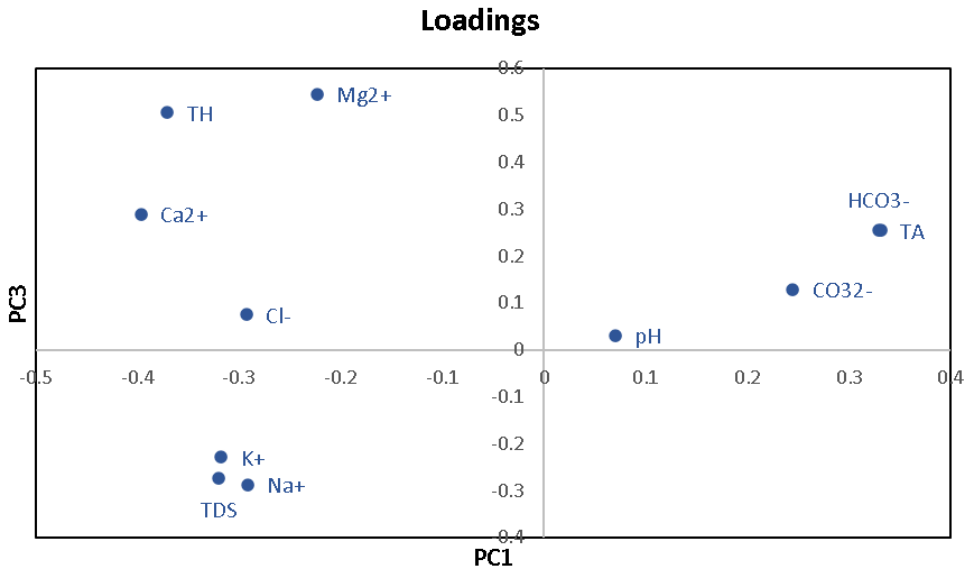


Figure 4. Plot of PCA loadings of physicochemical parameters in the months of July and November.

The correlation coefficient matrix between the physicochemical parameters evaluated is depicted in Table 4. The data from the correlation matrix evidence the similarities between variables presented in the PCA loading plot. There is a very significant correlation between the concentration of bicarbonate ions (HCO_3^-) and TA, reinforcing that this ion is the leading cause of the alkaline character of the samples. There are also strong positive correlations ($r > 0.7$) for the pairs CO_3^{2-} and pH, Na^+ and TDS, corroborating the PCA data.

A strong correlation between the concentrations of calcium (Ca^{2+}) and magnesium (Mg^{2+}) ions with the total hardness was also observed, which agrees with the PCA loading plot, emphasizing that the hardness is caused by the presence of multivalent metallic cations. This correlation also suggests the dissolution of calcite ($CaCO_3$) and dolomite [$CaMg(CO_3)_2$] in these aquifers, the main components of sedimentary rocks (Celestino *et al.*, 2018). Similar findings were reported by Charfi *et al.* (2013).

Table 4. Correlation matrix for the physicochemical parameters evaluated.

	TA	**CO_3^{2-}**	**HCO_3^-**	**TH**	**Mg^{2+}**	**Ca^{2+}**	**Cl^-**	**Na^+**	**K^+**	**pH**	**TDS**
TA	**1**										
CO_3^{2-}	0.5057	**1**									
HCO_3^-	**0.9997**	0.4856	**1**								
TH	–0.2287	–0.1955	–0.2265	**1**							
Mg^{2+}	–0.1171	–0.0066	–0.1185	**0.8397**	**1**						
Ca^{2+}	–0.2665	–0.3260	–0.2613	**0.8222**	0.3813	**1**					
Cl^-	–0.1126	–0.0369	–0.1131	0.3906	0.1721	0.4846	**1**				
Na^+	–0.1816	0.1502	–0.1880	0.1407	0.0463	0.1910	0.4746	**1**			
K^+	–0.2342	–0.4922	–0.2242	0.2553	0.0124	0.4216	0.0960	0.4623	**1**		
pH	0.0383	**0.8060**	0.0172	–0.0621	0.0927	–0.2030	0.0738	0.3757	–0.4316	**1**	
TDS	–0.2125	0.1243	–0.2187	0.1858	0.0214	0.2939	0.5521	**0.9709**	0.4720	0.3589	**1**

The observed results can be explained from the geological aspects of the sedimentary basin of the Rio do Peixe, described by Galvão *et al.* (2005), where the studied area is located, which is constituted by rocks of the Precambrian Crystalline Complex. This group is predominantly made up of clayey rocks (claystones and shales), which resulted in the formation of thin layers of salts in the form of films (mainly Ca^{2+}, Na^+, K^+ associated with HCO_3^- and CO_3^{2-}).

Figure 5 illustrates the dendrogram of the HCA of the data from the wells collected during the dry season. According to Ferreira (2015), the main purpose of HCA is to gather samples in such a way that those belonging to the same group are more similar to each other than to samples from other groups.

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

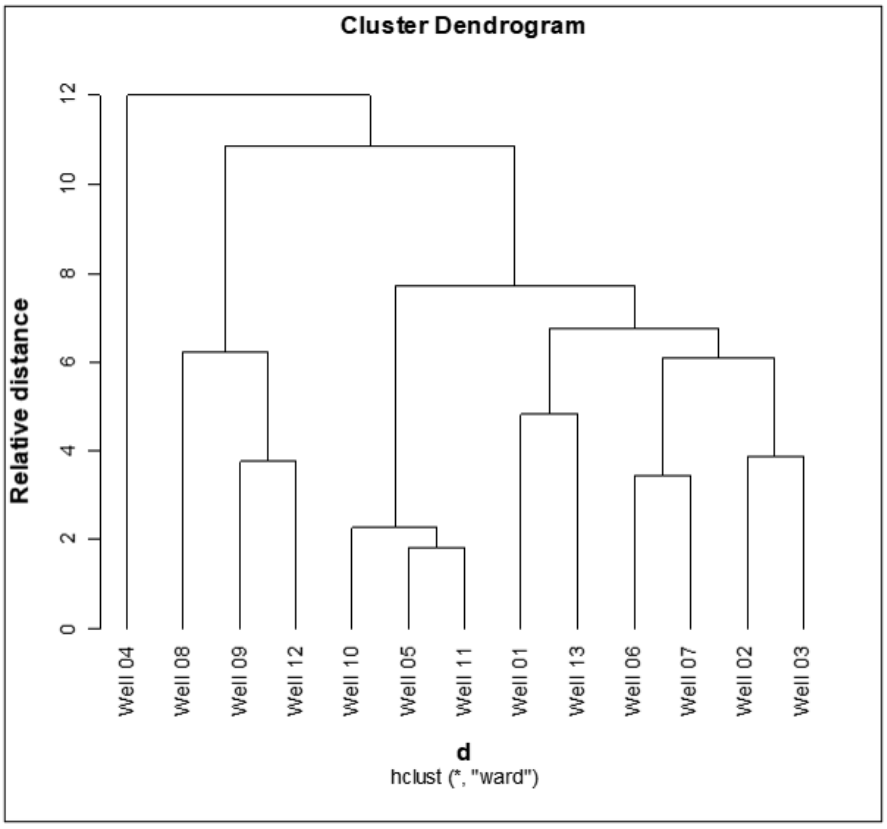


Figure 5. HCA dendrogram of the dry season for the sampled wells.

Figure 5 demonstrates some patterns of similarity between the samples. The wells 05, 11 and 10 present a high degree of similarity, as can be seen from the small relative distance between them in the dendrogram. This result implies that these samples have a uniform composition among themselves, which is corroborated by the PCA scores plot (Fig. 3), in which these wells are very close to each other.

It is also noteworthy the high relative distance of the wells 08 and 04 in relation to the others, suggesting that the waters of these points have a very different composition from the other wells. This result can be confirmed through Fig. 3. In relation to the other samples, from the analysis of the HCA dendrogram, it is verified that they present a certain degree of similarity to each other.

4. Conclusions

The findings of this study revealed that the waters of the tubular wells in the urban area of the municipality of Sousa have high levels of bicarbonate, chloride, and sodium ions, which also increases the values of total alkalinity and TDS, thereby making the consumption of some samples unfeasible, and limiting the use of these waters only for domestic cleaning activities.

PCA showed that there was a change in the patterns between the analyzed periods, July, and November, which proves the influence of the rainy season on the recharge and water composition of these wells. The first four principal components explain 87.48% of the total variance of the data. PCA also demonstrated that some parameters are well correlated, such as alkalinity, pH, HCO_3^- and CO_3^{2-}; total hardness, Ca^{2+} and Mg^{2+}; TDS, Na^+ and K^+. The correlation matrix corroborates the PCA data, showing the relationships between the physicochemical variables evaluated. HCA confirmed the correlations between the samples, thus allowing to assess the degree of similarity between the composition of the wells and between the parameters evaluated. From the HCA, it can be verified that the wells 04 and 08 have very different compositions from the others.

A possible limitation of this work was the time interval used for data gathering, which was performed in a single year. Collecting data over several years would provide a broader view of water quality and possibly provide other relationships between variables and wells in the multivariate analysis. In this context, it would also be possible to use supervised statistical analysis techniques, which use group discrimination criteria.

Eclética Química, vol. 48, n. 4, 2023, 37-47
ISSN: 1678-4618
DOI: 10.26850/1678-4618eqj.v48.4.2023.p37-47

Authors' contribution

Conceptualization: Gadelha, A. J. F.; Veras, G.;
Data curation: Gadelha, A. J. F.; Veras, G.; Rocha, C. O.;
Formal Analysis: Gadelha, A. J. F.; Veras, G.; Rocha, C. O.;
Funding acquisition: Not applicable;
Investigation: Gomes, M. A.; Gadelha, A. J. F.;
Methodology: Gadelha, A. J. F.; Veras, G.; Gomes, M. A.;
Project administration: Gadelha, A J. F.;
Resources: Gadelha, A. J. F.; Rocha, C. O.;
Software: Not applicable;
Supervision: Gadelha, A. J. F.; Veras, G.;
Validation: Gadelha, A. J. F.; Veras, G.;
Visualization: Gadelha, A. J. F.; Veras, G. Rocha, C. O.;
Writing – original draft: Gadelha, A. J. F.; Veras, G.; Rocha, C. O.; Gomes, M. A.;
Writing – review & editing: Gadelha, A. J. F.; Veras, G.; Rocha, C. O.

Data availability statement

All data sets were generated or analyzed in the current study.

Funding

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Grant No: 313579/2021-0

Instituto Nacional de Ciência e Tecnologia em Ciências Moleculares (INCT-CiMol). Grant No: 406804/2022-2.

Acknowledgments

Not applicable.

References

Baird, R. B.; Eaton, A. D. & Rice, E. W. (Eds.) Standard Methods for the Examination of Water and Wastewater. 23th ed. American Public Health Association, American Water Works Association, Water Environment Federation, Washington. 2017.

Carvalho, F. I. M.; Lemos, V. P.; Dantas Filho, H. A.; Dantas, K. G. F. Avaliação da Qualidade das Águas Subterrâneas de Belém a partir de Parâmetros Físico-Químicos e Níveis de Elementos Traço Usando Análise Multivariada. *Rev. Virtual Quím.* **2015**, *7* (6), 2221–2241. https://doi.org/10.5935/1984-6835.20150132

Celestino, A. E. M.; Cruz, D. A. M.; Sánchez, E. M. O.; Reyes, F. G.; Soto, D. V. Groundwater Quality Assessment: An Improved Approach to K-Means Clustering, Principal Component Analysis and Spatial Analysis: A Case Study. *Water*. **2018**, *10* (4), 437. https://doi.org/10.3390/w10040437

Chaves, H. S.; Morais, D. G.; Dantas Filho, H. A.; Dantas, K. G. F.; Beirao, A. T. M.; Silva, K.P.; Silva, J. N.; Silva, V. F. A.; Silva, P. A.; Carvalho, F. I. M. Aplicação estatística multivariada para a avaliação físico-química na qualidade da água subterrânea na cidade de Parauapebas (Sudeste do Estado do Pará). *Revista Ibero Americana de Ciências Ambientais.* **2020**, *11* (5), 261–272. https://doi.org/10.6008/CBPC2179-6858.2020.005.0025

Charfi, S.; Zouari, K.; Feki, S.; Mami, E. Study of variation in groundwater quality in a coastal aquifer in north-eastern Tunisia using multivariate factor analysis. *Quat. Int.* **2013**, *302*, 199–209. https://doi.org/10.1016/j.quaint.2012.11.002

Ferreira, M. M. C. *Quimiometria*: Conceitos, métodos e aplicações. Editora da Unicamp, 2015. https://doi.org/10.7476/9788526814714

Galvão, M. J. T. G.; Costa Filho, W. D.; Srinivasan, V. S.; Schuster, H. D. M.; Rego, J. C.; Albuquerque, J. P. T. *Comportamento das bacias sedimentares da região semiárida do Nordeste brasileiro. Hidrogeologia da Bacia Sedimentar do Rio do Peixe*. CPRM/UFCG/FINEP, 2005.

Gomes, M. C. R.; Cavalcante, I. N. Aplicação da análise estatística multivariada no estudo da qualidade da água subterrânea. *Águas Subterrâneas.* **2017**, *31* (1), 134–149. https://doi.org/10.14295/ras.v31i1.28617

Harris, D. C. *Análise Química Quantitativa*. LTC, 2017.

Jiang, Y.; Guo, H.; Jia, Y.; Cao, Y.; Hu, C. Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia. *Geochemistry*. **2015**, *75* (2), 197–205. https://doi.org/10.1016/j.chemer.2014.12.002

Lyra, W. S.; Silva, E. C.; Araújo, M. C. U; Fragoso, W. D.; Veras, G. Classificação periódica: um exemplo didático para ensinar análise de componentes principais.

Quím. Nova. **2010**, *33* (7), 1594–1597. https://doi.org/10.1590/S0100-40422010000700030

Lopes, T. S. A.; Santos, W. B.; Silva, G. A. B.; Silveira, T. N.; Ferreira, W. B.; Feitosa, P. H. C.; Lima, V. L. A. Effects of the transfer of the São Francisco River waters on the performance of the water treatment plant of Gravatá, Paraíba, Brazil. *Water Supply.* **2022**, *22* (3), 3297–3306. https://doi.org/10.2166/ws.2021.404

Mingoti, S. A. *Análise de dados através de métodos de estatística multivariada*. Editora da UFMG, 2007.

Nnorom, I. C.; Ewezie, U.; Eze S. O. Multivariate statistical approach and water quality assessment of natural springs and other drinking water sources in Southeastern Nigeria. *Heliyon.* **2019**, *5* (1), e01123. https://doi.org/10.1016/j.heliyon.2019.e01123

Pan, C.; Ng, K. T. W.; Richter, A. An integrated multivariate statistical approach for the evaluation of spatial variations in groundwater quality near an unlined landfill. *Environ Sci Pollut Res* **2019**, *26* (6), 5724–5737. https://doi.org/10.1007/s11356-018-3967-x

Rossiter, K. W. L.; Marques, E. A. T.; Oliveira, C. R.; Morais, M. M. Q. M. M. Spatial-temporal evaluation of water quality in Brazilian semiarid reservoirs. *Water Pract. Technol.* **2020**, *15* (1), 92–104. https://doi.org/10.2166/wpt.2020.001

Souza, A. M.; Poppi, R. J. Didactic chemometrics experiment for exploratory analysis of edible vegetable oils by mid-infrared spectroscopy and principal component analysis: a tutorial, part I. *Quím. Nova.* **2012**, *35* (1), 223–229. https://doi.org/10.1590/S0100-40422012000100039

Taşan, M.; Demir, Y.; Taşan, S. Groundwater quality assessment using principal component analysis and hierarchical cluster analysis in Alaçam, Turkey. *Water Supply.* **2022**, *22* (3), 3431–3447. https://doi.org/10.2166/ws.2021.390

Valderrama, L.; Paiva, V. B.; Março, P. H.; Valderrama, P. Proposta experimental didática para o ensino de análise de componentes principais. *Quím. Nova.* **2016**, *39* (2), 245–249. https://doi.org/10.5935/0100-4042.20150166