




## Cluster and Factor Analyses as Contributions to the Groundwater Quality Monitoring of the Marizal/São Sebastião Aquifer System, Alagoinhas (Bahia, Brazil)

*Análises de Cluster e Fatorial como Subsídio ao Monitoramento da Qualidade das Águas Subterrâneas do Sistema Aquífero Marizal/São Sebastião, Alagoinhas (Bahia, Brasil)*

Maíra Sampaio da Costa<sup>1</sup> , Maria da Conceição Rabelo Gomes<sup>2</sup>   
& Sérgio Augusto de Moraes Nascimento<sup>2</sup> 

<sup>1</sup>Universidade Federal da Bahia, Instituto de Geociências, Núcleo de Estudos Hidrogeológicos e do Meio Ambiente, Salvador, BA, Brasil

<sup>2</sup>Universidade do Estado do Pará, Departamento de Engenharia Ambiental e Sanitária, Marabá, PA, Brasil

E-mails: mairasampaio42@gmail.com; conceicaorabelo@yahoo.com.br; sergiomn@ufba.br

**Corresponding author:** Maíra Sampaio da Costa; mairasampaio42@gmail.com

### Abstract

The Marizal/São Sebastião aquifer system is the main water supply of the municipality of Alagoinhas in the state of Bahia. However, anthropic interventions contribute to soil and groundwater pollution, increasing the need for related research. Multivariate statistical analysis is a widely used tool, helping in the investigation of groundwater quality while being capable of simultaneously evaluating diverse variables of a sample set. In this study, factor analysis and multivariate cluster analysis methodologies were applied. Ten of the most influential variables for groundwater quality were selected and then grouped into two factors. The first factor included electrical conductivity, salinity, calcium, chloride, sulfate, manganese, and iron, which are indicators of water salinity. The second factor encompassed pH, bicarbonate, and phosphate, indicating anthropic interventions and alkalinity in the environment. The multivariate cluster analysis was applied to the parameters of both factors, resulting in dendrograms with four clusters. The present study showed that the multivariate statistical analysis is an efficient tool for monitoring and can contribute to the management of groundwater quality.

**Keywords:** Sedimentary aquifer; Factor analysis; Cluster analysis

### Resumo

O sistema aquífero Marizal/São Sebastião é a principal fonte de água do município de Alagoinhas no Estado da Bahia. Entretanto, as intervenções antrópicas podem contribuir para a poluição do solo e das águas subterrâneas, elevando a necessidade de pesquisas sobre o tema. Uma ferramenta amplamente utilizada para auxiliar na investigação da qualidade das águas é a análise estatística multivariada, que é capaz de avaliar simultaneamente diversas variáveis de um conjunto amostral. Neste estudo foram aplicadas a análise fatorial e a análise de agrupamento multivariada. Foram selecionadas as dez variáveis mais influentes para a qualidade das águas subterrâneas e agrupadas em dois fatores. O primeiro fator potencializou condutividade elétrica, salinidade, cálcio, cloreto, sulfato, manganês e ferro, que são indicadores de salinidade na água. O segundo fator potencializou pH, bicarbonato e fosfatos, que são indicadores de alcalinidade do meio e atividades antrópicas. A análise de agrupamento multivariada foi aplicada nos parâmetros dos dois fatores, apresentando como resultado dendrogramas com formação de quatro grupos. Esse estudo mostrou que a análise estatística multivariada é uma ferramenta eficaz no monitoramento e pode contribuir na gestão da qualidade das águas subterrâneas.

**Palavras-chave:** Aquífero sedimentar; Análise fatorial; Análise de agrupamento

# 1 Introduction

The Marizal/São Sebastião aquifer system presents a high hydrogeological potential and is considered the main source of water supply for the municipality of Alagoínhas (Moraes et al. 2004; Nascimento et al. 2006). However, urban and rural occupation contribute to soil and groundwater pollution, which have mobilized researchers from different fields of knowledge into working with methods capable of quantitatively and qualitatively characterize the hydrogeological environment. In general, these studies enable a better management of the water resources and land use and occupation.

In this approach, multivariate statistical analysis is presented as a widely used tool. It comprises a set of statistical procedures capable of simultaneously evaluating variables from a sample set, gathering and associating similar components, and also investigating their interdependence and testing hypotheses (Vicini 2005; Ferreira 1996).

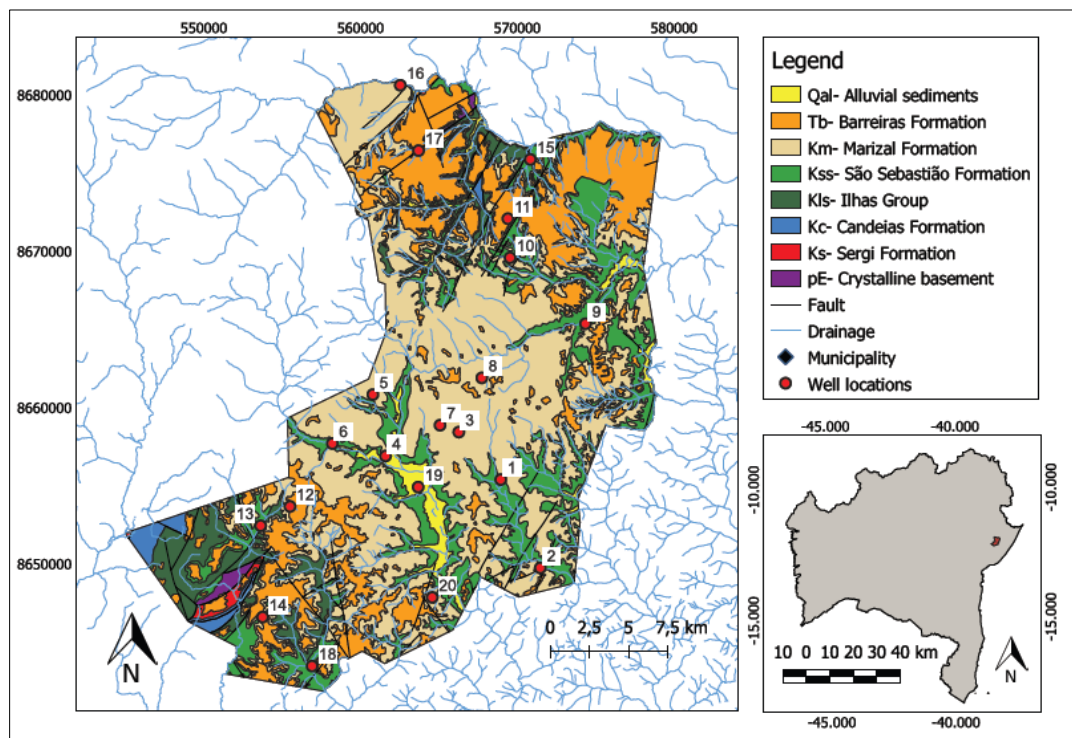
When studying the similarity between variables, two methodologies are of particular interest: the factor analysis (FA), using a principal components analysis (PCA), and the hierarchical cluster analysis (HCA). The first methodology reduces the number of original variables (factors) into a smaller set, with minimal loss of information; the latter groups a sample set considering the similarities between them.

By applying a multivariate analysis, it is possible to select the parameters that best characterize groundwater and define the physicochemical characteristics that should be monitored, reducing costs by excluding less relevant analyses. In this context, Gomes et al. (2017), Gomes and Cavalcante (2017), Gomes and Franca (2019), and Costa et al. (2020) used the FA and HCA techniques as contributions to the management of groundwater in the state of Ceará, Brazil, a semiarid region that presents high environmental vulnerability, similar to the present study area.

The objective of the present study was to select and identify the similarity of determinant variables for groundwater quality in the Marizal/São Sebastião aquifer system in the municipality of Alagoínhas, state of Bahia, Brazil, based on factor analysis and hierarchical cluster analysis methodologies.

## 1.1 Geological Setting

The study area is located in the northern part of the Recôncavo Sedimentary Basin, Bahia, Brazil, featuring the Marizal/São Sebastião Aquifer System, in the municipality of Alagoínhas (Figure 1). The Marizal Formation (Km) aquifer section is classified as unconfined and consists mainly of sandstones and conglomerates, and secondly of siltstones, shales, and limestones, deposited in an



**Figure 1** Geology of the area and location of wells analyzed (white squares with numbers). UTM projection - 24 S zone, SIRGAS 2000 Horizontal Datum. Source: modified geology map from Petrobrás (1969)

environment of alluvial fans and braided fluvial systems (Ribeiro 2008; Nascimento & Alves 2014).

The aquifer portion representing the São Sebastião Formation (Kss) is classified as unconfined and semi-confined and consists of three members divided as Joanes River (upper member), Passagem dos Teixeiras (medium member), and Paciência (lower member). It comprises gray-yellowish, pinkish, and yellow-reddish sandstones, with particle sizes varying from fine, to medium, and to coarse-grained. Sandstones are massive, show tabular geometry and low angle plane-parallel stratification and/or trough cross-stratification (Nascimento & Alves 2014; Alves 2015).

## 2 Methodology and Data

Twenty groundwater well samples from the Marizal/São Sebastião aquifer system were evaluated (Table 1). They were collected at the end of the rainy season (August 2021), with field support from SAAE (Water and Sewage Service Provider from Alagoinhas). Samples were chosen based on the wells that collect water from the Marizal/São Sebastião Formation and their spatial distribution.

In total, 18 physicochemical parameters were used in the study: pH, electrical conductivity (EC), salinity, calcium, sodium, magnesium, potassium, nitrate, chloride, bicarbonate, sulfate, fluoride, phosphate, manganese, aluminum, iron, copper, and lead. These parameters were selected considering the relevance of the major cations and anions in groundwater classification and minor elements listed in the contamination area, according to Nascimento et al. (2006), Pereira and Lima (2007), Nascimento and Alves (2014), and Alves (2015).

Physicochemical parameters were analyzed in the ALS (*Life Sciences Brazil*) laboratory and statistically processed using the software *SPSS Statistics* 17.0 version, *Excel*, and *PHREEQC Interactive*.

A factor analysis (FA) and a hierarchical cluster analysis (HCA) were applied using *SPSS Statistics software*. FA aims to reduce the number of initial variables, with the lowest possible loss of information, and group them into basic categories called factors. The analysis calculates the linear combination of the original variables, and then indicates a correlation between a new factor set. Together with the FA, a Principal Component Analysis (PCA) and a normalized varimax rotation were used. In the varimax rotation method, the variables are converted into standard scores by the *Z scores* method, in which data is reassigned to a simplified position on the Cartesian axis, where new factors have minimized loadings (Vicini 2005; Araújo et al. 2013).

Data correlation was subjected to the Bartlett sphericity test and KMO (*Kaiser–Meyer–Olkin Measure of Sampling Adequacy*). The Bartlett sphericity test assesses the null hypothesis regarding the absence of correlation between variables. The KMO evaluates the input value of the variables, which ranges between 0.5 and 0.9 and can be used to perform the FA (Hair Jr et al. 1998; Araújo et al. 2013; Franca et al. 2018). The KMO value is calculated according to Equation 1 as:

$$KMO = \frac{\sum_{i,j} r_{ij}^2}{\sum_{i,j} r_{ij}^2 + \sum_{i,j} a_{ij}^2} \quad (1)$$

where  $r_{ij}$ : correlation coefficient between  $i$  and  $j$  variables;  $a_{ij}$ : partial correlation coefficient between  $i$  and  $j$  variables

The Hierarchical Cluster Analysis (HCA) aims to find and sort objects into similar clusters. It was conducted considering the maximum number of variables strongly explained by one single factor. This was achieved through the Ward method, and the squared Euclidean distance was adopted as a measure of similarity. According to Vicini (2005), the Euclidean distance is obtained through the Pythagorean theorem for a multidimensional space, where there are  $n$  individuals and each one of them has values for  $p$  variables. Considering two individuals,  $I$  and  $I'$ , the distance between them is calculated according to Equation 2 as:

$$d_{ii'} = \sum_{j=1}^p \left[ (X_{ij} - X_{i'j})^2 \right]^{\frac{1}{2}} \quad (2)$$

The HCA resulted in dendrograms with cut-off points defining the number of groups formed. The cut-off point was determined from the adjusted variation distance of the agglomeration coefficient (Ferreira 1996).

The correlation level among attributes was assessed through Person's correlation matrix using *Excel software*. This parameter indicates the linear relationship degree between two variables that can be either positive, when an increase in one variable reflects an increase in the other, or negative, when a decrease in one variable implies in a decrease in the other. Coefficient values equal to 1 and -1 indicate a perfect correlation (either positive or negative), while a coefficient value of 0 indicates no linear relationship between the variables. In this study, the cut-off value adopted was 0.7, for a 95% confidence level (Silva 2014).

**Table 1** Physicochemical analyses from the Marizal/São Sebastião aquifer system (August 2021).

Well	pH	EC µS/cm	Salinity (mg/L)	Calcium (mg/L)	Sodium (mg/L)	Manganese (mg/L)	Potassium (mg/L)	Nitrate (mg/L)	Bicarbonate (mg/L)
P-01	5.78	147.8	0.07	0.68	19.0	2.6	2.9	20.55	5.0
P-02	4.71	144.3	0.07	0.87	23.0	3.6	2.5	15.97	5.0
P-03	5.92	47.5	0.02	0.5	8.1	1.1	2.8	6.55	5.0
P-04	6.2	85.5	0.04	1.3	11.0	2.2	9.9	4.43	20.2
P-05	4.98	70.0	0.03	0.79	34.0	3.9	2.6	0.2	5.0
P-06	5.5	178.9	0.09	1.0	31.0	5.0	2.6	26.71	5.0
P-07	6.17	67.1	0.03	0.53	6.3	1.3	7.4	4.34	10.4
P-08	4.97	35.5	0.02	0.5	6.8	0.74	0.49	1.31	5.0
P-09	5.34	141.5	0.07	6.6	57.0	5.3	7.0	1.54	5.11
P-10	5.06	133.5	0.07	0.5	20.0	2.6	1.8	7.2	5.0
P-11	5.9	100.9	0.05	1.1	12.0	2.3	5.3	0.29	10.3
P-12	4.71	162.0	0.08	0.91	21.0	4.5	3.0	22.15	5.0
P-13	6.99	243.0	0.12	3.8	38.0	6.4	7.0	5.17	64.8
P-14	5.68	545.0	0.27	69.0	25.0	7.0	5.0	5.64	19.7
P-15	5.77	129.0	0.06	0.61	16.0	3.4	7.6	2.45	10.8
P-16	5.84	87.8	0.04	0.5	15.0	1.0	5.4	0.2	10.1
P-17	4.6	164.0	0.08	0.79	28.0	2.5	1.3	3.37	5.0
P-18	6.36	364.0	0.18	6.6	39.0	12.0	15.0	1.0	40.4
P-19	5.0	155.6	0.08	0.99	25.0	2.7	5.3	6.54	5.0
P-20	5.53	131.1	0.06	0.81	20.0	3.9	6.6	9.55	5.5
Well	Chloride (mg/L)	Sulfate (mg/L)	Fluoride (mg/L)	Phosphate (mg/L)	Manganese (mg/L)	Aluminum (mg/L)	Iron (mg/L)	Copper (mg/L)	Lead (mg/L)
P-01	19.8	2.0	1.1	0.3	0.049	0.151	0.436	0.01346	0.005
P-02	27.5	2.0	0.35	0.3	0.109	0.17	0.05	0.01416	0.005
P-03	7.1	1.0	0.08	0.3	0.009	0.045	0.088	0.00315	0.005
P-04	11.0	1.2	0.15	0.3	0.005	0.036	0.05	0.00251	0.005
P-05	10.9	1.0	0.08	0.3	0.029	0.025	0.05	0.00873	0.005
P-06	35.1	2.0	0.22	0.3	0.055	0.13	0.05	0.00327	0.005
P-07	8.0	1.0	0.13	0.3	0.005	0.044	0.05	0.0025	0.005
P-08	6.7	1.0	0.18	0.3	0.005	0.089	0.05	0.00633	0.005
P-09	40.8	3.1	0.16	0.3	0.07	0.099	1.7	0.00351	0.005
P-10	31.1	2.0	0.41	0.3	0.105	0.037	0.05	0.01077	0.005
P-11	20.7	1.3	1.0	0.3	0.005	0.116	0.05	0.0025	0.005
P-12	32.2	2.0	0.16	0.3	0.06	0.071	0.086	0.00991	0.005
P-13	26.4	9.8	0.31	0.61	0.1	0.201	0.595	0.00294	0.005
P-14	134.0	10.0	0.8	0.3	0.243	0.164	7.4	0.0025	0.005
P-15	27.8	2.0	0.16	0.3	0.005	0.025	0.05	0.0025	0.005
P-16	16.0	1.0	0.1	0.3	0.005	0.037	0.098	0.0025	0.005
P-17	44.0	2.0	0.46	0.3	0.015	0.067	0.05	0.00674	0.005
P-18	61.2	5.0	0.4	0.3	0.167	0.025	0.05	0.0025	0.005
P-19	39.6	2.0	0.16	0.3	0.033	0.232	0.05	0.00564	0.005
P-20	27.6	2.0	0.16	0.3	0.015	0.038	0.05	0.00346	0.005

**Legend:** pH- potential of hydrogen; EC- electrical conductivity

Evaluation of the chemical reactions in groundwater was set by calculating the Saturation Index (SI) through *PHREEQC Interactive* 3.0 software. As stated by Nascimento and Alves (2011), SI (Equation 3) is defined by the ratio between the ion activity product (IAP) from cations and anions dissociated in an aqueous medium and the equilibrium constant (Ksp) at a given temperature, as follows:

$$SI = \frac{IAP}{K_{sp}} \text{ or } \log SI = \log IAP - \log K_{sp} \quad (3)$$

Lastly, the results obtained from the physicochemical analyses were compared to the maximum permissible limit (MPL) set by Ordinance No. 888/2021 of the Ministry of Health (Brasil 2021), which sets the procedures for controlling and monitoring water quality and its potability standards considered suitable for human consumption, and Ordinance No. 396/2008 of CONAMA (Brasil 2008), which establishes a classification and environmental directives for groundwater.

### 3 Results

The factor analysis through PCA was first carried out with all physicochemical variables: pH, electrical conductivity, salinity, calcium, sodium, magnesium, potassium, nitrate, chloride, bicarbonate, sulfate, fluoride, phosphate, manganese, aluminum, iron, copper, and lead. To obtain the most satisfactory result in the factor analysis, four simulations were necessary based on its criteria, which reduced in 55.5% the number of initial variables. The final simulation preserved 10 variables (pH, electrical conductivity, salinity, calcium, chloride, bicarbonate, sulfate, phosphate, manganese, and iron), which were then distributed between two factors, corresponding to 88% of the total accumulated variance. The result showed that the *KMO* was 0.72 and the *Bartlett* sphericity test was equal to  $p < 0.01$ .

The correlation matrix for the 10 variables selected during the last factor simulation is shown in Table 2. Of

the 45 correlation coefficients, 66.6% showed significant values and 51.1% were within the range of  $0.6 \leq |r| < 0.9$  which, according to Callegari-Jacques (2003), represents a strong correlation. From the remaining coefficients (48.9%), 11.1% were within the range of  $0.3 \leq |r| < 0.6$ , which characterizes a moderate correlation, and 37.8% presented values of  $r > 0.3$ , which characterizes a low correlation.

The correlation matrix exhibited a strong correlation between electrical conductivity and calcium, chloride, sulfate, manganese, and iron (Table 3), which indicates a possible influence of these ions on the salinity of the groundwater in the area. Bicarbonate showed a strong correlation with pH, considering its ability for alkalization (Kehew 2001). The matrix indicated moderate correlations between pH and sulfate and phosphorus; between bicarbonate and manganese and electrical conductivity; and between phosphorous and chloride. In addition to the low correlations between pH and electrical conductivity, calcium, chloride, manganese, and iron; between bicarbonate and calcium, chloride, and iron; and between phosphorous and electrical conductivity, calcium, chloride, manganese, and iron.

The first factor from the factor analysis contributed approximately 63.5% to explain the total variance of the sample set, showing a strong correlation among the following variables: electrical conductivity, salinity, calcium, chloride, sulfate, manganese, and iron (Table 4). In addition, chloride (0.981), electrical conductivity (0.945), and salinity (0.945) represent the variables with the highest factor loadings and are indicators of groundwater salinity (mineralization). Moreover, they showed the greatest influence in characterizing groundwater quality.

The constituent parameters of factor 1 are present in the lithological composition of the Marizal/São Sebastião System (Alves 2015), and can precipitate in groundwater, according to the saturation index (SI) through *PHREEQC* calculations, preferably as calcite minerals ( $\text{CaCO}_3$ ), dolomite [ $\text{CaMg}(\text{CO}_3)_2$ ], alunite [ $\text{KAl}(\text{SO}_4)_2(\text{OH})_6$ ], and jarosite [ $\text{KFe}_2(\text{SO}_4)_2(\text{OH})_6$ ].

**Table 2** KMO and Bartlett's test results.

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO)		0.72
Bartlett's Test of Sphericity	df	45
	Sig	<0.01
Cumulative total		88%

**Table 3** Correlation Matrix (Pearson) of the physicochemical variables selected through the factor analysis (August 2021).

Matrix correlation	EC ( $\mu\text{S/cm}$ )	Salinity (mg/L)	Calcium (mg/L)	Chloride (mg/L)	Sulfate (mg/L)	Manganese (mg/L)	Iron (mg/L)	pH (mg/L)	Bicarbonate (mg/L)	Phosphate (mg/L)
Electrical conductivity	1									
Salinity	0.99	1								
Calcium	0.83	0.82	1							
Chloride	0.94	0.95	0.89	1						
Sulfate	0.85	0.85	0.70	0.72	1					
Manganese	0.90	0.91	0.74	0.84	0.78	1				
Iron	0.78	0.76	0.98	0.86	0.70	0.71	1			
pH	0.20	0.19	0.09	0.02	0.43	0.09	0.074	1		
Bicarbonate	0.47	0.47	0.18	0.20	0.75	0.42	0.141	0.74	1	
Phosphate	0.17	0.17	0.017	0.04	0.64	0.17	0.006	0.54	0.82	1

**Legend:** pH- potential of hydrogen; EC- electrical conductivity.

**Table 4** Factor loadings variance explained by factors and accumulated variance of ten variables, after the varimax rotation method (August 2021). In bold, the strong correlations from both factors.

Variable	Factor 1	Factor 2
pH	0.15	<b>0.81</b>
EC	<b>0.95</b>	0.24
Salinity	<b>0.95</b>	0.24
Calcium	<b>0.94</b>	0.16
Bicarbonate	0.24	<b>0.95</b>
Chloride	<b>0.98</b>	0.35
Sulfate	<b>0.76</b>	0.62
Phosphate	0.04	<b>0.91</b>
Manganese	<b>0.88</b>	0.19
Iron	<b>0.91</b>	0.27
<b>Variance explained by factors</b>	63.5%	24.5%
<b>Accumulated variance</b>	<b>88%</b>	

**Legend:** pH- potential of hydrogen; EC- electrical conductivity ( $\mu\text{S/cm}$ ); salinity (mg/L); calcium (mg/L); bicarbonate (mg/L); chloride (mg/L); sulfate (mg/L); phosphate (mg/L); manganese (mg/L); iron (mg/L).

The second factor from the factor analysis represented about 24.5% of the sample set total variance and showed a strong correlation among the variables pH, bicarbonate, and phosphate (Table 4). As stated by Nossa (2011), Peixinho (2016), and Gomes and Franca (2019), pH and bicarbonate show a strong correlation between each other and are water alkalinity indicators, whereas phosphate has a secondary participation in the process (Pina 2012). These variables may be related to the presence of calcium carbonate levels in the Marizal and São Sebastião Formations. Another explanation is related to the distance

covered by groundwater, considering the recharge zone as a starting point, which is responsible for incorporating more bicarbonates and carbonates in the environment (Nascimento & Alves 2011). Moreover, the influence of phosphate on the environment may be associated with anthropic activities. As reported by Nascimento and Barbosa (2005), some fertilizers and pesticides used in agriculture are derived from phosphate compounds that can accumulate in the unsaturated zone of the soil, representing a source of pollution to groundwater. According to the same study, phosphate is less concerning than nitrate, as it is less

movable in the aqueous medium and is easily adsorbed by the aquifer lithology during ionic exchange, which eases its dispersion in the water. It is also found alongside feces and cleaning products, such as soaps and detergents coming from domestic sewage.

The cluster analysis used the components selected from the factor analysis to group chemically similar samples. According to Franca et al. (2018), in this case, the number of groups is defined considering the first major difference between the rescaled coefficients from the cluster. In the present study, these coefficients elucidate cut-off point 3, which resulted in two dendrograms from factors 1 and 2, and also formed 4 well clusters (Figure 2).

In the first dendrogram (Figure 2A), cluster 1 comprised 50% (10 wells) of the samples analyzed, among them, 60% met the potability standards set by CONAMA Ordinance No. 396/2008 and Ordinance No. 888/2021, regarding the concentration of chloride (MPL = 250 mg/L), sulfate (MPL = 250 mg/L), manganese (MPL = 0.1 mg/L), and iron (MPL = 0.3 mg/L); 40% of the remaining samples showed either manganese or iron concentrations above the maximum permissible limit.

Iron concentration was above the maximum permissible limit for human consumption in wells P-01 (0.44 mg/L) and P-09 (1.7 mg/L). According to data acquired from the saturation index (SI), this result reflects the existence of iron oxide in the environment and facilitates precipitation of goethite and hematite in the groundwater of the Marizal/São Sebastião system. Excessive iron intake, as reported by Nutti et al. (2006),

can harm the immune system and cause developmental delay, anemia, reduced work capacity, and death.

Manganese concentration was above the maximum permissible limit for human consumption in wells P-02 (0.11 mg/L) and P-10 (0.11 mg/L). The saturation index (SI) value reflects the presence of manganese carbonate in the water, enabling the precipitation of rhodochrosite in the groundwater of the Marizal/São Sebastião system. As stated by Ishimine (2002), excessive manganese intake can cause headaches, irritability, and emotional instability, which directly affects the central nervous system.

Regarding water quality for irrigation, the samples from cluster 1 met the maximum permissible limit established by CONAMA Ordinance No. 396/2008, except for sulfate, which does not have a defined maximum concentration. Cluster 1 was also characterized by waters with low levels of salinity and electrical conductivity ranging from 19.8 to 40.8  $\mu\text{S}/\text{cm}$ .

Cluster 2 comprised 35% (7 wells) of the samples analyzed, in which 100% met the potability standards established by Ordinance No. 888/2021 of the Ministry of Health and by CONAMA Ordinance No. 396/2008 regarding the concentration of chloride, sulfate, manganese, and iron. Thus, this cluster showed the best quality level for the waters analyzed. The water from this cluster is suitable for irrigation, with the concentration of chloride, manganese, and iron within the maximum permissible limit set by CONAMA Ordinance No. 396/2008. Lastly, it showed electrical conductivity values varying between 35.5 and 100.9  $\mu\text{S}/\text{cm}$ , which indicates low salinity.

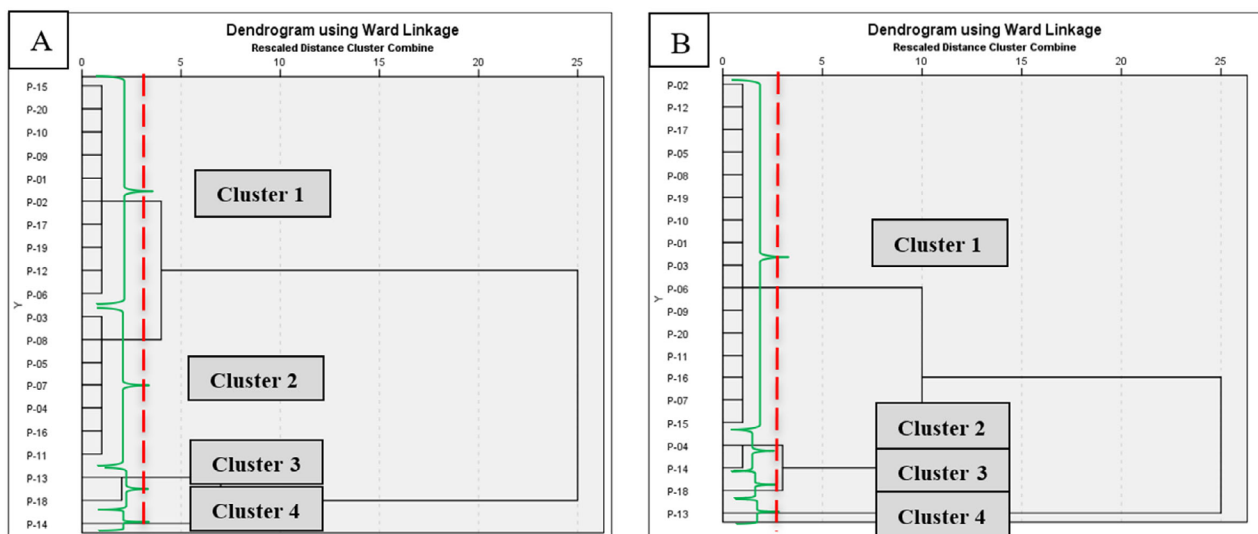


Figure 2 Dendrograms resulting from a hierarchical cluster analysis of variables: A. Factor 1; B. Factor 2.

Cluster 3 comprised 10% (2 wells) of the samples analyzed, with one sample (P-13) showing an iron concentration (0.59 mg/L) above the level allowed by Ordinance No. 888/2021 of the Ministry of Health and CONAMA Ordinance No. 396/2008. Despite that, the water from this cluster was considered appropriate for irrigation, meaning that chloride, manganese, and iron concentrations met the maximum permissible limit set by CONAMA Ordinance No. 396/2008.

Cluster 4 represented 5% (P-14) of the samples and showed manganese (0.24 mg/L) and iron (7.4 mg/L) concentrations above the maximum permissible limit set by CONAMA Ordinance No. 396/2008. Thus, the water from this cluster was considered inappropriate for irrigation.

In addition, clusters 3 and 4 revealed electrical conductivity values ranging from 243 to 545  $\mu\text{S}/\text{cm}$ , which characterize low and medium salinity, respectively. The increase in the salinity shown in cluster 4, according to the results of the saturation index (SI), calculated through *PHREEQC*, reflects the presence of ions released by minerals of calcite ( $\text{CaCO}_3$ ), dolomite [ $\text{CaMg}(\text{CO}_3)_2$ ], alunite [ $\text{KAl}(\text{SO}_4)_2(\text{OH})_6$ ], and jarosite [ $\text{KFe}_2(\text{SO}_4)_2(\text{OH})_6$ ] in the groundwater system.

In the second dendrogram (Figure 2B), cluster 1 comprised 80% (16 wells) of the samples analyzed, corresponding to clusters 1 and 2 of the first dendrogram, except for sample P-04. Cluster 1 was characterized as having acidic waters (100% of samples) and showed pH values varying from 4.6 to 6.2. From this set, only sample P-07 showed pH (6.2) within the limits set by Ordinance No. 888/2021 of the Ministry of Health (Brasil 2008) (6.0-9.5). It also showed the lowest bicarbonate concentrations (maximum of 10.8 mg/L). Similar results were found by Andrés (2011) and Alves (2015).

The main reason for the pH values of the water to be lower than 6.0 is the significant amount of clay minerals that, when dissolved, release silica and aluminum to the environment (Hounslow 1995). As stated in Alves (2015), the groundwater of the São Sebastião Formation is saturated, in some places, with silica, which can be related to the Medium Member, Passagem dos Teixeiras, composed of sandstones interbedded with siltstones, fine clayey sandstones, and rare limestones.

Cluster 2 included 10% of the samples (2 wells), all of which presented acidic waters. Among them, sample P-14 presented a pH concentration (5.7) beyond the permissible limit set by Ordinance No. 888/2021 of the Ministry of Health, whereas sample P-04 showed pH concentrations (6.2) within the set limit. The group also showed intermediate concentrations of bicarbonate (19.7 and 20.2 mg/L).

Clusters 3 and 4 equally comprised 5% of the samples (one well for each cluster) and showed waters with pH values of 6.4 and 7.0, respectively, thus meeting the limit set by Ordinance No. 888/2021 of the Ministry of Health. Also, both groups showed the highest concentrations of bicarbonate, 40.4 and 64.6 mg/L respectively. Finally, cluster 4, with sample P-13, was the only one that showed phosphate concentration (0.6 mg/L) above the permissible limit. In the municipality of Boa União, where the sample was taken, it is common to observe the discharge of untreated wastewater in the local river and the use of pesticides in agriculture. These actions may contribute to the increase of phosphate concentration in the local groundwater.

## 4 Conclusions

Groundwater properties from the Marizal/São Sebastião system are related primarily to the geological formation of the area and secondarily to anthropic contamination. Waters were characterized considering quality variables (pH, EC, salinity, calcium, sodium, chloride, bicarbonate, sulfate, phosphate, manganese, and iron), which were identified through factor analysis.

The cluster analysis using the first factor (pH, EC, salinity, calcium, sodium, chloride, bicarbonate, sulfate, phosphate, manganese, and iron) showed that most of the samples analyzed met the potability standards (70%) and the irrigation standards (95%) set by Ordinance No. 888/2021 of the Ministry of Health and by CONAMA Ordinance No. 396/2008. The inappropriate water samples showed high concentrations of iron ions and manganese ions, both associated with the lithology of the area. These can oxidate and precipitate in the form of goethite and hematite minerals and in the form of rhodochrosite, respectively.

The cluster analysis using the second factor (pH, bicarbonate, and phosphate) showed that most samples had acidic waters (95%) and their pH was beyond the permissible limit (85%) set by Ordinance No. 888/2021 of the Ministry of Health. This condition is probably related to the dissolution of clay minerals from the Medium Member of the São Sebastião Formation.

The results showed that the use of multivariate statistical analyses is relevant for the monitoring and management of groundwater quality, especially in semiarid regions, such as the study area, characterized by high socio-environmental vulnerability. Furthermore, the use of statistical tools promotes cost reduction, by selecting exclusively the most relevant physicochemical variables to characterize the area.

In summary, the present study can contribute to future research developments that aim to carry out new



monitoring of the groundwater quality of the Marizal/São Sebastião aquifer located in Alagoinhas, state of Bahia, Brazil.

## 5 Acknowledgments

The authors would like to thank the Federal University of Bahia (UFBA), the Center for Hydrogeological and Environmental Studies (NEHMA), the SAAE (Water and Sewage Service Provider from Alagoinhas), and the National Council for Scientific and Technological Development (CNPq) for the financial support granted to the present study.

## 6 References

- Alves, J.E. 2015, 'Estudo hidrogeoquímico comparativo entre os membros da Formação São Sebastião, Recôncavo Norte – BA', Dissertação de Mestrado, Universidade Federal da Bahia.
- Andrés, H.L.J. 2011, 'Avaliação dos problemas de contaminação urbano-industrial na área de Alagoinhas, Bahia', Dissertação de Mestrado, Universidade Federal da Bahia.
- Araújo, A.O., Mendonça, L.A.R., Lima, M.G.S., Feitosa, J.V., Silva, F.J., Ness, R.L.L., Frischkorn, H., Simplício, A.A.F. & Kerntopf, M.R. 2013, 'Modificações nas propriedades dos solos de uma área de manejo florestal na Chapada do Araripe', *Revista Brasileira Ciência do Solo*, vol. 3, no. 1, pp. 754-62, DOI:10.1590/S0100-06832013000300022.
- Brasil 2008, CONAMA – Conselho Nacional do Meio Ambiente Nº396, de 3 de abril de 2008, Brasília, visto 12 Fevereiro 2022, <<https://www.mpf.mp.br/atuacao-tematica/ccr4/dados-da-atuacao/projetos/qualidade-da-agua/legislacao/resolucoes/resolucao-conama-no-396-de-3-de-abril-de-2008/view>>.
- Brasil 2021, Portaria Nº 888 de 04 de maio de 2021, Ministério da Saúde, Brasília, visto 12 Fevereiro 2022, <<https://www.in.gov.br/en/web/dou/-/portaria-gm/ms-n-888-de-4-de-maio-de-2021-318461562>>.
- Callegari-Jacques, S.M. 2003, *Bioestatística: princípios e aplicações*, Artemed, Porto Alegre.
- Costa, C.F.T., Bezera, J.E.C.D., de Paula Filho, F.J., Firmino, P.R.A. & Souza, R.B. 2020, 'Análise multivariada aplicada ao estudo hidroquímico das águas subterrâneas na bacia sedimentar do Araripe – CE', *Revista Águas Subterrâneas*, vol. 34, no. 2, pp. 01-06, DOI:10.14295/ras.v34i2.29874.
- Ferreira, D.F. 1996, *Análise multivariada*, Lavras, Minas Gerais.
- Franca, M.R., Gomes, R.C.M., Silva, A.J.F., Lima, S.G.M. & Frischkorn, H. 2018, 'Análise multivariada dos parâmetros de qualidade das águas subterrâneas em Juazeiro do Norte – CE (Brasil)', *Revista Águas Subterrâneas*, vol. 32, no. 1, pp. 109-10, DOI:10.14295/ras.v32i1.28890.
- Gomes, R.C.M. & Cavalcante, N.I. 2017, 'Aplicação da análise multivariada do estudo da qualidade da água subterrânea', *Revista Águas Subterrâneas*, vol. 31, no. 1, pp. 134-49, DOI:10.14295/ras.v31i1.28617.
- Gomes, R.C.M., Mendonça, R.A.L., Lima, S.G.M. & Araújo, O.A. 2017, 'Similaridade de atributos físico-químicos com uso de análise multivariada aplicada à gestão qualitativa de águas subterrâneas', *Revista Geociências*, vol. 36, no. 2, pp. 325-37, DOI:10.5016/geociencias.v36i2.12591.
- Gomes, R.C.M. & Franca, M.R. 2019, 'Uso da análise multivariada para subsidiar no monitoramento da qualidade da água subterrânea na bacia sedimentar do Araripe, sul do Estado do Ceará', *Revista Geociências - Unesp*, vol. 38, no. 1, pp. 195-205, DOI:10.5016/geociencias.v38i1.13052.
- Hair Jr, J.F., Anderson, R.E., Tathan, R.L. & Black, W.C. 1998, *Multivariate data analysis*, New Jersey, Prentice Hall.
- Hounslow, A.W. 1995, *Water quality data: analysis and interpretation*, Lewis Publishers, New York.
- Ishimine, V. 2002, 'Avaliação hidrogeológica e hidrogeoquímica em área contaminada por manganês na região de Suzano – SP', Dissertação de Mestrado, Universidade de São Paulo.
- Keheew, A.E. 2001, *Applied chemical hydrogeology*, Prentice-Hall, Inc. Upper Saddle River, New Jersey.
- Moraes, S.R.L., Reis, M.G.C., Zanta, V.M., Luz, D.L., Orrico, M.R.S., Machado, N.L., Agra Filho, S.S. & Nascimento, S.A.M. 2004, *Plano Municipal de Saneamento ambiental de Alagoinhas*, vol. 2, viewed 19 January 2022, <[https://www.academia.edu/4455765/Plano\\_Municipal\\_de\\_Saneamento\\_Ambiental\\_de\\_Alagoinhas\\_Brasil\\_metodologia\\_e\\_resultados](https://www.academia.edu/4455765/Plano_Municipal_de_Saneamento_Ambiental_de_Alagoinhas_Brasil_metodologia_e_resultados)>.
- Nascimento, M.A.S. & Barbosa, F.S.J. 2005, 'Qualidade da água do aquífero freático no Alto Cristalino de Salvador, Bacia do Rio Lucaia, Salvador, Bahia', *Revista Brasileira de Geociências*, vol. 35, no. 4, pp. 543-50, DOI:10.25249/0375-7536.200535543550.
- Nascimento, M.A.S., Moraes, S. R.L. & Santana, A.V.A. 2006, 'Estudo quali-quantitativo das águas subterrâneas no Município de Alagoinhas-Bahia como componente do Plano Municipal de Saneamento Ambiental', *Anais do XIV Congresso Brasileiro de Águas Subterrâneas*, Brasil, viewed 15 December 2021, <<https://aguassubterraneas.abas.org/asubterraneas/article/view/22647/14866>>.
- Nascimento, S.A.M. & Alves, J.E. 2011, 'Caracterização hidrogeoquímica e evolução das águas subterrâneas na área do Polo Industrial de Camaçari – Bahia', *Cadernos de Geociências*, vol. 98, no. 2, pp. 54-62.
- Nascimento, S.A.M. & Alves, J.E. 2014, 'Estudo hidrogeoquímico comparativo entre os aquíferos Barreiras, Marizal e São Sebastião na região do Recôncavo Norte, Estado da Bahia', *Cadernos de Geociências*, vol. 11, no. 1-2, pp. 72-3.
- Nutti, M., Carvalho, V.L.J. & Watanabe, E. 2006, 'A biofortificação como ferramenta para combate a deficiências em micronutrientes', in R.C. Silva, R.B. Figueiredo, M.E. Capitani & G.F. Cunha (eds), *Geologia Médica no Brasil*, CPRM, pp. 43-7.
- Nossa, T.C.B. 2011, 'Avaliação da vulnerabilidade do aquífero cárstico Salitre-Bahia, através de análises hidroquímicas, isotópicas e aplicação da metodologia COP', Tese de Doutorado, Universidade Federal da Bahia, viewed 12 February 2022, <<https://repositorio.ufba.br/handle/ri/21489>>.
- Peixinho, L.A.M. 2016, 'Hidrogeologia do sistema aquífero Marizal/São Sebastião na sub-bacia do Tucano Central, Estado da Bahia', Dissertação de Mestrado, Universidade Federal da

- Bahia, viewed 12 February 2022, <<https://repositorio.ufba.br/handle/ri/22501>>.
- Pereira, A.P. & Lima, K.A.O. 2007, 'Estrutura elétrica da contaminação hídrica provocada por fluidos provenientes dos depósitos de lixo urbano e de um curtume no município de Alagoinhas, Bahia', *Revista Brasileira de Geofísica*, vol. 25, no. 1, pp. 6-8, DOI:10.1590/S0102-261X2007000100001.
- Pina, A.P.S.A. 2012, 'Fundamentos hidrogeoquímicos aplicados na Bacia Hidrográfica de Santa Cruz, Ilha de Santiago – Cabo Verde, como instrumento para a gestão de recursos hídricos, Fortaleza', *Dissertação de Mestrado*, Universidade Federal do Ceará, viewed 12 February 2022, <[http://bdtd.ibict.br/vufind/Record/UFC-7\\_ca9c79bccd234495aee45cebb7e40a3f](http://bdtd.ibict.br/vufind/Record/UFC-7_ca9c79bccd234495aee45cebb7e40a3f)>.
- Ribeiro, L.G. 2008, 'Avaliação geoelétrica da contaminação urbano-industrial do aquífero Recôncavo no entorno de Alagoinhas, Bahia', *Monografia de Graduação*, Universidade Federal da Bahia, viewed 12 February 2022, <<http://www.cpgg.ufba.br/gr-geof/geo213/trabalhos-graduacao/Gilson-Ribeiro.pdf>>.
- Silva, S.A.C. 2014, 'Aspectos Hidrogeológicos dos Municípios de Aramiri e Ouriçangas-Bahia', *Monografia de Graduação*, Universidade Federal da Bahia, viewed 12 February 2022, <[https://nehma.ufba.br/sites/nehma.ufba.br/files/Mon\\_Silva2013\\_2.pdf](https://nehma.ufba.br/sites/nehma.ufba.br/files/Mon_Silva2013_2.pdf)>.
- Vicini, L. (ed.) 2005, *Análise multivariada da teoria à prática*, Santa Maria.

#### Author contributions

**Maíra Sampaio da Costa:** conceptualization; formal analysis; methodology; validation; writing-original draft; writing – review and editing; visualization. **Maria da Conceição Rabelo Gomes:** methodology; validation; supervision. **Sérgio Augusto de Moraes Nascimento:** validation; supervision.

#### Conflict of interest

The authors declare no conflict of interest.

#### Data availability statement

Scripts and code are available on request.

#### Funding information

CNPq, Process N° 140126/2020-1

#### Editor-in-chief

Dr. Claudine Dereczynski

#### Associate Editor

Dr. Gerson Cardoso da Silva Jr.

#### How to cite:

Costa, M.S., Gomes, M.C.R. & Nascimento, S. A. M 2023, 'Cluster and Factor Analyses as Contributions to the Groundwater Quality Monitoring of the Marizal/São Sebastião Aquifer System, Alagoinhas (Bahia, Brazil)', *Anuário do Instituto de Geociências*, 46:54180. [https://doi.org/10.11137/1982-3908\\_2023\\_46\\_54180](https://doi.org/10.11137/1982-3908_2023_46_54180)