

STATISTICAL MULTIVARIATE ANALYSIS APPLIED TO ENVIRONMENTAL CHARACTERIZATION OF SOIL IN SEMIARID REGION¹

ANTÔNIO ITALCY DE OLIVEIRA JÚNIOR^{2*}, LUIZ ALBERTO RIBEIRO MENDONÇA², SÁVIO DE BRITO FONTENELE³, ADRIANA OLIVEIRA ARAÚJO⁴, MARIA GORETHE DE SOUSA LIMA BRITO²

ABSTRACT - Soil is a dynamic and complex system that requires a considerable number of samples for analysis and research purposes. Using multivariate statistical methods, favorable conditions can be created by analyzing the samples, i.e., structural reduction and simplification of the data. The objective of this study was to use multivariate statistical analysis, including factorial analysis (FA) and hierarchical groupings, for the environmental characterization of soils in semiarid regions, considering anthropic (land use and occupation) and topographic aspects (altitude, moisture, granulometry, PR, and organic-matter content). As a case study, the São José Hydrographic Microbasin, which is located in the Cariri region of Ceará, was considered. An FA was performed using the principal component method, with normalized varimax rotation. In hierarchical grouping analysis, the “farthest neighbor” method was used as the hierarchical criterion for grouping, with the measure of dissimilarity given by the “square Euclidean distance.” The FA indicated that two factors explain 75.76% of the total data variance. In the analysis of hierarchical groupings, the samples were agglomerated in three groups with similar characteristics: one with samples collected in an area of the preserved forest and two with samples collected in areas with more anthropized soils. This indicates that the statistical tool used showed sensitivity to distinguish the most conserved soils and soils with different levels of anthropization.

Keywords: Factor analysis. Cluster analysis. Soil variables.

ESTATÍSTICA MULTIVARIADA APLICADA NA CARACTERIZAÇÃO AMBIENTAL DE SOLOS EM REGIÃO SEMIÁRIDA

RESUMO - O solo é um sistema dinâmico e complexo ao ser estudado, que demanda para fins de pesquisa um considerável número de amostras e análises. Com os métodos estatísticos multivariados podem-se criar condições mais favoráveis ao analisar as amostras, i.e. redução e simplificação estrutural dos dados. Este trabalho tem por objetivo empregar a ferramenta de análise estatística multivariada, por meio da análise fatorial (AF) e de agrupamentos hierárquicos, na caracterização ambiental de solos em regiões semiáridas, considerando aspectos antrópicos (uso e ocupação do solo) e topográficos (altitude e declividade), além de atributos físicos e químicos (umidade, granulometria, resistência à penetração e matéria orgânica), utilizando como estudo de caso a Microbacia Hidrográfica do São José (MHSJ), localizada na Região do Cariri cearense. A AF foi realizada pelo método dos componentes principais, empregando a rotação varimax normalizada. Na análise de agrupamento hierárquico foi utilizado o método do “vizinho mais distante” como critério hierárquico de agrupamento, com medida de dissimilaridade dada pela “distância euclidiana quadrada”. A AF demonstrou que dois fatores explicam 75,76% da variância total dos dados. A análise de agrupamentos hierárquicos aglomerou as amostras em três grupos com características similares: um com amostras coletadas em uma área de floresta preservada e dois com amostras coletadas em áreas com solos mais antropizados. Isso indica que a ferramenta estatística utilizada apresentou sensibilidade para distinguir os solos mais conservados e os solos com diferentes níveis de antropização.

Palavras-chave: Análise fatorial. Análise de agrupamentos Hierárquicos. Solos antropizados.

*Corresponding author

¹Received for publication in 11/01/2017; accepted in 09/27/2018.

²Department of Civil Engineering / Water Resources and Sanitation, Universidade Federal do Cariri, Juazeiro do Norte, CE, Brazil; antonioitalcy@hotmail.com – ORCID: 0000-0002-8297-5068, larm@ufc.br – ORCID: 0000-0002-8166-3337, goretthe.lima@ufca.edu.br – ORCID: 0000-0001-8977-1116.

³Department of Production Engineering, Faculdade Paraíso do Ceará, Juazeiro do Norte, CE, Brazil; savio.fontenele@hotmail.com – ORCID: 0000-0002-5098-762X.

⁴Department of Environment, Instituto Federal da Paraíba, Princesa Izabel, PB, Brazil; adrianasaneamento050@gmail.com – ORCID: 0000-0003-3372-5668.

INTRODUCTION

The use and occupation of soil, especially the modification of natural vegetation and increasing urbanization, may leave it unprotected, increasing its impermeability. Consequently, the soil becomes compacted and more susceptible to surface runoff and sediment loss, leading to environmental degradation.

As an open system, soil is dynamic and is the major component of terrestrial ecosystems, influencing the energy balance, the water cycle, the cycling of nutrients, and the productivity of the ecosystem (MACHADO, 2005). Thus, soil degradation has a direct influence on environmental sustainability.

In semiarid regions such as northeastern Brazil, whose main characteristics are shallow soils and marked water scarcity (ARAÚJO, 2011), the quality and quantity of water reserves are very sensitive to changes in the water basin soil characteristics. Thus, the characterization of the soils in these regions is critical for the preparation of management plans that contribute to socioeconomic and environmental sustainability.

The characterization of soils in a given microbasin can be accomplished by studying the anthropic (land use and occupation) and topographic aspects (altitude and declivity), in conjunction with the physical and chemical properties (moisture, granulometry, penetration resistance (PR), organic-matter content, etc.). However, to ensure the technical quality of the information obtained from this characterization, a large number of collections and analyses is required (RIOS et al., 2008). Consequently, these surveys are sometimes unfeasible for financial, logistical, or operational reasons, giving rise to uncertainties regarding the sampling procedure.

Multivariate statistical analysis can be used as an auxiliary tool in sample scaling and in selecting variables for soil characterization. This tool has recently been used by several researchers for exploratory data analysis. Research indicates this approach is effective for studying soil, sediment, and water samples (ARAÚJO et al., 2013; KHWEDIM et al., 2015; SINGH; RAMANATHAN, 2015). For investigating how the studied variables are related or how close they are, two multivariate techniques are particularly effective: factorial analysis (FA) and grouping analysis (HAIR JR et al., 2005; SALGADO et al., 2011).

Because multivariate analysis techniques are capable of solving a problem by constructing a smaller set of variables and classifying samples with similar or dissimilar features, they may allow soil characterization in semiarid regions whose complexity necessitates a large number of samples and analyses.

The objective of this study is to use multivariate statistical analysis methods, including FA and hierarchical grouping methods, for the environmental characterization of semiarid soils, considering anthropic and topographic aspects, as well as physical and chemical features, in the São José Hydrographic Microbasin (SJHM), which is located in the Cariri region of Ceará.

MATERIAL AND METHODS

Area of study and sampling points

The study was performed in the SJHM, which is located in the southernmost region of Ceará and comprises an area of approximately 40 km² (Figure 1). The microbasin is part of Araripe sedimentary basin and is one of the sub-basins of Salgado, which is one of the main sources of the Jaguaribe River—the largest river in the state of Ceará.

According to the Köppen–Geiger classification (1928), the region has a semiarid hot tropical climate, with 925 mm of annual rainfall. The rainy season—from January to May—is responsible for approximately 70% of the annual rainfall (FUNCEME, 2006). Although the typology is varied, with significant spatial diversification of soils, the studied area contains only three types of soils: Red-Yellow Latosols, Neosols, and Argisols (FUNCEME, 2006).

A total of 39 geo-referenced points in the SJHM (Figure 1) were selected for obtaining data. The data were obtained from the 0.0–0.2-m surface layer.

Anthropic Aspects

The uses and occupations of SJHM lands were classified by Costa (2013) into the following categories: ground vegetation, urban area, arboreal-shrubby caatinga, agriculture, rainforest, savannah, and exposed soil (Figure 2), which occupy 43.6%, 6.7%, 22.7%, 7.9%, 5.5%, 8.9%, and 4.7% of the SJHM area, respectively.

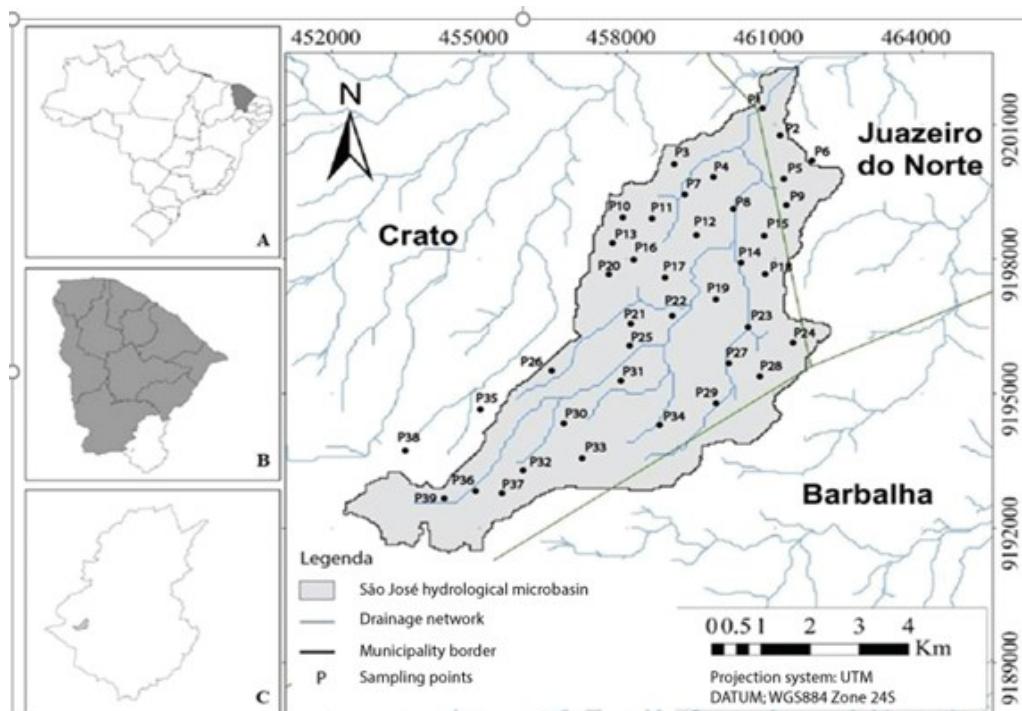


Figure 1. SJHM location. A) Map of Brazil, highlighting the state of Ceará; B) Ceará divided into water basins, highlighting the Salgado river basin in the southeast; and C) the SJHM in 2014. Source: Adapted from Costa (2013).

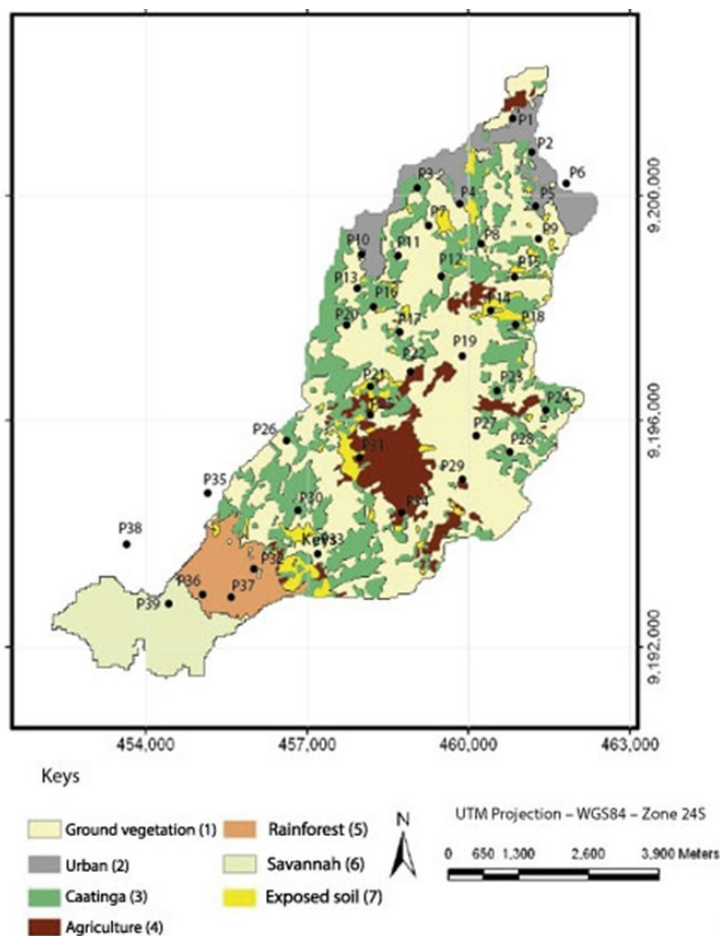


Figure 2. Map of SJHM land use and occupation in 2014. Source: Adapted from costa (2013)

Topographical Aspects and Soil Physical and Chemical Properties

Topographical data for the SJHM were obtained using Digital Terrain Models from Shuttle Radar Topography Mission data, with a spatial resolution of 92 m × 92 m, made available by the Brazilian Agricultural Research Corporation (EMBRAPA, 1979).

The soil samples were placed in plastic bags and taken to the Sanitation and Soil Mechanics Laboratories of the Federal University of Cariri for analysis.

The moisture content of the samples was determined via gravimetry. The dry masses were determined by weighing the samples after drying in an oven at approximately 105 °C (EMBRAPA, 1997).

Granulometry was performed using the pipette method (EMBRAPA, 1997), with sodium hydroxide (NaOH 1 mol·L⁻¹) as a chemical dispersant. In this analysis, a volume of soil suspension was pipetted and used to determine the clay content via oven drying and weighing. The sand content was obtained after segregation, oven drying, and weighing. The silt content was determined by calculating the complement of the 100% percentage.

The PR was measured using a Stolf impact penetrometer (STOLF; FERNANDES; FURLANI NETO, 1983). In this procedure, the impact penetration was read at the penetrometer-graduated rod. The results indicated the number of impacts per decimeter of perforated soil (N), which was converted into MPa using Equation 1.

$$PR = 0.547 + (0.675 \times N) \quad (\text{Eq. 1})$$

Here, the PR has units of MPa, and N is the number of impacts required to perforate 1 decimeter of soil (impacts/dm).

The organic-matter content was determined via wet oxidation (Walkley–Black modified method) (EMBRAPA, 1997). This method is based on the oxidation of organic matter with potassium dichromate (Cr₂O₇) in an acidic medium (H₂SO₄) as an oxidizing agent. In this analysis, the potassium dichromate remaining after the oxidation reaction was titrated with a standard solution of ammoniacal ferrous sulfate (FeSO₄·7H₂O) in an acidic medium, using diphenylamine or ferroin as an indicator. Thus, the organic-matter content was determined.

Statistical Analysis

A sample composed of 39 data points for each studied variable was used in the statistical analysis. The following 10 variables were analyzed: declivity, altitude, clay content, silt content, coarse-sand content, fine-sand content, organic-matter content, moisture, PR, and land use and occupation.

We noticed that the declivity, altitude, silt content, clay content, organic matter, moisture, and land use and occupation data did not follow a univariate normal distribution. In contrast, the data for the coarse-sand content, fine sand content, and PR followed a univariate normal distribution.

Regarding the multivariate statistics, the initial procedure was the standardization of the processed data—for excluding the scale effect and units—by converting the variables into standard scores called Z scores. The next step was to construct a Pearson linear correlation matrix and use this matrix to evaluate the degree of association between the variables for analyzing the applicability of the multivariate techniques to the data. Subsequently, the FA technique was applied, using the principal component approach with the standardized varimax rotation (ARAÚJO et al., 2013). All the points and variables sampled in the SJHM were used in this step. We used the software SPSS version 20.0 to process the information.

The data were subjected to Kaiser–Meyer–Olkin (KMO) and Bartlett (Sig.) sphericity tests in order to verify their suitability for FA. According to Bezerra and Corrar (2006), a KMO test result higher than 0.5 indicates that the data are adequate for FA. Regarding the sphericity test (Sig.) result, the authors recommend that it not exceed 0.05.

After the FA, we identified the covariance relationships between the correlated variables, taking into account a small number of random variables, which were defined as factors. The explanatory power of each variable was revealed through communalities, considering all the factors found in the FA.

Afterwards, we conducted a hierarchical grouping analysis, with the purpose of spatially grouping the points collected from the maximum number of variables strongly explained by one single factor obtained through FA (ARAÚJO et al., 2013). For the hierarchical grouping analysis, we used the “furthest neighbor” method as the hierarchical criterion, with the measure of dissimilarity given by the “squared Euclidean distance.” This criterion was selected because it is one of the most frequently used hierarchical algorithms when the objective of the analysis is to identify groups with elements of similar characteristics (SALGADO et al., 2011; ARAÚJO et al., 2013). The procedure resulted in a dendrogram indicating the similarity of the elements within the groups. The dendrogram “cut-off point” for the determination of the number of groups formed was based on the technique presented by Hair Jr. et al. (2005); it is reached when the rescaled distance of the agglomeration coefficient exhibits greater variation.

The linear combinations of the standardized variables were then subjected to a one-way analysis of variance (ANOVA), for each generated group, to

check whether a stable group set was formed. Then, each group was submitted to descriptive statistical analysis, and the maximum, minimum, mean, standard deviation, and coefficient of variation (CV) of all the variables used in the hierarchical grouping analysis were calculated.

RESULTS AND DISCUSSION

Table 1 presents the variables used for the environmental characterization of the SJHM soils.

Table 1. Data from variables used for the environmental characterization of the SJHM soils for 2014.

POINT	Dec. (%)	Alt. (m)	CS (%)	FS (%)	S (%)	C (%)	OM (g/kg)	M (%)	PR (MPa)	UO
P1	3.12	388.90	0.63	0.288	0.064	0.019	11.17	2.19	3.18	1.00
P2	3.37	418.10	0.55	0.410	0.039	0.002	2.28	0.83	3.02	2.00
P3	7.78	441.90	0.36	0.387	0.111	0.143	8.38	4.61	5.38	3.00
P4	6.71	425.50	0.59	0.338	0.028	0.042	1.86	1.53	5.51	2.00
P5	4.46	418.90	0.51	0.430	0.056	0.002	1.55	1.45	2.17	2.00
P6	6.89	433.70	0.18	0.374	0.163	0.284	7.65	14.02	14.87	2.00
P7	1.46	426.00	0.47	0.459	0.025	0.060	1.45	1.67	9.12	1.00
P8	0.90	404.10	0.14	0.227	0.358	0.278	19.86	9.36	12.02	1.00
P9	3.18	419.20	0.47	0.390	0.068	0.072	7.55	4.78	2.17	1.00
P10	3.30	481.20	0.42	0.380	0.064	0.141	8.59	4.79	7.08	1.00
P11	6.64	463.70	0.41	0.307	0.086	0.197	9.41	5.76	10.14	1.00
P12	9.36	434.70	0.33	0.506	0.037	0.127	5.07	3.60	9.88	3.00
P13	3.36	497.20	0.51	0.263	0.068	0.160	6.30	5.84	6.84	3.00
P14	2.01	411.60	0.55	0.405	0.028	0.020	1.76	1.23	3.50	1.00
P15	3.09	419.00	0.45	0.457	0.035	0.056	1.97	2.24	3.61	1.00
P16	6.79	465.70	0.50	0.410	0.073	0.022	3.31	1.81	6.88	3.00
P17	8.08	436.90	0.35	0.514	0.102	0.037	4.45	2.24	11.59	3.00
P18	4.53	431.60	0.42	0.434	0.052	0.094	5.17	3.00	4.36	1.00
P19	3.76	434.60	0.19	0.717	0.035	0.054	4.45	1.75	2.43	1.00
P20	6.72	509.80	0.56	0.249	0.053	0.144	9.52	5.18	6.84	3.00
P21	3.49	466.70	0.19	0.184	0.521	0.106	6.41	4.35	3.17	3.00
P22	3.32	426.60	0.54	0.387	0.039	0.034	4.76	1.88	3.17	1.00
P23	5.19	425.30	0.57	0.346	0.040	0.042	5.59	2.44	6.29	3.00
P24	4.41	446.70	0.26	0.642	0.057	0.044	2.79	2.78	9.39	3.00
P25	10.75	448.60	0.38	0.292	0.153	0.179	11.48	5.32	3.17	1.00
P26	11.28	570.30	0.33	0.219	0.170	0.278	12.00	7.91	16.63	3.00
P27	1.93	429.60	0.65	0.287	0.045	0.022	3.62	0.88	6.29	1.00
P28	4.32	442.10	0.30	0.556	0.073	0.068	4.65	2.89	5.04	1.00
P29	1.99	438.80	0.30	0.595	0.078	0.029	4.34	2.25	3.15	1.00
P30	9.61	602.50	0.41	0.323	0.107	0.157	9.72	2.95	7.08	1.00
P31	1.26	449.70	0.61	0.306	0.033	0.050	1.76	1.90	10.15	4.00
P32	6.10	695.50	0.38	0.285	0.095	0.241	13.34	6.90	9.12	5.00
P33	13.35	644.70	0.39	0.334	0.137	0.138	10.65	3.93	13.34	1.00
P34	2.79	447.70	0.36	0.371	0.105	0.168	10.65	4.55	2.62	4.00
P35	8.99	644.30	0.40	0.344	0.104	0.151	17.07	4.60	3.89	1.00
P36	35.23	918.90	0.36	0.253	0.111	0.272	26.79	11.56	2.65	5.00
P37	26.38	816.60	0.28	0.206	0.173	0.337	37.86	12.58	2.39	5.00
P38	27.05	938.40	0.28	0.152	0.152	0.419	35.07	15.93	2.17	6.00
P39	0.22	952.20	0.30	0.143	0.542	0.012	37.55	14.24	2.33	6.00

Dec: declivity; Alt: altitude; CS: coarse sand; FS: fine sand; S: silt; C: clay; OM: organic matter; M: moisture; PR: soil penetration resistance; UO: land use and occupation according to the map shown in Figure 2.

Table 2 shows the correlation matrix of the variables. Among the 25 correlation coefficients, approximately 56% of the total pairs presented significant values. Moreover, 32% of them exhibited

“strong” correlations ($0.6 \leq r < 0.9$), while the other 68% exhibited moderate correlations ($0.3 \leq r < 0.6$), in accordance with the results of Araújo et al. (2013).

Table 2. Pearson's linear correlation matrix of the analyzed variables.

	Dec.	Alt.	UO	CS	FS	S	C	OM	M	PR
Dec.	1									
Alt.	0.708**	1								
UO	0.386*	0.630**	1							
CS	-0.232	-0.276	-0.077	1						
FS	-0.335*	-0.539**	-0.424**	-0.098	1					
S	0.026	0.415**	0.307	-0.562**	-0.536**	1				
C	0.677**	0.549**	0.276	-0.511**	-0.511**	0.248	1			
OM	0.595**	0.864**	0.616**	-0.177	-0.517**	0.390*	0.425**	1		
M	0.447**	0.697**	0.578**	-0.088	-0.407*	0.270	0.318*	0.850**	1	
PR	-0.259	-0.270	-0.075	0.208	0.066	-0.135	-0.188	-0.195	0.115	1

**The correlation is significant at the 0.01 level; *The correlation is significant at the 0.05 level; Other correlations were not significant at 0.05; Dec: declivity; Alt: altitude; UO: land use and occupation; CS: coarse sand; FS: fine sand; S: silt; C: clay; OM: organic matter; M: moisture; PR: soil penetration resistance.

The coarse-sand content exhibited significant negative correlations with the silt ($r \cong -0.56$) and clay ($r \cong -0.51$) contents. There were no significant positive correlations between the coarse-sand content and the other studied variables. In the case of the fine-sand content, we observed significant negative correlations with the altitude ($r \cong -0.54$), silt content ($r \cong -0.54$), organic-matter content ($r \cong -0.52$), clay content ($r \cong -0.51$), land use and occupation ($r \cong -0.42$), and moisture ($r \cong -0.41$).

As organic matter and moisture have significant interaction in fine-particle soils (silt and clay) and the micropores of fine sand alone cannot retain these particles (HICKMANN ; COSTA, 2012), there was no correlation of the fine sand with the organic matter and moisture. This explains why the fine-sand content exhibited significant negative correlations with the organic-matter content and moisture.

The negative correlation between the fine-sand content and land use and occupation may be associated with the fact that certain soil-management procedures contribute to soil compaction, with a decrease in micropores, reducing the infiltration capacity and increasing the surface flow, resulting in the transport of the finer sand fractions (MELO et al., 2008). This process occurs, for instance, in the classes of use and occupation characterized by urbanization and exposed soil.

The use of FA requires checking whether there are sufficient links among the data. Although 10 variables were analyzed, the correlation matrix shows that the PR and the coarse-sand and silt contents have insignificant correlations with the other variables. These variables exhibit correlations below 0.5 with at least three other variables, which is not significant for explaining the total variance, according to Andrade et al. (2007). Thus, two FAs

were performed. In the first, all 10 variables were used, and in the second, the PR and the contents of coarse sand and silt were excluded.

In the first FA, the KMO index was 0.423, and the result of the Bartlett sphericity statistical test was significant, at $p < 0.01$. Although the result of the Bartlett test proved to be acceptable for FA, that of the KMO test was not—it was below 0.5, as reported by Bezerra and Corrar (2006), owing to the presence of the variables for the PR and the contents of coarse sand and silt. On the other hand, for the second FA—which included the variables for the declivity, altitude, land use and occupation, fine-sand content, clay content, organic-matter content, and moisture—the KMO test resulted in an index of 0.799 (good). Additionally, the result of the Bartlett sphericity statistical test was highly significant, at $p < 0.01$, indicating that the factors found in the FA can adequately describe the original data variation.

Thus, Table 3 shows the results of the second FA. Here, the principal component technique was applied, where two factors were extracted with eigenvalues greater than 1, according to Norusis (1990) principles. Then, these factors were subjected to the standardized *varimax* rotation, indicating that they explain 61.47% and 14.29% of the total variance of variables, bringing together, in two components, 75.76% of the data, which were previously diluted among seven variables. These results reinforce the results of Araújo et al. (2013), who obtained two factors explaining 76.15% of the total variance when studying the modifications in soil properties in an area of native vegetation management in the Cariri region of Ceará.

The factor loading of the two factors in Table 3 identifies the variables that are most interrelated in each factor. Values equal to or higher than 0.6 indicate variables with a significant degree of

interrelation (ANDRADE et al., 2007), and values slightly below 0.6 indicate a moderate degree of interrelation.

The variables for the altitude, land use and occupation, organic-matter content, and moisture exhibited weights higher than 0.73 in the first factor, indicating that they are the most important variables for defining the degree of anthropization or conservation of the soils in the studied microbasin. In contrast, the declivity and fine-sand content are variables of moderate interrelations. This factor

presents a higher correlation diversity, comprising topographic aspects (altitude), anthropic aspects (use and occupation), and physical (moisture) and chemical aspects (organic matter) of soils. In this context, it appears that Factor 1 is related to variables capable of controlling soil weathering, i.e., those associated with relief (altitude), water retention capacity (moisture), and edaphic fauna (organic matter) (BRODNJAK-VONCINA et al., 2002), as well as anthropic processes (land use and occupation).

Table 3. Factor loading, communalities, and variance explained by the FA of the eight investigated variables, after varimax rotation.

Variable	Factor Loading		Communality
	Factor 1	Factor 2	
Declivity	0.332	<u>0.795⁽²⁾</u>	0.742
Altitude	<u>0.740⁽¹⁾</u>	0.557	0.858
Use and occupation	<u>0.797⁽¹⁾</u>	0.155	0.659
Fine sand	-0.413	<u>-0.545⁽²⁾</u>	0.510
Clay	0.101	<u>0.933⁽²⁾</u>	0.88
Organic matter	<u>0.864⁽¹⁾</u>	0.378	0.889
Moisture	<u>0.877⁽¹⁾</u>	0.195	0.807
% of variance explained by the factor	61.47	14.29	-
% of accumulated variance explained	61.47	75.76	-

⁽¹⁾Candidate variables to be part of factor 1, ⁽²⁾Candidate variables to be part of factor 2.

The variables explained by Factor 1 have a specific relationship, as in more conserved soils, the type of land use and occupation is generally associated with preserved vegetation, where there is more organic matter and soil moisture retention, leading to increased infiltration and reduction in erosion susceptibility, owing to the surface flow (TRINDADE; VALENTE; MOURÃO JÚNIOR, 2012). Altitude is significant owing to its influence on solar incidence, which is a source of energy in the soil-plant-atmosphere system and plays a significant role in evaporation and evapotranspiration processes. Lower-relief regions tend to be regions of higher temperature and consequently of greater evaporation and evapotranspiration, which accelerates the processes of the soil-plant-atmosphere system (BOHNER; ANTONIC, 2009).

The second factor has strong correlations with the declivity and clay content and moderate correlations with the altitude and fine-sand content, all the other variables exhibiting weak correlations. The variables explaining this factor are related to the soil topography and granularity. Consequently, it becomes evident that Factor 2 is related only to the

granule structure and size, as well as the topographic slope of the region containing the soil. Importantly, variables correlated in this factor are generally used in soil-loss studies.

The communalities represent the total variances explained by the factors in each variable. In this regard, the communalities obtained through the FA indicate that 74% and 86% of the topographic aspect variances (declivity and altitude, respectively) were explained by both factors. The physical properties of the soils (fine-sand, clay, and moisture contents) accounted for between 51% and 88% of the variance, while the chemical property (organic matter) explained 89%. The anthropic aspects (land use and occupation) explained 66%.

Therefore, the altitude, land use and occupation, and organic-matter and moisture contents, which correlated with Factor 1, were sufficient to explain approximately 61.5% of the total variance of the data used to environmentally characterize the soils of the SJHM. This information is relevant for sample size determination, considering that the number of assays and financial costs can be reduced.

The technique of hierarchical grouping analysis linked the soil samples of similar characteristics by synthesizing the variables strongly correlated with Factor 1 of the FA, i.e., the factor best explaining the total data variation (approximately 61.5%). For a cut-off distance of five, three homoge-

neous groups were obtained. The results of this analysis are presented in the dendrogram of Figure 3.

The groups obtained and the linear combinations of the variables generating them were subjected to a one-way ANOVA and exhibited significance levels lower than 5%, indicating the formation of a relatively stable set of groups.

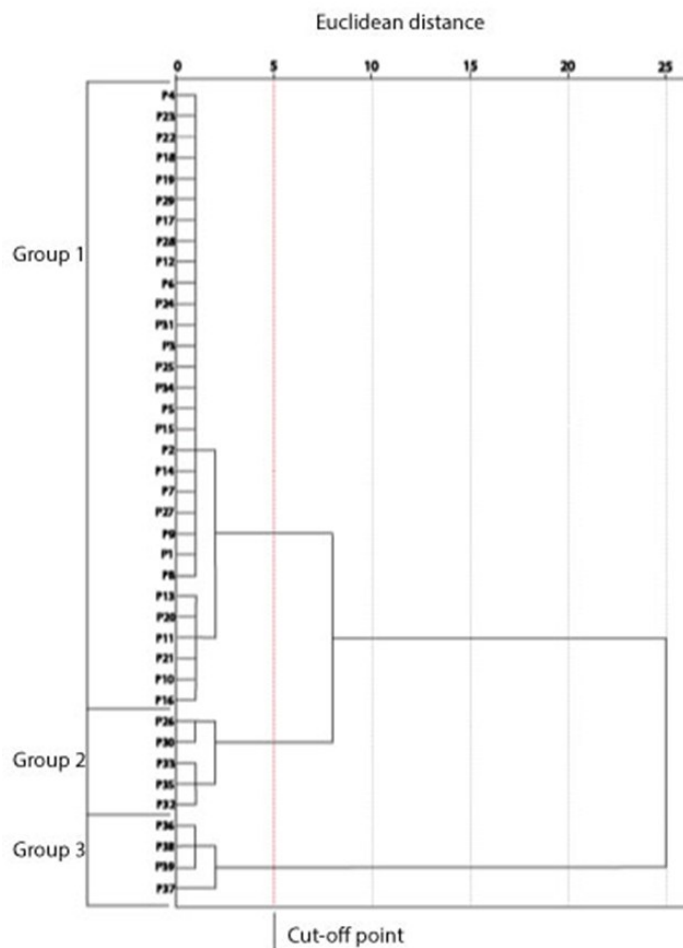


Figure 3. Dendrogram resulting from the hierarchical grouping analysis of the SJHM soils from 2014.

Group 1 exhibited the largest quantity of grouped points (~77%). Figure 3 shows that Groups 1 and 2 were close to each other and distant from Group 3. Groups 2 and 3 exhibited a small quantity of grouped points (approximately 12.8% and 10.3%, respectively).

According to Table 4, 50% of the points in Group 1 were classified as ground vegetation, 30% as arboreal-shrubby caatinga, 13% as urban areas, and 7% as agricultural areas. The points in Group 2 were mostly classified as ground vegetation (60%), followed by arboreal-shrubby caatinga (20%), and rainforest (20%). Group 3 contained only points classified as preserved savannah (50%) and rainforest (50%). Using the technique of hierarchical grouping, regarding land use and occupation, the

samples were classified into groups according to the state of environmental conservation of the system. Group 3 contained soils with preserved native vegetation, protecting the soil against compaction and mass losses. Group 2 contained soils with ground vegetation and shrubs with twisted branches, which in drier periods lose their leaves (a property useful for preventing water loss due to evaporation). This offers reduced soil protection against natural compaction and mass losses due to surface flow (ALVES; ARAÚJO; NASCIMENTO, 2009). Group 1 contained soil ranging from classes of vegetal formation, similar to Group 2, to classes of soil utilization with a more anthropic characteristic, such as urbanization and agriculture.

Table 4. Descriptive statistics of the variables and land use and occupation for the groups in the SJHM for 2014.

Group 1 Descriptive statistics with n=30				Land Use and Occupation	
	Alt. (m)	OM (g/kg)	M (%)	UO (%)	
Maximum	509.8	19.86	14.02	Ground Vegetation	50
Minimum	388.9	1.45	0.83	Urban	13
Mean	439.47	5.93	3.57	Caatinga	30
Standard deviation	26.03	4.03	2.74	Agriculture	7
CV (%)	5.92	68.04	76.83		
Group 2 Descriptive statistics with n=5				Land Use and Occupation	
	Alt. (m)	MO (g/kg)	U (%)	UO (%)	
Maximum	695.5	17.07	7.91	Ground Vegetation	60
Minimum	570.3	9.72	2.95	Caatinga	20
Mean	631.46	12.56	5.26	Rainforest	20
Standard deviation	47.49	2.87	2.08		
CV (%)	7.52	22.87	39.51		
Group 3 Descriptive statistics with n=4				Land Use and Occupation	
	Alt. (m)	MO (g/kg)	U (%)	UO (%)	
Maximum	952.2	37.86	15.93	Preserved savannah	50
Minimum	816.6	26.79	11.56	Rainforest	50
Mean	906.53	34.32	13.58		
Standard Deviation	61.49	5.17	1.92		
CV (%)	6.78	15.07	14.13		

Alt: altitude; OM: organic matter; M: moisture; UO: land use and occupation.

Regarding the organic-matter content, Groups 1, 2, and 3 exhibited mean values of 5.93, 12.56, and 34.32 g/kg (Table 4), respectively, corresponding to soil compositions of 0.59%, 1.26%, and 3.43%, respectively. According to the classification used by Fageria, Barbosa Filho and Zimmermann (1994), the soils of Groups 1 and 2 exhibited low organic-matter contents, with values lower than 1.5%. The soils of Group 3 exhibited an intermediate organic-matter content, with values between 1.5% and 4.5%. Regarding the moisture content, Groups 1, 2, and 3 exhibited mean values of 3.57%, 5.26%, and 13.58%, respectively (Table 4). These results are possibly related to the type of use and occupation of the soils in each group, considering that in the rainforest and preserved savannah areas, in Group 3, there is a greater input of organic matter and consequently greater water retention in the soil, favoring infiltration and reducing the risk of water erosion.

These results corroborate those obtained by Mendonça et al. (2009) and Araújo et al. (2013), who studied soils of different management types in Chapada do Araripe. According to the authors, preserved soils in rainforest and preserved savannah areas tend to have higher moisture and organic-matter contents and a lower degree of compaction of the surface layers compared with more anthropized

soils located in areas of forest management, exposed soil, agriculture, etc.

Groups 1, 2, and 3 had average altitudes of 439.47, 631.46, and 906.53 m, respectively (Table 4). Regarding this parameter, the groups were distinguished among the following regions of the SJHM: the top of Chapada do Araripe (Group 3), the slope (Group 2), and the valley (Group 1). This is evident in Figure 4, where we observe the spatialization of these groups and the types of soils in the SJHM. We also observe that the points of Group 3 occur in Neosol, Litholic, and Yellow Latosol soils, which are located at the top of the plateau and on the steepest part of the slope, in the surroundings of the topographic divider that limits the SJHM, with altitudes between 952 and 816 m.

The points of Group 2 are found in Litholic Neosol soils only, in an intermediate region between Groups 3 and 1, located in the slope of Araripe plateau, with altitudes varying between 696 and 570 m. The points of Group 1 appear in Yellowish Red Latosol, Fluvisol Neosol, Litholic Neosol, and Yellowish Red clayey soils and Urbanization, in the SJHM valley, in an area between the exutory and the slope of the plateau, with altitudes varying between approximately 509 and 389 m.

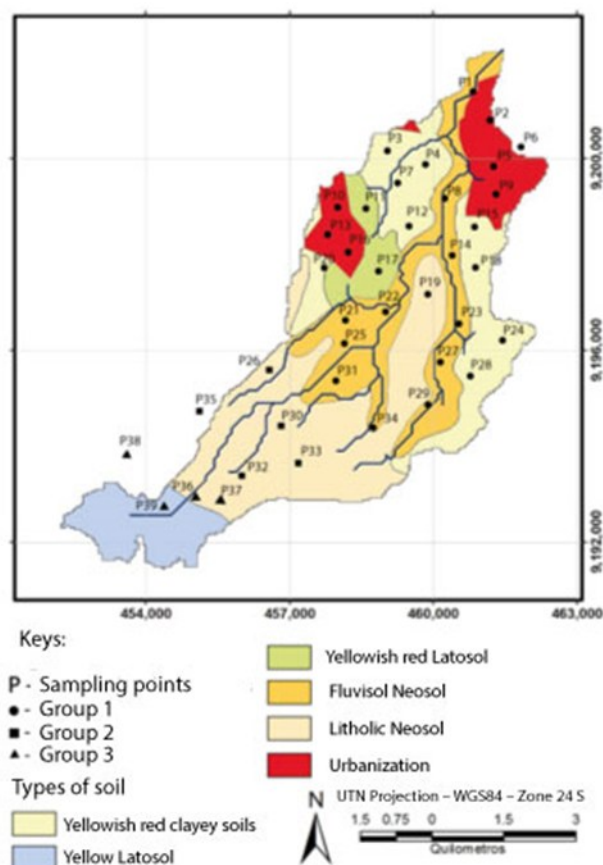


Figure 4. Spatialization of groups and types of soils in the SJHM in 2014.
Source: Adapted from Costa (2013)

CONCLUSIONS

Both the FA and the hierarchical grouping analysis contributed to the environmental understanding of the SJHM soil characteristics.

The FA reduced the number of SJHM soil characteristics from 10 to seven, resulting in two factors explaining 75.76% of the total data variance. The first factor is an indicator of variables controlling soil weathering, as well as anthropic processes associated with use and occupation. The second is an indicator of the structure and size of the soil granules and the topographic slope.

The hierarchical grouping analysis, which was applied to variables strongly correlated to the first FA factor, evidenced the occurrence of three groups. One of them, with samples at the top of the Chapada do Araripe in a preserved vegetation area, exhibited higher contents of organic matter and moisture, which are typical of environmentally preserved soils. Another group containing samples at the valley exhibited lower contents of organic matter and moisture, indicating more anthropized soils. Finally, the third group of samples was from the slope area between the top of the Araripe plateau and the valley, where the land use and occupation exhibit smaller vegetation and intermediate organic-matter and moisture contents.

The use of multivariate statistics was satisfactory with regard to its ability to classify, synthesize, and dynamically group samples according to the anthropic and topographical aspects, as well as the physical and chemical properties, of soils, allowing the characterization of areas with soils ranging from most preserved to most anthropized in a hydrological microbasin in a semiarid region.

REFERENCES

- ALVES, J. J. A.; ARAÚJO, M. A.; NASCIMENTO, S. S. Degradação da caatinga: Uma investigação ecogeográfica. *Revista Caatinga*, v. 22, n. 3, p. 126–135, 2009.
- ANDRADE, E. M. de et al. Fatores determinantes da qualidade das águas superficiais na Bacia do Alto Acaraú, Ceará, Brasil. *Ciência Rural*, v. 37, n. 6, p. 1791–1797, 2007.
- ARAÚJO, A. O. et al. Modificações nas propriedades dos solos de uma área de manejo florestal na Chapada do Araripe. *Revista Brasileira de Ciência do Solo*, v. 37, n. 3, p. 754–62, 2013.

- ARAÚJO, S. M. S. A Região Semiárida do Nordeste do Brasil: Questões ambientais e possibilidades de uso sustentável dos recursos. **Rios Eletrônica**, v. 5, n. 5, p. 89–98, 2011.
- BEZERRA, F. A.; CORRAR, L. J. Utilização da factorial analysis na identificação dos principais indicadores para a avaliação do desempenho financeiro: uma aplicação nas empresas de seguros. **Revista Contabilidade e Finanças**, v. 17, n. 42, p. 50–62, 2006.
- BÖHNER, J., ANTONIĆ, O. Land-surface parameters specific to topo-climatology. In: Hengl, T., Reuter, H.I. (Eds.), **Geomorphometry: Concepts, Software, Applications**. Elsevier Amsterdam, 2009, p. 195–226.
- BRODNJAK-VONCINA, D. et al. Chemometrics characterisation of the quality of river water. **Analytica Chimica Acta**, v. 462, n. 1, p. 87–100, 2002.
- COSTA, K. V. M. **Parâmetros hidrossedimentológicos da bacia hidrográfica do São José, no Cariri cearense**. 2013. 103 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Federal do Ceará, Fortaleza, Ceará, 2013.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA – EMBRAPA. **Centro Nacional de Pesquisa de Solos**. Manual de métodos de análises de solo. 2ª. ed. Rio de Janeiro: 1997.
- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA – EMBRAPA. **Serviço Nacional de Levantamento e Conservação de Solos (Rio de Janeiro, RJ)**. In: 10ª Reunião Técnica de Levantamento de Solos; 1979; Rio de Janeiro. Rio de Janeiro: 1979.
- FAGERIA, N. K., BARBOSA FILHO, M. P.; ZIMMERMANN, F. J. P. Caracterização química e granulométrica de solos de várzea de alguns estados brasileiros. **Pesquisa Agropecuária Brasileira**, v. 29, n. 2, p. 267–274, 1994.
- FUNDAÇÃO CEARENSE DE METEOROLOGIA E RECURSOS HÍDRICOS - FUNCEME, **Zoneamento Geoambiental do Estado do Ceará – Parte II: Mesorregião do Sul Cearense**. Fortaleza: Funceme: 2006.
- HAIR, J. F. et al. **Multivariate data analysis**. 6 ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2005. 593 p.
- HICKMANN, C.; COSTA, L. M. da. Estoque de carbono no solo e agregados em Argissolo sob diferentes manejos de longa duração. **Revista Brasileira De Engenharia Agrícola e Ambiental**, v. 16, n. 10, p. 1055–1061, 2012.
- KHWEDIM, K. et al. Trace metals in topsoils near the Babylon Cement Factory (Euphrates River) and human health risk assessment. **Environmental Earth Sciences**, v. 74, n. 1, p. 665–73, 2015.
- KÖPPEN, W.; GEIGER, R. **Klimate der Erde**. Gotha: **Verlag Justus Perthes**, 1928. Wall-map 150cmx200cm.
- MACHADO, P. L. O. A. Carbono do solo e a mitigação da mudança climática global. **Química Nova**, v. 28, n. 2, p. 329–334, 2005.
- MELO, R. O. et al. Susceptibilidade à compactação e correlação entre as propriedades físicas de um Neossolo sob vegetação de caatinga. **Revista Caatinga**, v. 21, n. 5, p. 12-17, 2008.
- MENDONÇA, L. A. R. et al. Avaliação da capacidade de infiltração de solos submetidos a diferentes tipos de manejo. **Engenharia Sanitária e Ambiental**, v. 14, n. 1, p. 89–98, 2009.
- NORUSIS, M. J. **SPSS Base System user's guide**. Chicago: SPSS, 1990. 520 p.
- RIOS, A. J. W. et al. Densidade de amostragem e complexidade de área em levantamentos pedológicos executados no Brasil. **Pesquisa Agropecuária Tropical**, v. 38, n. 1, p. 44–52, 2008.
- SALGADO, E. V. et al. Similaridade das variáveis hidroquímicas com o uso da análise multivariada, na Bacia do Salgado, Ceará. **Revista Caatinga**, v. 24, n. 3, p. 158–66, 2011.
- SINGH, V. B.; RAMANATHAN, A. L. Assessment of solute and suspended sediments acquisition processes in the Bara Shigri glacier meltwater (Western Himalaya, India). **Environmental Earth Sciences**, v. 74, n. 3, p. 2009–2018, 2015.
- STOLF, R.; FERNANDES, J.; FURLANI NETO, V. L. **Recomendações para uso de penetrômetro de impacto, modelo IAA/Planalsucar- Stolf**. São Paulo: MIC/IAA/PNMC – Planasulcar; 1983. (Série Penetrômetro de Impacto, BT1).
- TRINDADE, E. F. S.; VALENTE, M. A.; MOURÃO JÚNIOR, M. Propriedades físicas do solo sob diferentes sistemas de manejo da capoeira no nordeste paraense. **Agroecossistemas**, v. 4, n. 1, p. 50–67, 2012.