

K-means Based Soil Classification System Applicable to a Brazilian Mineral Province

Pablo dos Santos Cardoso Coelho

Universidade Federal de Ouro Preto - Campus Morro do Cruzeiro: Universidade Federal de Ouro Preto

Gustavo Henrique Nogueira

Universidade Federal de Ouro Preto - Campus Morro do Cruzeiro: Universidade Federal de Ouro Preto

Leonardo Alberto Sala

Universidade Federal de Ouro Preto - Campus Morro do Cruzeiro: Universidade Federal de Ouro Preto

Tatiana Barreto Santos (✉ tatiana.santos@ufop.edu.br)

Universidade Federal de Ouro Preto <https://orcid.org/0000-0001-5484-6675>

Research Article

Keywords: Soil classification system, tropical zone, data mining, geotechnics

Posted Date: December 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-484798/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

K-MEANS BASED SOIL CLASSIFICATION SYSTEM APPLICABLE TO A BRAZILIAN MINERAL PROVINCE

PABLO DOS SANTOS CARDOSO COELHO¹; GUSTAVO HENRIQUE NOGUEIRA¹; LEONARDO ALBERTO SALA¹; TATIANA BARRETO DOS SANTOS¹

¹Universidade Federal de Ouro Preto
School of Mines
Núcleo de Geotecnologia
Ouro Preto, Minas Gerais
Brazil

Corresponding author: Tatiana Barreto dos Santos, tatiana.santos@ufop.edu.br

Abstract: This article presents a geotechnical soil classification system proposed for application on soils of a tropical mineral province, located in *Minas Gerais* state, Brazil. The system was constructed using data mining techniques, i.e., principal component analysis and k-means cluster analysis, which were applied to a dataset composed of 101 geotechnical characterization laboratory test results of soils from the Province of *Quadrilátero Ferrífero*. The main objective of the proposed soil classification method was to establish a regional soil classification system, which encompass the interpretability of the main geotechnical parameters of soils by means of the classification, given the little explanatory capacity of the Unified Soil Classification System classification system for the performance of such task. It was possible to establish a chart for soil classification capable of explaining 81.68% of the variability of the analyzed parameters, being established the soil classes A, B and C for the studied soils.

Keywords: Soil classification system, tropical zone, data mining, geotechnics.

1. INTRODUCTION

Soil classification systems, including Unified Soil Classification System (USCS) (Das and Sobhan 2013) has high relevance in geotechnical engineering practice. They are highly widespread tools in geotechnical engineering for soil typification, used to the prediction and evaluation of different soil behaviors, such as evaluation of its applicability on earthworks and civil construction.

According to Fookes (1994), traditional geotechnical classification systems in general consider characteristics that are relevant in temperate soils, characterized by the absence of weathering present in tropical soils.

The Unified Soil Classification (USCS) system is largely applied to different types of soils. It takes into account, for soil classification, the granulometric and plastic characteristics. According to Fookes (1994), these variables are capable of satisfactorily typifying soils from temperate climates, being not equally sufficient to typify tropical soils, in which saprolitic or lateritic residual natures are constantly observed. These soils frequently present residual characteristics from the matrix rock and grain cementation. Geotechnical behaviors induced by these characteristics cannot be explained by means of plasticity and granulometry characteristics.

In the given context, Pinto (2006) emphasizes the relevance of regional soil characterization systems, such as the MCT (Miniature, Compacted, Tropical) system proposed by Nogami and Villibor (1981), which is applicable to classification

of tropical natural soils. Besides, there are informal soil classification systems used in local character, such as “red porous clay”, used in the city of São Paulo, Brazil. The great advantage of local classification systems use is its assertiveness, whereas the study area is restricted and, consequently, the variability of soil types grouped into classes is reduced.

Considering this discussion, this article presents a local soil classification system, in order to precisely understand the coverage of conventional soil classification systems. The classification system was developed for soils from mineral province of *Quadrilátero Ferrífero*, Minas Gerais state, Brazil. Laboratory tests were carried out on 101 natural and compacted soils samples from *Quadrilátero Ferrífero* and a dataset were constructed. It has as variables effective friction angle (ϕ'), cohesion (c'), plasticity index (PI), particle specific gravity (Gs) and fine content.

Principal component analysis (PCA) and k-means analysis were carried out in the dataset in order to construct the proposed classification system. PCA was applied in order to comprehending the variable interdependence, reducing the dimensionality of the data and graphically representing the data into two dimensions. The application of k-means clustering algorithm shows the occurrence of three well-defined groups of soils with similar behaviors. Then, a soil classification chart capable of typifying three distinct geotechnical behaviors was proposed. Carvalho and Ribeiro (2020) proposed a similar approach for the classification of partially saturated soils trough cone penetration test, suggesting that multivariate statistics is a useful tool for that purpose.

The use of data mining techniques (including on the techniques used in this research) is being successfully used to recognize patterns and purpose models in geotechnical and geological study topics, as shown the published research of Debnath and Dey (2017), Santos et al. (2018), Santos et al. (2019), Hou et al. (2020) and Do et al. (2021).

The present article is organized as follows. Topic 1 presents the introduction of the article. Topic 2 presents the dataset, the methodology used for the development of the proposed classification system. Topic 3 presents the background of techniques used to propose the soil system classification. Topic 4 shows the results and discussion and Topic 5 shows the conclusions of the research.

2. MATERIAL AND METHODS

2.1. Data organization

The dataset used to the elaboration of the proposed soil classification system is composed by 101 samples in which fine content, cohesion, solid density, plastic index and friction angle were measured. Laboratory tests was carried out on these samples of natural and compacted soils from different areas located in province of *Quadrilátero Ferrífero*. The carried-out testes were granulometric characterization, CD and CIU triaxial compression tests, determination of solid density (Gs) and Atterberg limits. These tests were carried out according to the recommendations of the standards NBR 7181:2016 (ABNT 2016a), NBR 6459:2016 (ABNT 2016b), NBR 7180:2016 (ABNT 2016c), NBR 6458:2016 (ABNT 2016d) and ISO 17892-9: 2018 (ISO 2018).

The samples used to compose the dataset are very varied in terms of granulometric, strength, density and plasticity characteristics, encompassing residual soils of phyllites, quartzites and itabirites, alluvial and colluvial soils, as well compacted embankments built with the same materials.

According to Das and Sobhan (2013), minerals with iron in its composition, mainly oxides, tend to have higher values of Gs when compared to silicate minerals, such as kaolinite and quartz. Stefanou and Papazafeiriou (2013) also discuss about the significant influence of the fine content and Gs on the strength of five types of soils, where a positive correlation between these two variables and the penetration strength was established.

Considering a preliminary geotechnical evaluation of a determined soil, it is known that the generalization of soils with higher granular contents (sands and gravels) tend to be harder than soils with higher plasticity and fine content (clay and silt) is not recommended, as stated by Das and Sobhan (2013). Then, this information should not be taken as a prerogative in design methods, as it is used in some empirical methods of highway design.

Therefore, the dataset was built based on 5 variables, i.e., cohesion (c') and friction angle (φ'), responsible to representing the strength characteristics of the analyzed materials, and plasticity (PI), fine content and particle specific gravity (Gs), responsible to representing physical characteristics of the analyzed materials.

Table 1 presents the first ten samples of the dataset. In order to carry out the propose the soil classification system, the data were standardized, as shown in Table 2.

Table 1. First ten samples of the dataset.

ID	Fine content (%)	c' (kPa)	Gs	PI (%)	φ' (°)
1	50.10	5.00	2.69	16.20	18.20
2	65.00	27.00	3.24	23.00	27.10
3	76.00	16.00	3.00	17.70	29.00
4	68.14	13.00	2.96	16.00	30.00
5	42.67	16.60	2.86	9.00	25.53
6	74.00	10.60	2.91	13.00	31.15
7	67.00	0.00	2.69	9.00	35.00
8	63.00	14.29	2.97	17.00	29.00
9	58.00	20.00	3.19	21.00	27.94
10	42.00	8.00	2.82	12.00	32.00

Table 2. First ten samples of the standardized dataset.

ID	Fine content (%)	c' (kPa)	Gs	PI (%)	φ' (°)
1	-0,91	-0,75	-0,84	0,30	-2,39
2	-0,07	1,59	0,98	1,27	-0,54
3	0,56	0,42	0,20	0,51	-0,15
4	0,11	0,10	0,04	0,27	0,06
5	-1,34	0,48	-0,27	-0,73	-0,87
6	0,44	-0,16	-0,10	-0,16	0,30
7	0,05	-1,29	-0,84	-0,73	1,10

8	-0,18	0,23	0,09	0,41	-0,15
9	-0,46	0,84	0,81	0,99	-0,37
10	-1,37	-0,44	-0,40	-0,30	0,48

2.2. Methodology

The methodology for obtaining the proposed soil classification system was divided in four steps, see Fig. 1. The first step consisted of the organization of data, including the removal of multivariate outliers and descriptive statistics of the variable dataset in order to define and understand the characteristics of the dataset. The second step consisted of the application of PCA in order to reduce de dimensionality of the dataset. The third step consisted of performing k-means cluster analysis, in which the number of clusters (k) was defined through a majority rule composed by 30 index proposed in the literature (Charrad et al. 2015). The fourth step consisted of the analysis of the generated clusters, definition of the classificatory limits of the proposed soil classification system and presentation and discussion of the results.

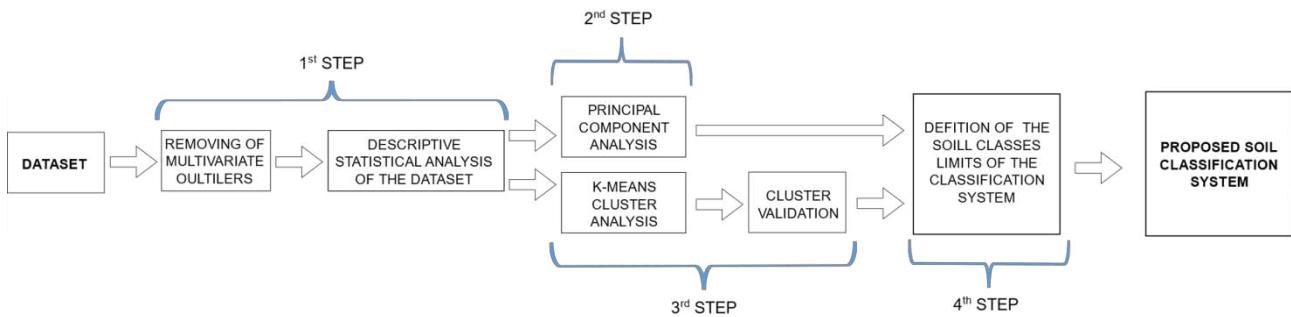


Fig. 1. Methodology

3. THEORY

3.1. Principal component analysis

Principal component analysis (PCA) consists of a statistical capable of reduce the dimensionality of a multidimensional dataset through linear combination of the original variables in a number equal to or less than the original named principal components, as proposed by Pearson in 1901 and formally presented by Hotelling (1933). These principal components are orthogonal, and, there is no correlation between them.

According to Hair et al. (2018), this multivariate technique presupposes the correlation between the variables. The technique can be carried out through covariance or correlation matrixes. Principal component analysis via correlation matrix was carried out in this research, in order to avoid bias in the analysis.

The i^{th} principal component (Y_i) is given by Eq.1.

$$Y_i = e_i^t Z = [e_1 \ e_2 \ e_3 \ \dots \ e_p] \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ \vdots \\ Z_p \end{bmatrix} \quad \text{Eq. 1}$$

Where e_i^t is the transposed eigenvector i of the correlation matrix of the data and Z is the vector of standardized original variable.

The explained variability of the i^{th} principal component is given by the i^{th} eigenvalue (λ_i) associated to the i^{th} eigenvector.

The number of principal components must to be retained is defined through Kaiser criterion (Kaiser 1960) and scree-plot (Cattell 1966). Kaiser (1960) suggests maintaining components corresponding to eigenvalues greater than 1, in the case of using the correlation matrix. Scree-plot consists in retaining the number of components defined by the inflection point of the graph.

3.1. K-means method

Cluster analysis is a statistical technique used to group elements by statistical similarity by means of a definition of a distance function used to determining the statistical distance between the samples.

K-means is a non-hierarchical method of cluster analysis, which consists of grouping n data points into predefined k clusters. The distance between the individuals of the same cluster must to be minimal, while the distance between clusters must be as high as possible (Sark and Gaben 2014). In general, the method can be described in the following four steps:

- selection of k random points, which represent the center of each cluster;
- allocation of each point of the input set into the nearest center k ;
- recalculation of the center of each cluster;
- iterative repetition of steps 2 and 3, until distances obtained in each iteration converge to a constant value.

The most used distance function is the Euclidean distance, given by Eq. 3:

$$d_{ij} = \sqrt{(x_i - x_j)^t(x_i - x_j)} \quad \text{Eq. 2}$$

Where:

d_{ij} is the Euclidean distance between i^{th} and j^{th} individuals;

x_i is the variable vector of i^{th} individual and

x_j is the variable vector of j^{th} individual.

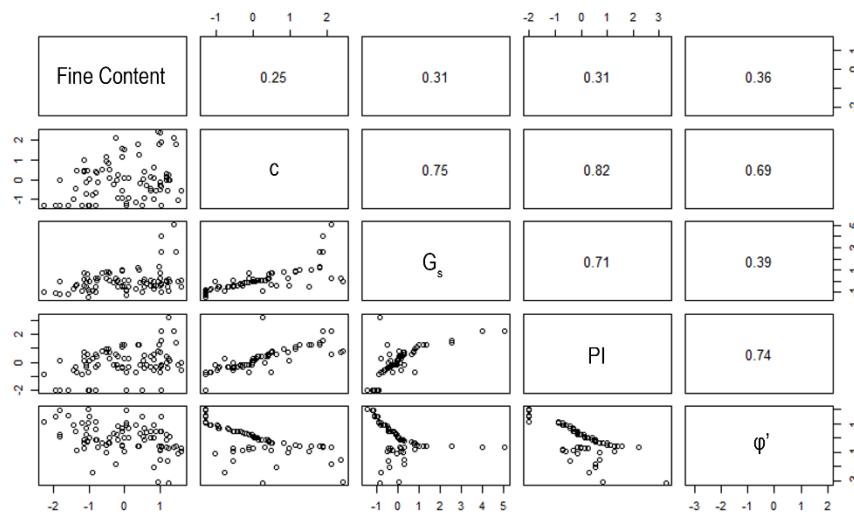
4. RESULTS AND DISCUSSION

Removal of multivariate outliers was performed through the analysis of Mahalanobis distance of the data points to the dataset centroid in order to obtain a realistic dataset through the exclusion of unreliable data. For that end, analysis of multivariate outliers was carried with the objective of determining the samples that should be taken from the dataset. 22 outliers were identified, of which 12 showed unreasonable behaviors for natural and compacted soils. Values of φ' above 40° were discarded on predominantly fine soils and c' values above 35 kPa were discarded on sandy soils, seeking adherence to values present in the literature, such as Das and Sobhan (2013) and Maiolino (1985). The outlier extraction resulted in a dataset with fine contents between 26% and 94.37%, cohesion values between 0 and 35 kPa, friction angle values varying between 14.60° and 39.40° , PI values between 0 and 37% and G_s values between 2.50 and 4.46. With the inconsistent data withdrawn from the dataset, 88 samples were considered for posterior analysis, which was carried out in order to propose a soil classification system. Table 3 presents the descriptive statistics of the variable dataset.

Table 3. Descriptive statistical summary of the analyzed dataset.

	Fine content (%)	c' (kPa)	G_s	PI (%)	φ' ($^{\circ}$)
Average	66.18	12.09	2.94	14.10	29.70
Standard deviation	17.60	9.40	0.30	7.00	4.82
1º quartile	51.52	4.61	2.80	11.0	26.97
3º quartile	83.12	16.39	3.00	18.00	32.84
Median	66.50	11.96	2.92	14.00	30.00
Minimum	26.00	0.00	2.50	0.00	14.60
Maximum	94.37	35.00	4.46	37.00	39.40

Analysis of correlations between the variables adopted in the present study was carried out, see Fig. 2. Significant correlation between the variables were observed in the correlation matrix, with values reaching 0.82, classified as a strong correlation (Callegari-Jaques 2003). Bartlett's test (Bartlett 1951) was performed and it was possible to conclude that the data present sufficient correlation to apply multivariate statistical techniques, with a p-value equal to 9.25×10^{-53} .

*Fig. 2. Scatter plot.*

PCA via correlation matrix was carried out in order to reduce the dimensionality of the data. Table 4 shows the loadings (coefficients of linear combinations) of the five generated principal components.

Table 4. Principal components loadings.

PC	Fine content	c'	G_s	PI	φ'	Percentage of explained variance	Cumulative percentage of the explained variance

PC ₁	0.266	0.509	0.451	0.516	-0.448	64,578	64,578
PC ₂	0.938	-0.254	-0.166	-0.150	-0.070	17,098	81,676
PC ₃	0.165	0.016	0.677	-0.073	0.713	12,097	93,773
PC ₄	0.056	0.679	-0.035	-0.728	-0.070	3,545	97,317
PC ₅	0.134	0.464	-0.557	0.419	0.530	2,683	100

Kaiser criterion (Kaiser 1960) and the scree-plot analysis (Cattell 1966) were used to determining the number of principal components that should be kept in the analysis, see Fig. 3. They suggested the retention of only one principal component. Although there is an indication of the retention of only one variable, the proposed soil classification system also considers the second principal component in its conception, once the second principal component has a high explanatory character of the geotechnical behavior of the data. Principal components 1 and 2 are given by Eq. 6 and Eq. 7, which take standardized variables as input parameters.

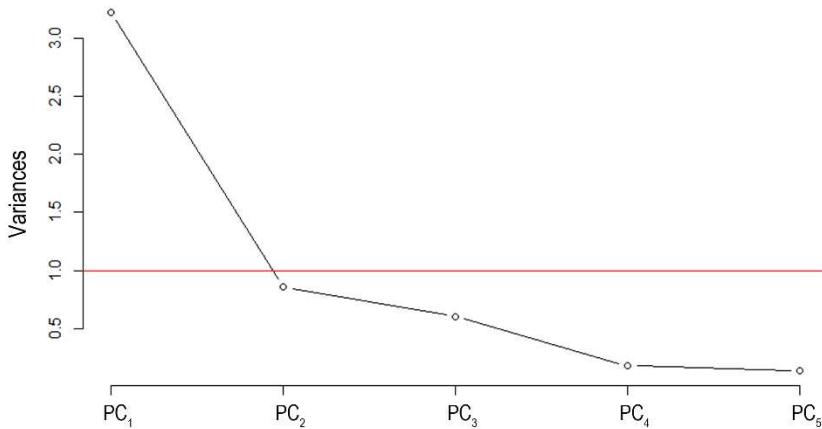


Fig. 3. Scree plot.

$$PC_1 = 0,265 \times Fine\ Content + 0,509 \times c' + 0,450 \times G_S + 0,516 \times PI - 0,450 \times \varphi' \quad Eq. \ 3$$

$$PC_2 = 0,938 \times Fine\ Content - 0,254 \times c' - 0,166 \times G_S - 0,150 \times PI - 0,070 \times \varphi' \quad Eq. \ 4$$

First principal component has a positive correlation with fine content, cohesion, GS and PI, and a negative correlation with the friction angle. It is also important to verify that the correlations present, in module, similar values, except for fines content, which is about half of the others. Second principal component has a very high correlation with fines content, which is the main factor controlling its behavior. The other variables present minor importance for the definition of this principal component. Fig. 4 resumes these information in a biplot graph.

Usage of these two components encompasses the importance of all variables, being capable of explaining 81.68% of original data variability.

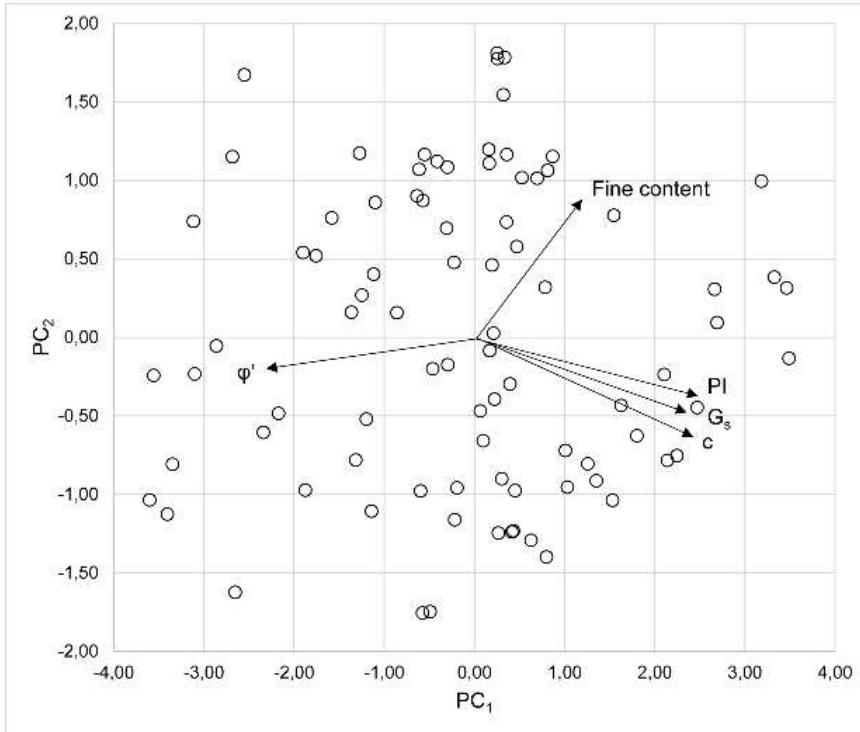


Fig. 4. Biplot graph.

The clusterization of the data was performed through k-means method, see Fig. 5. The distance used to calculation of dissimilarity of the samples was Euclidean distance and Mahalanobis distance. The first one presented best performance and fit and was used to proposed the new soil classification system. The number of clusters adopted was equal to three, as recommended by the majority rule (Charrad et al. 2015), based on 30 index used to determine the optimal number of clusters, see Fig. 6. The validation of the cluster analysis was based visual inspection of the graph of the Fig. 5. The groups are well defined with a very small overlapping region

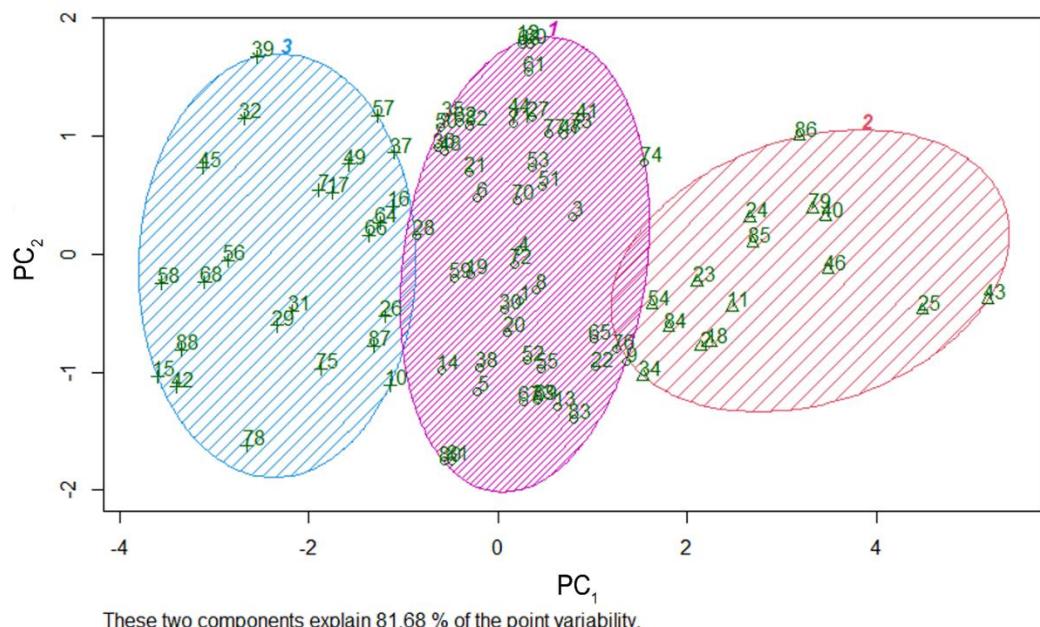


Fig. 5. Result of the cluster analysis.

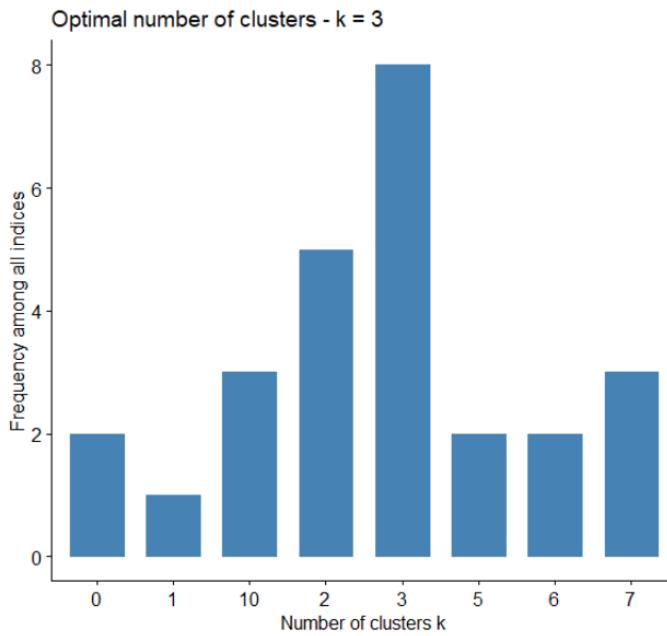


Fig. 6. Application of majority rule to determining the optimal number of clusters for the data.

The characteristics of the samples that composes each cluster were analyzed in order to label each cluster, defining three material classifying regions in the Cartesian space of principal components 1 and 2. Linear contours, α and β , given by Eq. 6 and Eq. 7, were proposed for the definition of the boundary between the soil classes. The proposed contours consist of linear limits that intercept the intersection points of the clusters. The soils classes were labeled as A, B and C classes. The proposed classification graph is shown in Fig. 7.

$$\alpha: PC_2 = 12,780 \times PC_1 + 12,198 \quad Eq. 5$$

$$\beta: PC_2 = 4,768 \times PC_1 + 7,278 \quad Eq. 6$$

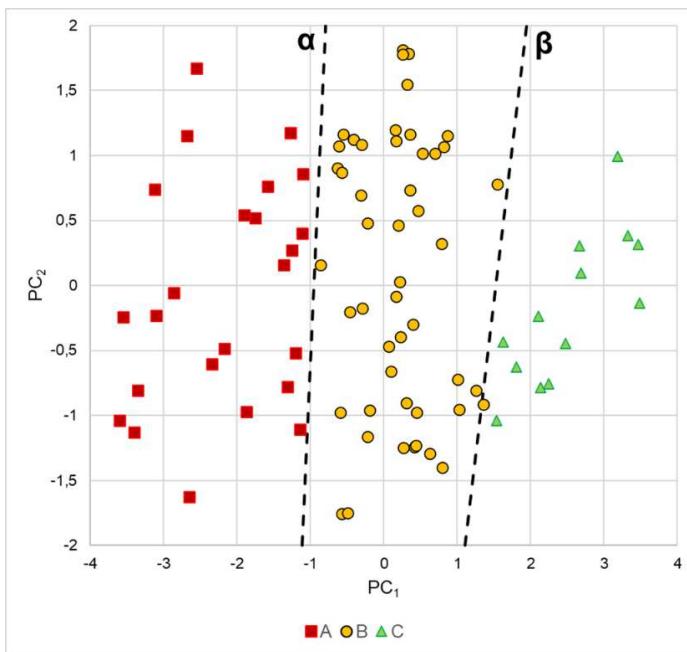


Fig. 7. Proposed soil classification system.

The description of the proposed soil classes of the classification system are presented below.

- Class A: soils of this classes are the one with the lowest values of PI and Gs. They present values of Gs lower than 2.82, which indicate the predominance of minerals with lower unit weight, such as quartz ($G_s = 2.65$) and kaolinite ($G_s = 2.60$). According to (Das and Sobhan 2013), these type of soils presents low activity values and, therefore, low plasticity. Moreover, the predominance of low unit weight minerals in class A, indicates low iron content and, consequently, low cementation. This information is corroborated by Stefanou and Papazafeiriou (2013) and Cruz et al. (2013), which discuss the positive influence of cementation on soil cohesion. Besides, this type of soil presents a range of fine content 26% to 85%, where 75% of the samples have fine content less than 67.25%. The fine content of these soil samples and the presence of minerals with G_s lower than 2.82 indicate the predominance of sandy characteristics of these materials. These characteristics lead to highest values of ϕ' observed (Das and Sobhan 2013). It is, therefore, possible to infer that class A soils correspond to materials with a sandy to inactive fine matrix, with concentration of low unit weight minerals and low cementation.
- Class B: soils with transitory behavior between classes A and C. For these materials, intermediate values of Gs and PI are indicative of the occurrence of a higher percentage of active clay minerals and higher density – when compared to the observable values for class A. It is noticeable the occurrence of more expressive cohesion for these materials, indicative of cementation between the grains. The occurrence of moderate friction angles, however, corroborates the observation of granular fractions of not less than 16% in 75% of the observations. Class B soils, therefore, can be understood as transitive materials between classes A and C, with intermediate behaviors.
- Class C: Materials encompassed in class C present the opposite behavior to class A materials for the five variables analyzed. It can be seen, therefore, that class A tends to encompass predominantly fine materials with clayey and active characteristic, for which the design concerns often inherent to CH type materials, distributed between classes A, B and C, are applicable. The higher cohesion values observed for class C materials can be associated with pre-consolidation and cementation phenomena (Cruz et al. 2013), the latter derived from the presence of high-density minerals, such as hematite and goethite, justified by observation of G_s values greater than 3.05 in 75% of the analyzed observations. The high values of c' perceived for class C, however, are contrasted by materials of low geotechnical competence, for which the possibility of non-processing of the previously mentioned phenomena is indicated.

The samples of the dataset were classified according to the Unified Soil Classification System (USCS) and compared with the proposed soil classification system, see Fig. 8.

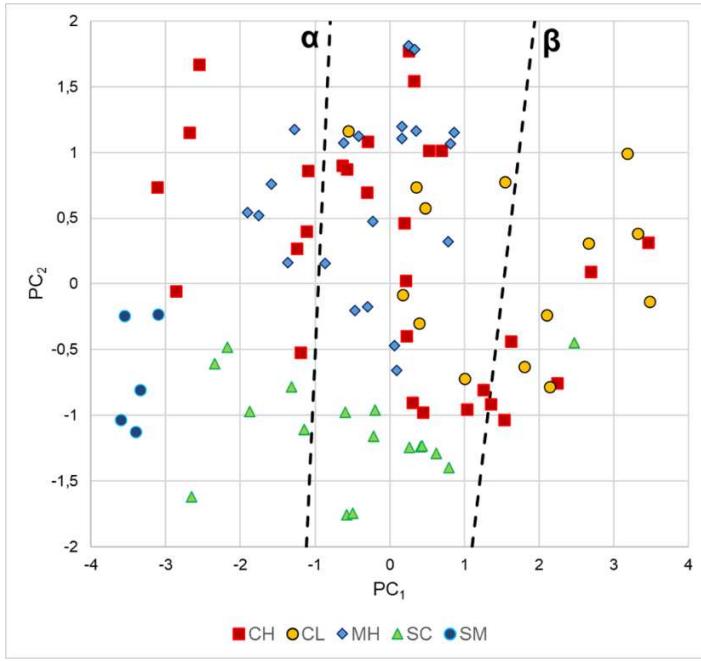


Fig. 8. USCS versus proposed soil classification system of the dataset samples.

Soils classified by USCS as silty sands (SM) were allocated in the region corresponding to class A of the proposed classifications system. Soils with high and low plasticity clays (CH and CL), high plasticity silts (MH) and clayey sands (SC), considering USCS, presented a significant dispersion in the proposed classification system. Each USCS class were allocated in at least two classes of the proposed chart, see Fig. 8. CH and CL soils were allocated in the region corresponding to class 3 of the proposed classification system. Therefore, the mentioned observations point out that the USCS is not efficient in the differentiation of the geotechnical behavior of the the analyzed soils, as, for example, high plasticity clays, high plasticity silts, low plasticity clays, and clayey sands have similar behaviors in the region between the α and β lines.

Regarding the dataset, boxplots of the A, B and C classes of proposed classification system and boxplots of SM, SC, MH, CH and CL classes of USCS system were drawn, considering φ' , c' , G_s , PI and fine content variables. Fig. 9 to Fig. 13 present the boxplots.

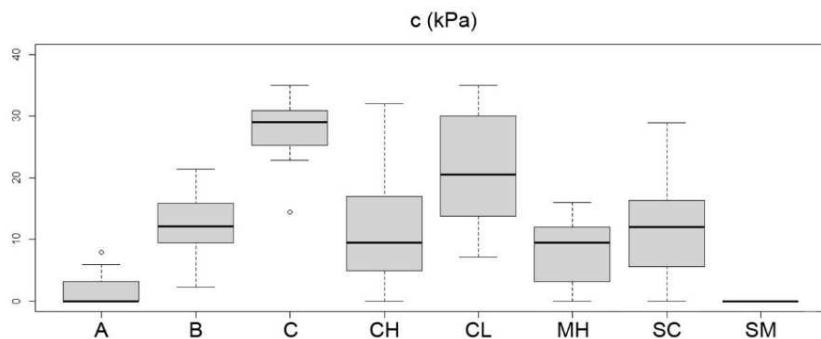


Fig. 9. Boxplots of c' values for the analyzed classes.

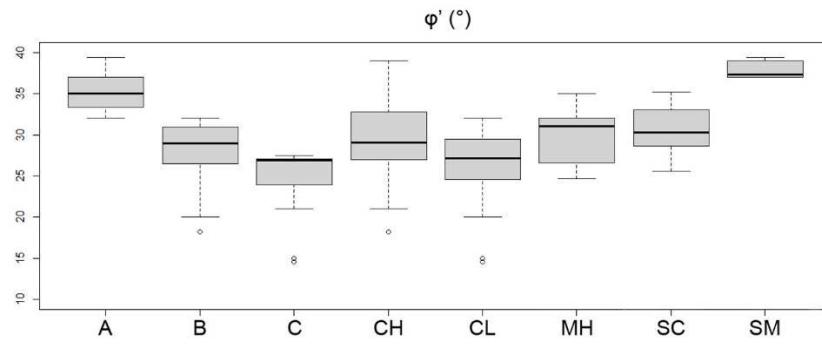


Fig. 10. Boxplots of ϕ' values for the analyzed classes.

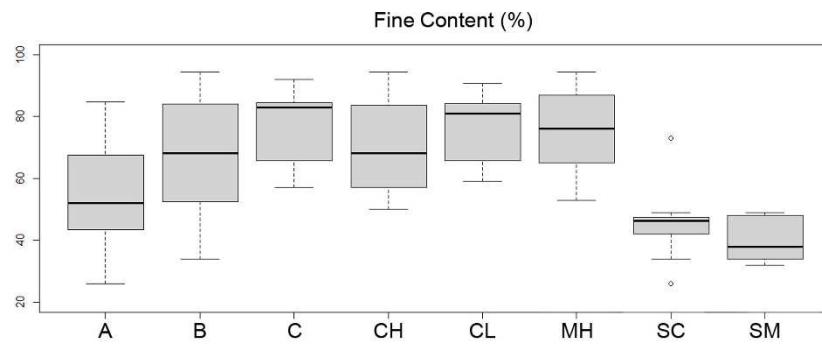


Fig. 11. Boxplots of fine content values for the analyzed classes.

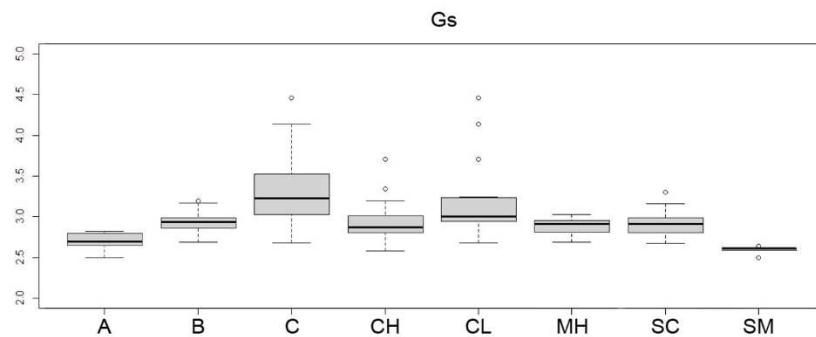


Fig. 12. Boxplots of Gs values for the analyzed classes.

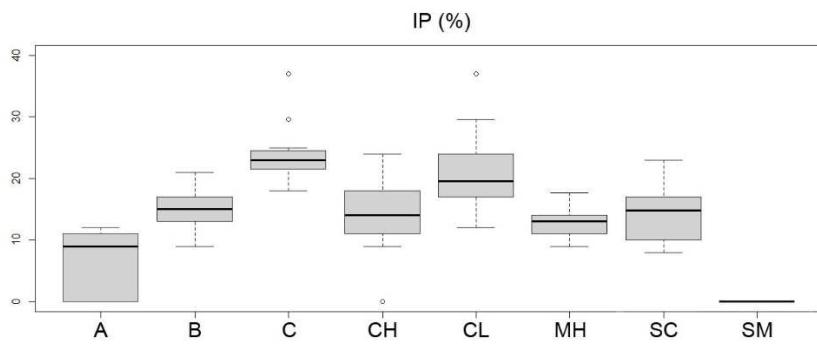


Fig. 13. Boxplots of PI values for the analyzed classes.

A, B and C classes of the proposed classification system presented distinct geotechnical behaviors, which were well demarcated by the different ranges of ϕ' , c' , PI, Gs and fine content for each class. In case of USCS classes, they present random scattered values of ϕ' , c' , PI, Gs and fine content for each class, see Fig. 6 to 10. It is possible to note that clays, silts and sands are from different classes considering USCS, but they present similar characteristics, from strength and behavior point of view, indicating the non-applicability of UCSC for the soils of mineral province of Quadrilátero Ferrífero. The boxplots presented in Fig.6 to 10 are capable of demonstrate the capability of the proposed soil system in discriminating the soils according to its strength behaviors.

6. CONCLUSIONS

Multivariate statistical techniques are powerful tools for proposition of accurate methods and validation of techniques consolidated in geotechnical engineering practice. To propose the soil classification system applicable to Quadrilátero Ferrífero province, principal component analysis was used in order to reducing the dimensionality of the data and analyze the interdependence between the variables. It was also used k-means cluster analysis, which allows the definition of soil classes (generated clusters) and the boundaries α and β of the proposed classification system, which defines different behaviors from strength point of view.

The analyzes conducted in this article show that the USCS system is not efficient for the classification of soils in the Quadrilátero Ferrífero province, as by Fortes et al. (2002). Whereas, the proposed classification system is very efficient in classifying soils from *Quadrilátero Ferrífero* according to its geotechnical behaviors and it is capable of represent the variability of the soils from this mineral province.

The adoption of the proposed classification system for soil classification could allow the adoption of specific design guidelines for classes A, B and C, which would contribute to geotechnical engineering practices. This study is regional and not very restrictive, and it should be carefully used in engineering projects, due to the limited number of samples considered.

It is recommended the conduction of studies that correlates other geotechnical behaviors to the A, B and C classes and that allow the practical improvement of geotechnical engineering with its application.

REFERENCES

- ABNT (2016a) Nbr 7181:2016 versão corrigida 2:2018: Soil - grain size analysis. 12 p.
- ABNT (2016b) Nbr 6459:2016 versão corrigida 2017: Soil - soil - liquid limit determination. 5 p
- ABNT (2016c) Nbr 7180:2016: Soil — plasticity limit determination. 3 p.
- ABNT (2016d) Nbr 6458:2016 versão corrigida 2:2017: Gravel grains retained on the 4,8 mm 269 mesh sieve - determination of the bulk specific gravity, of the apparent specific gravity and of 270 water absorption. 10 p
- Bartlett MS (1951) The effect of standardization on a chi square approximation in factor analysis. *Biometrika* 38:337–344
- Callegari-Jaques SM (2003) Bioestatística: princípios e aplicações. Artesmed
- Carvalho LO, Ribeiro DB (2020) Application of kernel k-means and kernel x-means clustering to obtain soil classes from cone penetration test data. *Soils and Rocks* 43:607–618. <https://doi.org/https://doi.org/10.28927/SR.434607>
- Cattell RB (1966) The Scree Test For The Number Of Factors. *Multivariate Behav Res* 1:245–276
- Charrad M, Ghazzali N, Boiteau V, Niknafs A (2015) Nbclust: Determining the best number of clusters in 285 a data set. *J Stat Softw* 61:1–36
- Cruz N, Rodrigues C, Foneca A V. (2013) An Approach to Derive Strength Parameters of Residual Soils from DMT Results. *Soils and Rocks* 37:195–209
- Das BM, Sobhan K (2013) Principles of geotechnical engineering, 8th edn. Cengage Learning, Stamford
- Debnath P, Dey AK (2017) Prediction of Laboratory Peak Shear Stress Along the Cohesive Soil–Geosynthetic Interface Using Artificial Neural Network. *Geotech Geol Eng* 35:445–461. <https://doi.org/10.1007/s10706-016-0119-2>
- Do H-D, Pham V-N, Nguyen H-H, et al (2021) Prediction of Unconfined Compressive Strength and Flexural Strength of Cement-Stabilized Sandy Soils: A Case Study in Vietnam. *Geotech Geol Eng*. <https://doi.org/https://doi.org/10.1007/s10706-021-01805-z>
- Fookes PG (1994) A review: Genesis and classification of tropical residual soils engineers. In: *Geotechnics in the African Environment*. pp 423–442
- Fortes RM, Merighi J V., Neto AZ (2002) No TitleMétodo das Pastilhas para Identificação Expedita de Solos Tropicais. In: *Anais do 2º Congresso Rodoviário Português*. Lisboa
- Hair JF, Black WC, Anderson RE, Babin BJ (2018) Multivariate Data Analysis, 8th Editio
- Hottelling H (1933) Analysis of a complex of statistical variables into principal components. . *J Educ Psychol* 24:417 – 441
- Hou M lei, Ding W tao, Wang X gang, et al (2020) Analysis of the Soil Loss and Settlement of Silty Clay with Different Subclassifications. *Geotech Geol Eng* 38:5819–5838. <https://doi.org/10.1007/s10706-020-01395-2>
- ISO (2018) ISO 17892-9:2018: Gravel grains retained on the 4,8 mm mesh sieve - geotechnical 297 investigation and testing — laboratory testing of soil — part 9: Consolidated triaxial compression 298 tests on water saturated soils.

25 p.

- Kaiser HF (1960) The application of electronic computers to factor analysis. *Educ Psychol Meas* 20:141–151
- Maiolino AL (1985) Resistência ao Cisalhamento de Solos Compactados: Uma Proposta de Tipificação. Universidade Federal do Rio de Janeiro
- Nogami JS, Villibor DF (1981) Uma Nova Classificação de Solos Para Finalidades Rodoviárias. In: Proc. Simpósio Brasileiro de Solos Tropicais em Engenharia. ABMS. pp 30–41
- Pinto CS (2006) Curso Básico de Mecânica dos Solos em 16 Aulas. Oficina de Textos, São Paulo
- Santos AEM, Lana MS, Cabral IE, et al (2018) Evaluation of Rock Slope Stability Conditions Through Discriminant Analysis. *Geotech Geol Eng* 37:775–802
- Santos TB, Lana MS, Pereira TM, Canbulat I (2019) Quantitative hazard assessment system (Has-Q) for open pit mine slopes. *Int J Min Sci Technol* 29:419–427
- Sark S, Gaben M (2014) Large Scale and Big Data: Processing and Management. CRC Press, Boca Raton
- Stefanou S, Papazafeirou A (2013) Effects of spatial variability on soil liquefaction: Some design recommendations. *Eurasian J Soil Sci* 2:122–130
- Zare Naghadehi M, Jimenez R, KhaloKakaie R, Jalali SME (2013) A new open-pit mine slope instability index defined using the improved rock engineering systems approach. *Int J Rock Mech Min Sci* 61:1–14. <https://doi.org/10.1016/j.ijrmms.2013.01.012>

7. DECLARATIONS

Funding: not applicable

Conflicts of interest/Competing interests: not applicable

Availability of data and material: Yes, the data can be made available after acceptance of the article

Code availability: Yes, the data can be made available after acceptance of the article