ISSN: 2447-3359

REVISTA DE GEOCIÊNCIAS DO NORDESTE

Northeast Geosciences Journal

v. 11, nº 1 (2025)

https://doi.org/10.21680/2447-3359.2025v11n1ID37229

A new soil classification system for the quadrilátero ferrífero province using multivariate statistical analysis

Um novo sistema de classificação de solos para a província do quadriláte- ro ferrífero usando análise estatística multivariada

Gustavo Henrique Nogueira[1]; Leonardo Alberto Sala[2]; José Matheus Vieira Matos[3] ; Tatiana Barreto dos Santos[4]

[1] Departamento de Engenharia de Minas, Universidade Federal de Ouro Preto. Morro do Cruzeiro Campus, s/n – Bauxita, CEP: 35400-000, Ouro Preto, Minas Gerais, Brazil. Email: ghnogueira89@gmail.com
ORCID: https://orcid.org/0000-0003-2203-727X

[2] Mining Engineer at Vale S/A, Geomechanical and Hydrogeological Management, Mina Águas Claras, CEP: 34000-000, Nova Lima, Minas Gerais, Brazil. Email: leonardo.sala@aluno.ufop.edu.br
ORCID: https://orcid.org/0000-0002-6052-8427

[3] Departamento de Engenharia de Minas, Universidade Federal de Ouro Preto. Morro do Cruzeiro Campus, s/n – Bauxita, CEP: 35400-000, Ouro Preto, Minas Gerais, Brazil. Email: jose.matos@aluno.ufop.edu.br
ORCID: https://orcid.org/0000-0002-3700-166X

[4] Departamento de Engenharia de Minas, Universidade Federal de Ouro Preto. Morro do Cruzeiro Campus, s/n – Bauxita, CEP: 35400- 000, Ouro Preto, Minas Gerais, Brazil. Email: tatiana.santos@ufop.edu.br
ORCID: https://orcid.org/0000-0001-5484-6675

Abstract: The Unified Soil Classification System (USCS) is widely used in geotechnical engineering, but it has limitations in classifying tropical soils due to characteristics such as the presence of saprolites or laterites. This study proposes a regional soil classification system for the mineral province of the Quadrilátero Ferrífero, in Brazil, aiming to enhance the understanding of soil behavior in the region. A total of 101 soil samples, both natural and compacted, were analyzed, including variables such as effective friction angle, cohesion, plasticity index, specific gravity of particles, and fines content. Principal Component Analysis (PCA) and kmeans clustering were used to develop the proposed classification system. This categorizes the soils into three distinct classes (A, B, and C) based on their geotechnical parameters. The analysis showed that the proposed system outperforms the USCS in differentiating soil behavior in the Quadrilátero Ferrífero province, establishing a classification chart that explains 81.68% of the variability of the analyzed parameters. Compared to the USCS, the new system provides a more accurate tool for predicting soil behavior, being useful in foundation engineering, excavation projects, and other geotechnical applications in the region.

Keywords: Soil classification system; Quadrilátero Ferrífero; Multivariate statistical.

Resumo: O Sistema Unificado de Classificação de Solos (USCS) é amplamente utilizado na engenharia geotécnica, mas tem limitações na classificação de solos tropicais, devido a características como a presença de saprólitos ou lateritas. Este estudo propõe um sistema regional de classificação de solos para a província mineral do Quadrilátero Ferrífero, no Brasil, com o objetivo de aprimorar a compreensão do comportamento dos solos na região. Foram analisadas 101 amostras de solo, tanto naturais quanto compactadas, incluindo variáveis como ângulo de atrito efetivo, coesão, índice de plasticidade, gravidade específica das partículas e teor de finos. Utilizou-se a Análise de Componentes Principais (PCA) e a clusterização k-means para desenvolver o sistema de classificação proposto. Este categoriza os solos em três classes distintas (A, B e C) com base em seus parâmetros geotécnicos. A análise mostrou que o sistema proposto supera o USCS na diferenciação do comportamento dos solos na província do Quadrilátero Ferrífero, estabelecendo um gráfico de classificação que explica 81,68% da variabilidade dos parâmetros analisados. Comparado ao USCS, o novo sistema oferece uma ferramenta mais precisa para prever o comportamento dos solos, sendo útil em engenharia de fundações, projetos de escavação e outras aplicações geotécnicas na região.

Palavras-chave: Sistema de classificação de solos; Quadrilátero Ferrífero; Estatística multivariada.

Received: 06/08/2024; Accepted: 21/02/2025; Published: 08/04/2025.

1. Introduction

Soil classification systems are essential tools in geotechnical engineering for characterizing soil behavior. These systems, including the widely used Unified Soil Classification System (USCS), aid in predicting soil performance for various applications in earthworks and civil construction. However, traditional systems, as noted by Fookes (1994), often focus on characteristics relevant to temperate soils, which lack the intense weathering processes common in tropical regions.

The USCS, for example, relies on grain size and plasticity for classification. While effective for temperate soils (FOOKES, 1994), these properties may not adequately capture the behavior of tropical soils, which are frequently composed of saprolites or laterites. These residual soils inherit characteristics from the parent rock, including residual textures and cementation between grains. These unique features can significantly influence geotechnical behavior, and their impact cannot be fully explained by just plasticity and grain size.

Pinto (2006) emphasizes the importance of regional classification systems that account for these specificities. The MCT (Miniature, Compacted, Tropical) system proposed by Nogami and Villibor (1981) exemplifies such an approach, specifically designed for tropical soils. Additionally, informal local systems, like the "red porous clay" classification used in São Paulo, Brazil, can be valuable for specific regions due to their focus on a limited area with reduced soil type variability.

In light of these limitations of conventional systems, this study proposes a local soil classification system specifically tailored to soils from the *Quadrilátero Ferrífero* mineral province in *Minas Gerais*, Brazil. This system aims to provide a more precise understanding of soil behavior compared to existing methods. A dataset of 101 natural and compacted soil samples from the region was compiled, with variables including effective friction angle, cohesion, plasticity index, particle specific gravity, and fine content.

Principal Component Analysis (PCA) and k-means clustering were employed to analyze the dataset and develop the proposed classification system. PCA helped identify relationships between variables, reduce data dimensionality, and visually represent the data in two dimensions. K-means clustering then revealed three distinct groups of soils with similar geotechnical behaviors. Based on this analysis, a soil classification chart was developed to categorize soils into these three distinct behavioral groups. This approach is similar to the one used by Carvalho and Ribeiro (2020) for classifying partially saturated soils using cone penetration tests, further highlighting the effectiveness of multivariate statistics in soil classification.

The remaining sections of this article are structured as follows: Section 2 details the dataset, methodology used for system development, and reviews PCA and k-means clustering methods. Section 3 provides background information on these techniques. Section 4 presents the results and discussion, and Section 5 concludes 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 mines 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 (ABNT, 2016a; ABNT, 2016b; ABNT, 2016c; ABNT, 2016d; 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 kwon 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. The first ten samples of the dataset are presented in Table 1. To implement the proposed soil classification system, the data were standardized (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

Source: Author (2024).

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

Source: Author (2024).

2.2. Methodology

The methodology for developing the proposed soil classification system was structured into four stages (see Figure 1). In the first stage, the data were organized and subjected to a descriptive statistical analysis to understand their main characteristics. Additionally, multivariate outliers were removed using the Mahalanobis distance (MAHALANOBIS, 1928), a metric that considers the correlation between variables and allows the identification of outliers in multidimensional spaces. This approach, based on the mean and covariance matrix of the variables, enables a more accurate assessment of how distant a data point is from the overall distribution. In the second stage, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the dataset, facilitating interpretation and optimizing subsequent processes.

The third stage of the study consisted of applying the k-means clustering analysis, a widely used method for data segmentation based on their similarities (AHMED, SERAJ, and ISLAM, 2020). The k-means algorithm aims to partition a dataset into *k* clusters, associating each data point with the cluster whose centroid is closest. However, as highlighted by

Ahmed, Seraj, and Islam (2020), defining the optimal number of clusters (*k*) is one of the main challenges of this approach, as an inadequate value can compromise the classification quality.

To determine this optimal number, a majority rule was adopted, based on the evaluation of 30 different statistical indices recommended in the literature (CHARRAD et al., 2015). These indices are quantitative metrics that assist in analyzing the compactness and separation of the clusters, contributing to the selection of the most appropriate value for *k*. Among the 30 indices used, the Silhouette Index stands out, measuring how well a data point fits within its cluster and how distinct it is from the others, assigning a value between -1 and 1, where higher values indicate better clustering quality (DINH, FUJINAMI, and HUYNH, 2019).

Additionally, the Calinski-Harabasz Criterion was employed to calculate the ratio between inter-cluster dispersion and intra-cluster dispersion, indicating better-defined clusters as its value increases. The Davies-Bouldin Index was also considered, assessing cluster compactness and separation, where lower values indicate better-defined clusters. Finally, the Within-cluster Sum of Squares (WSS) method, which measures the internal variability of clusters, was used. This method is often combined with the Elbow Method, a technique employed to identify the point at which increasing the number of clusters no longer significantly improves segmentation.

The majority rule was applied by identifying the value of *k* that was most frequently indicated as optimal among the different indices. Thus, the final number of clusters was determined in a more robust manner, minimizing subjectivity and ensuring that the clustering process accurately represented the structure of the data.

Finally, in the fourth stage, the generated clusters were analyzed, the classification limits of the proposed soil classification system were established, and the results were presented and discussed.

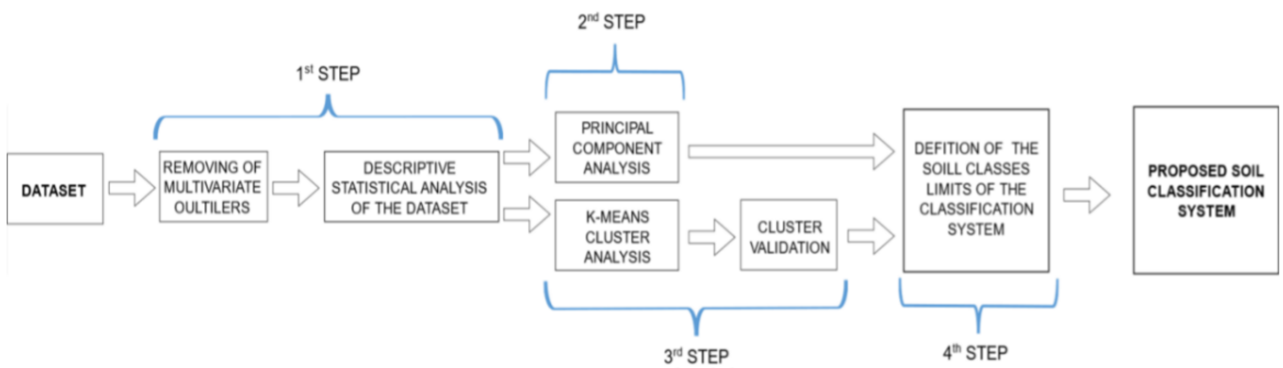

Figure 1 – Methodology.
Source: Author (2024).

3. 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 (DAS and SOBHAN, 2013; 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 14.60º and 39.40º, PI values between 0 and 37% and G_S values between

2.50 and 4.46. With the inconsistent data withdrawn from the dataset, 89 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)	Gs	PI (%)	φ' (°)
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

Source: Author (2024).

Analysis of correlations between the variables adopted in the present study was carried out (Figure 2). Significant correlation between the variables were observed in the correlation matrix, with values reaching 0.82, classified as a strong correlation (DOWDY, WEARDON and CHILKO, 2004). Bartlett's test (HAIR et al., 2009) 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}. 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.

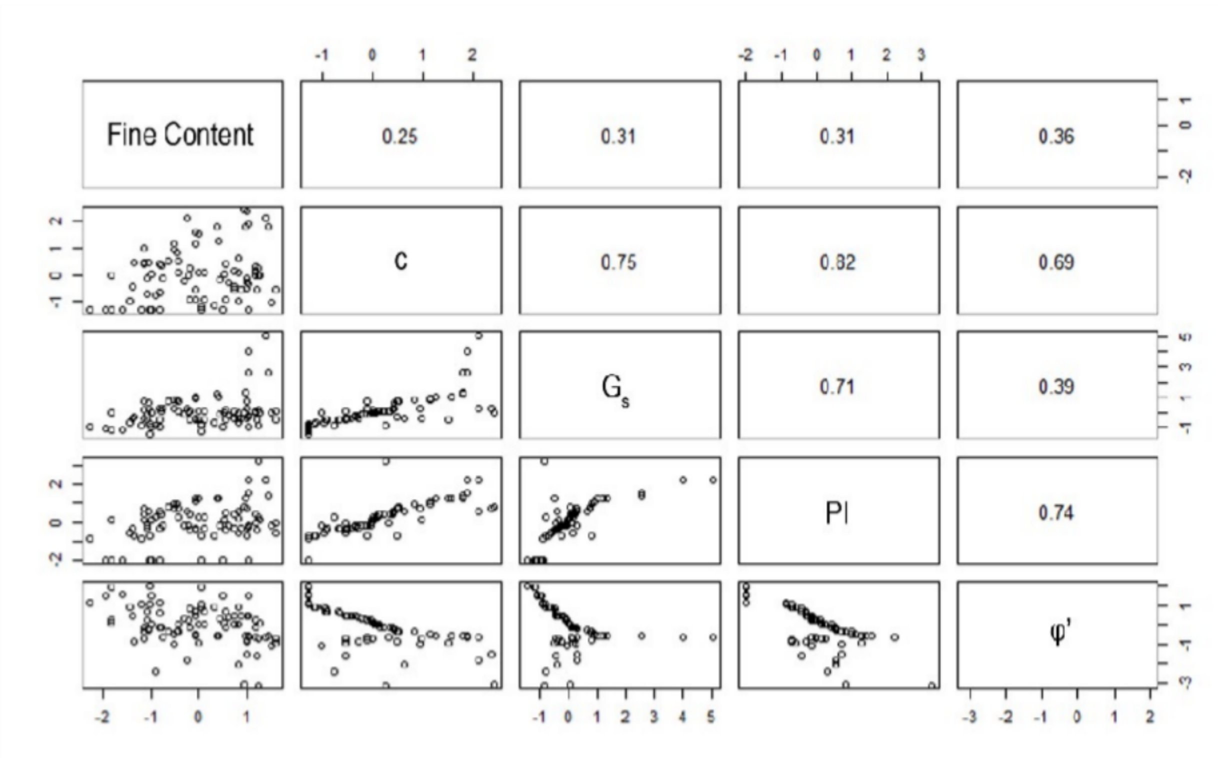

Figure 2 – Scatter plot matrix of the study variables.
Source: Author (2024).

Table 4 – Principal components loadings.

PC	Fine content	c'	Gs	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

Source: Author (2024).

Kaiser criterion (KAISER, 1970) and the scree-plot analysis (CATTELL, 1966) were used to determining the number of principal components that should be kept in the analysis (see Figure 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.

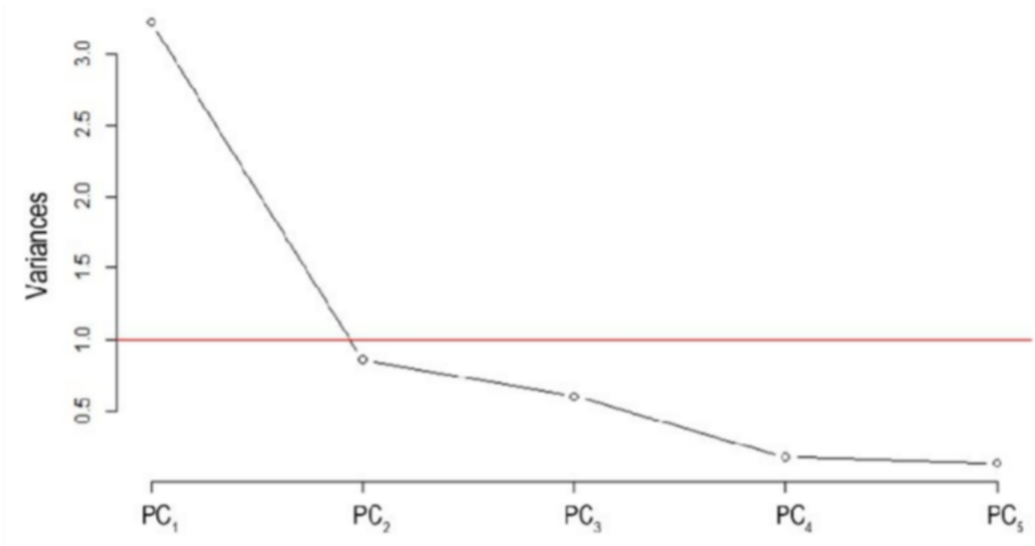

Figure 3 – Scree plot for determining the number of principal components.
Source: Author (2024).

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. Figure 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.

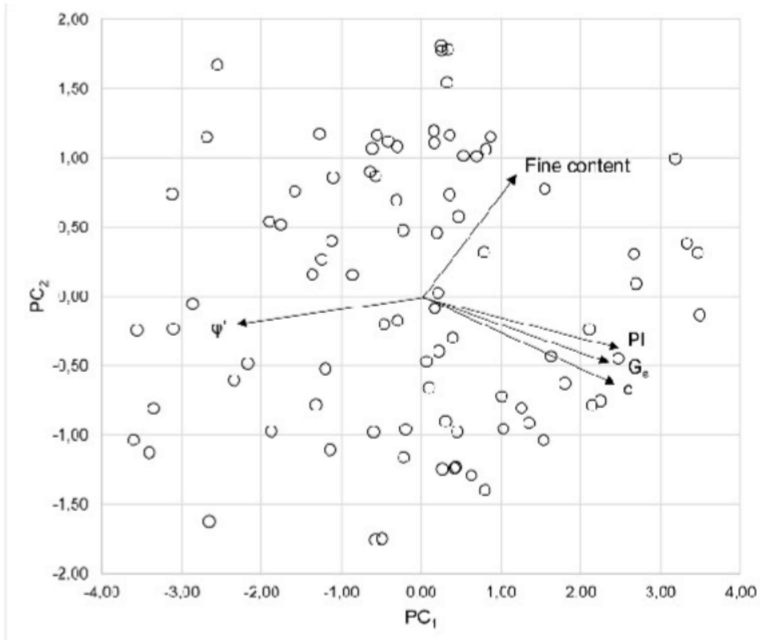

Figure 4 – Biplot graph.
Source: Author (2024).

The clusterization of the data was performed through k-means method (Figure 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 (Figure 6). The validation of the cluster analysis was based visual inspection of the Figure 5. The groups are well defined with a very small overlapping region.

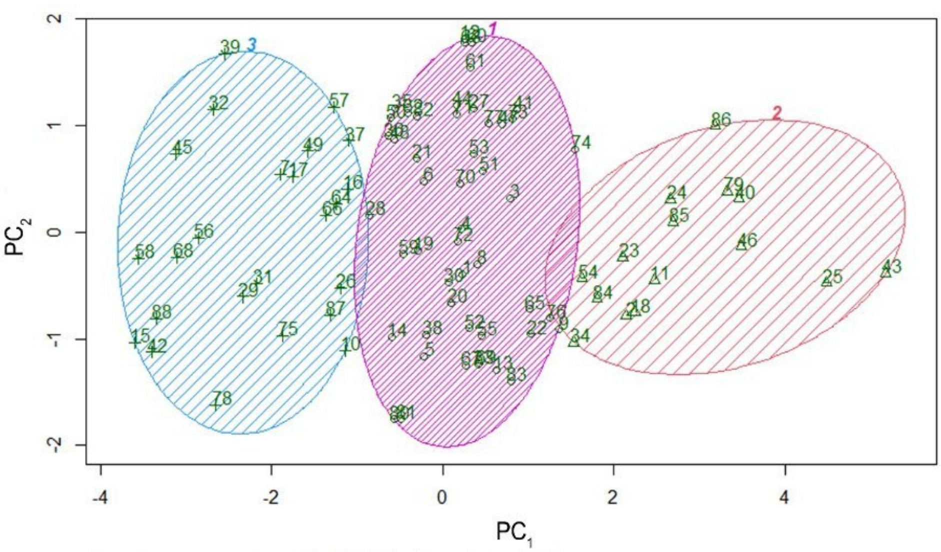

Figure 5 – Result of the cluster analysis.
Source: Author (2024)

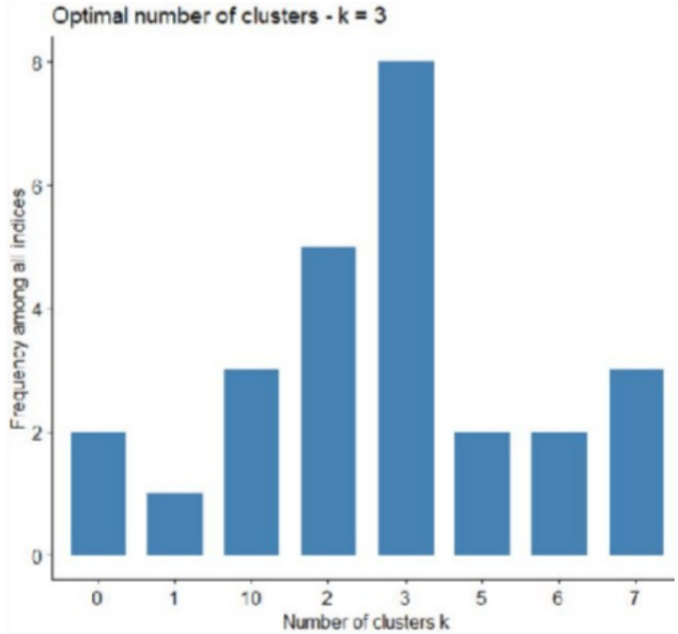

Figure 6 – Application of majority rule to determining the optimal number of clusters for the data.
Source: Author (2024).

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, 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 Figure 7.

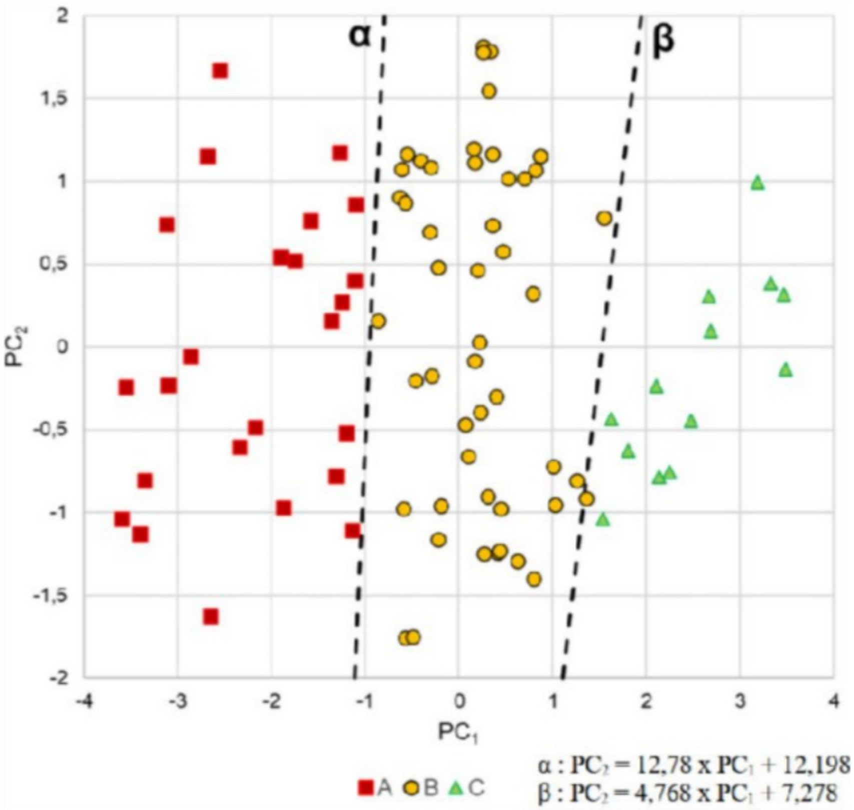

Figure 7 – Proposed soil classification system.
Source: Author (2024).

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

Class A includes soils with the lowest PI and Gs values. These soils have Gs values less than 2.82, indicating the predominance of minerals with lower unit weight, such as quartz (Gs = 2.65) and kaolinite (Gs = 2.60). According to Das and Sobhan (2013), these soils have low activity values and, consequently, low plasticity. Additionally, the predominance of low unit weight minerals in Class A indicates low iron content and, consequently, low cementation. This information is supported by Stefanou and Papazafeiriou (2013) and Cruz, Rodrigues and Foneca (2013), who discuss the positive in- fluence of cementation on soil cohesion. Furthermore, this type of soil has a fine content ranging from 26% to 85%, with 75% of samples having a fine content less than 67.25%. The fine content of these samples and the presence of minerals with Gs less than 2.82 indicate predominantly sandy characteristics of these materials. These characteristics lead to higher observed φ' values (DAS and SOBHAN, 2013). Therefore, it is possible to infer that Class A soils correspond to materials with a sandy to inactive fine matrix, with a concentration of low unit weight minerals and low cementation.

Class B includes soils with transitional behavior between Classes A and C. For these materials, intermediate Gs and PI values indicate the presence of a higher percentage of active clay minerals and greater density compared to the values observable for Class A. More pronounced cohesion is observed for these materials, indicative of cementation between grains. However, the occurrence of moderate friction angles corroborates the observation of granular fractions of no less than 16% in 75% of the observations. Thus, Class B soils can be understood as transitional materials between Classes A and C, with intermediate behaviors.

Class C includes materials that exhibit the opposite behavior to that of Class A materials for the five analyzed variables. It can be observed, therefore, that Class C tends to encompass predominantly fine materials with clayey and active characteristics, for which the design concerns often inherent to CH-type materials (high-plasticity clays), distributed among Classes A, B, and C, are applicable. The higher cohesion values observed for Class C materials may be associated with pre-consolidation and cementation phenomena (CRUZ, RODRIGUES and FONECA, 2013), with the latter derived from the presence of high-density minerals, such as hematite and goethite, justified by observing Gs values greater than 3.05 in 75% of the analyzed observations. However, the high c' values observed for Class C are contrasted by materials of low geotechnical competence, indicating the possibility of non-processing of the aforementioned phenomena.

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

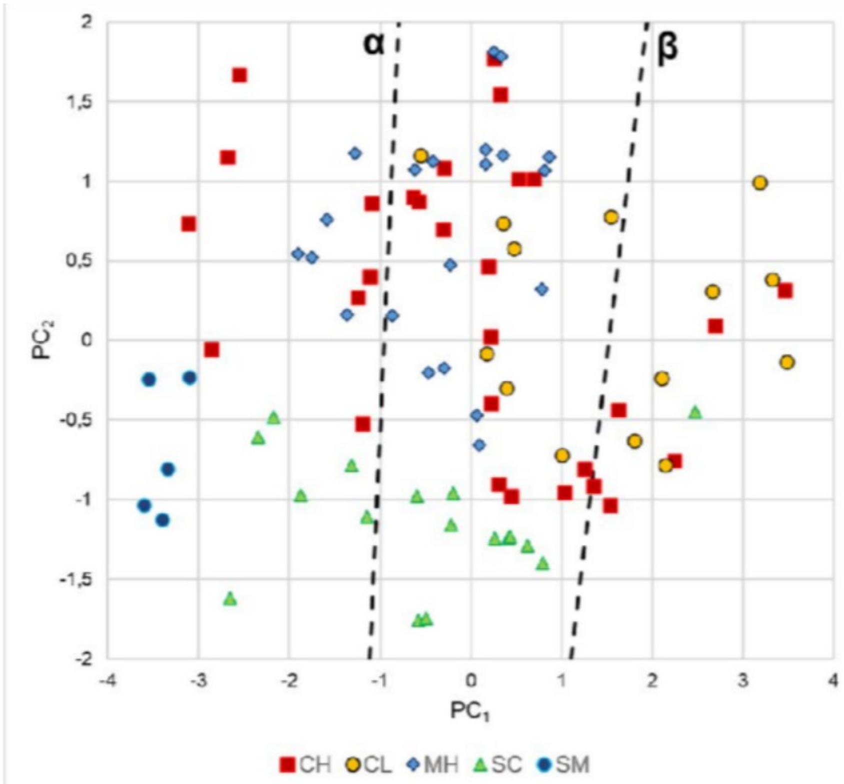

Figure 8 – USCS versus proposed soil classification system of the dataset samples.
Source: Author (2024).

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 (Figure 8). CH and CL soils were allocated in the region corresponding to class 3 of the proposed classification system. Therefore, the mentioned observations points 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', Gs, PI and fine content variables. Figure 9 to Figure 13 present the boxplots.

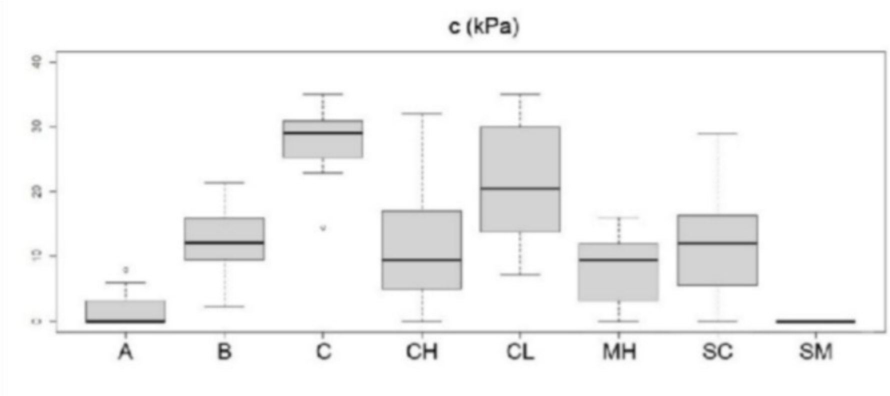

Figure 9 – Boxplots of c' values for the analyzed classes.
Source: Author (2024).

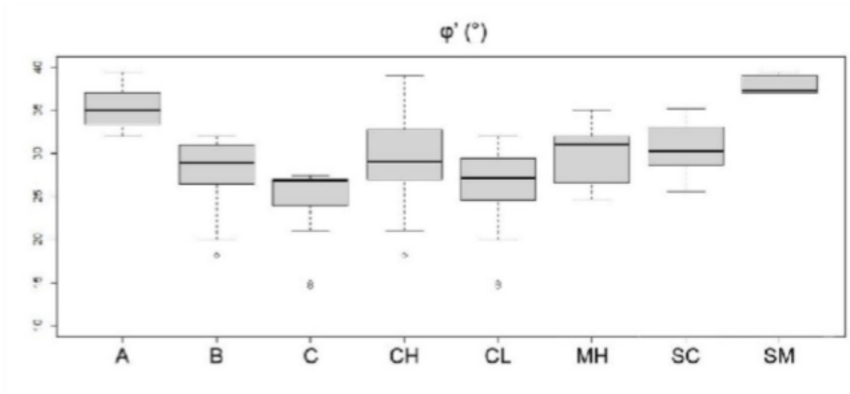

Figure 10 – Boxplots of φ' values for the analyzed classes.
Source: Author (2024).

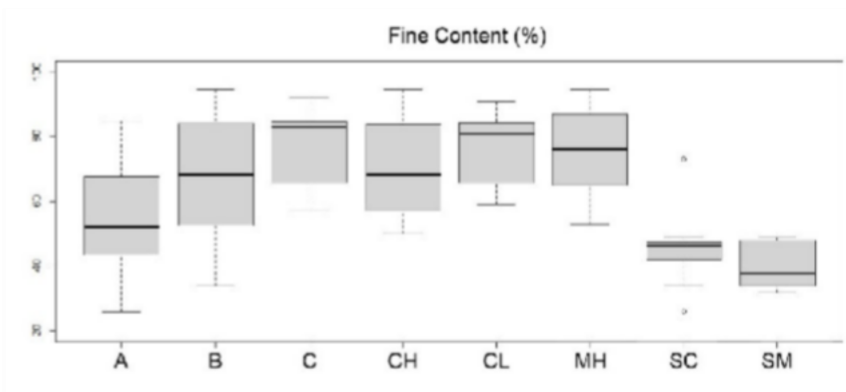

Figure 11 – Boxplots of fine content values for the analyzed classes.
Source: Author (2024).

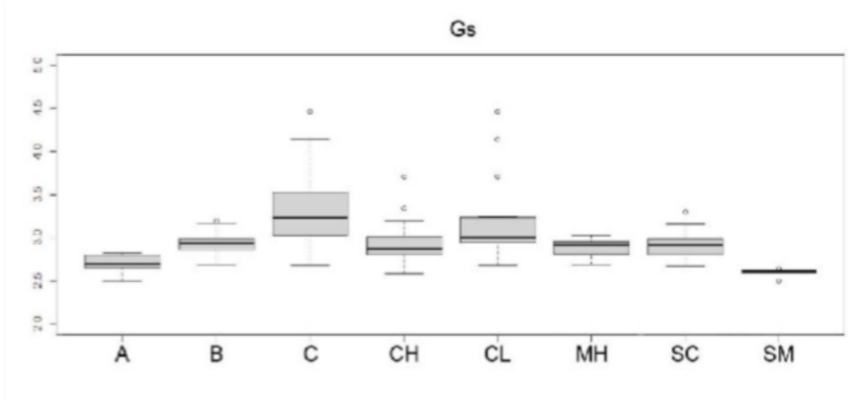

Figure 12 – Boxplots of Gs values for the analyzed classes.
Source: Author (2024).

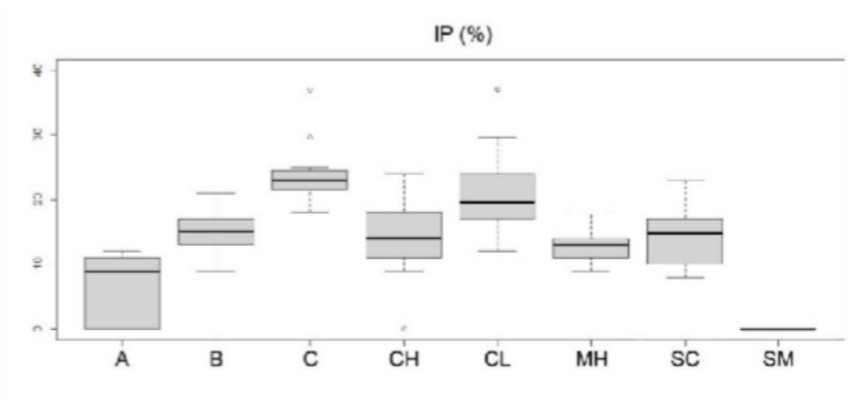

Figure 13 – Boxplots of PI values for the analyzed classes.
Source: Author (2024).

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. 9 to 13. 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 (Figure 9 to 13) are capable of demonstrate the capability of the proposed soil system in discrimi- nating the soils according to its strength behaviors.

4. Conclusions

This study proposes a novel soil classification system specifically tailored for the Quadrilátero Ferrífero province in Brazil. Multivariate statistical techniques, including principal component analysis (PCA) and k-means cluster analysis, were employed to analyze a comprehensive dataset of soil samples. PCA effectively reduced data dimensionality while identifying key relationships between soil properties. Subsequently, k-means clustering efficiently grouped soils based on their geotechnical behavior, defining distinct classes (A, B, and C). These classes are further characterized by establishing α and β limits, which directly translate to soil resistance variations.

The findings reveal limitations of the widely used USCS system in accurately classifying Quadrilátero Ferrífero soils (FORTES, MERIGHI and NETO, 2002). In contrast, the proposed system demonstrably offers superior efficiency in categorizing soils based on their geotechnical behavior, reflecting the unique soil variability within this mineral province.

The adoption of this novel classification system paves the way for the development of specific design guidelines tailored for classes A, B, and C. These guidelines hold significant promise for enhancing geotechnical engineering practices in the region. However, due to the study's regional scope and limited sample size, it is crucial to exercise caution when ap- plying it to individual engineering projects. Further research is recommended to explore correlations between these soil classes (A, B, and C) and other relevant geotechnical behaviors, ultimately leading to a more comprehensive and practical application of this classification system in geotechnical engineering.

Acknowledgments

The authors would like to thank the Federal University of Ouro Preto (UFOP) that, through UFOP grant, had funded this research.

References

AHMED, M.; SERAJ, R.; ISLAM, S. M. S. 2020. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. Electronics. 9(8):1295. DOI: https://doi.org/10.3390/electronics9081295

Associação Brasileira de Normas Técnicas (ABNT). NBR 6458:2016: *Soil - Determination of Specific Gravity of Grains*. Rio de Janeiro: ABNT, 2016a.

Associação Brasileira de Normas Técnicas (ABNT). NBR 6459:2016: *Soil - Determination of Liquid Limit*. Rio de Janeiro: ABNT, 2016b.

Associação Brasileira de Normas Técnicas (ABNT). NBR 7180:2016: *Soil - Determination of Plastic Limit*. Rio de Janeiro: ABNT, 2016c.

Associação Brasileira de Normas Técnicas (ABNT). NBR 7181:2016: *Soil - Particle Size Analysis*. Rio de Janeiro: ABNT, 2016d.

CARVALHO, L. O.; RIBEIRO, D. B. Application of kernel k-means and kernel x-means clustering to obtain soil classes from cone penetration test data. Soils and Rocks, 43(4), 607-618, 2020. Disponível em: https://doi.org/10.28927/SR.434607. Acesso em: 12/01/2024.

Cattell, R. B. The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, 1(2):245-276, 1966.

CHARRAD, M.; GHAZZALI, N.; BOITEAU, V.; NIKNAFS, A. NbClust: Determining the Best Number of Clusters in a Data Set. R Package, version 3.0, 2015. Disponível em: https://CRAN.Rproject. Acesso em: 15/01/2024.

Cruz, N.; Rodrigues, C.; Foneca, A. V. An Approach to Derive Strength Parametersnof Residual Soils from DMT Results. *Soils and Rocks*, 37 (3): 195–209, 2013.

Das, B. M; Sobhan, K. *Principles of geotechnical engineering*. 8ª. ed. Cengage Learning, Stamford, CT, 2013.

DINH, D. T.; FUJINAMI, T. & HUYNH, V. N. 2019. Estimating the Optimal Number of Clusters in Categorical Data Clustering by Silhouette Coefficient. In: Chen, J., Huynh, V., Nguyen, GN., Tang, X. (eds) Knowledge and Systems Sciences. KSS 2019. Communications in Computer and Information Science, vol 1103. Springer, Singapore. DOI: https://doi.org/10.1007/978-981-15-1209-4_1

Dowdy, S.; Weardon, S.; Chilko, D. Statistics for Research. 3ª ed. Wiley Interscience, Hoboken, NJ, 2004.

Fookes, P. G. A review: Genesis and classification of tropical residual soils engineers. *Geotechnics in the African Environ- ment.* 423-442, 1994.

Fortes, R. M.; Merighi, J. V.; Neto, A. Z. Método de Teste de Azulejo para Identificação Rápida de Solos Tropicais. Anais do 2º Congresso Português de Estradas, Lisboa, Portugal, 18–20 novembro de 2002. Centro Rodoviário Português, 2002.

Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL. *Multivariate Data Analysis*. 7ª. ed. Bookman, 2009. International Organization for Standardization (ISO). ISO 17892-9:2018: *Geotechnical investigation and testing - Laboratory testing of soil - Part 9: Triaxial compression tests for saturated cohesive soils.* Geneva: ISO, 2018.

Kaiser, H. F. A second generation little Jiffy. *Psychometrika*, 35(4):401–415, 1970. MAHALANOBIS, P. C. 1928. A statistical study oftbe Chinese hud, Man in India, 8, p.107-122.

MAIOLINO, A. L. *Resistência ao Cisalhamento de Solos Compactados: Uma Proposta de Tipificação*. Dissertação de Mestrado, Programa de pós-graduação em Engenharia. Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 1985.

Nogami, J. S.; Villibor, D. F. A New Soil Classification for Road Purposes. Proceedings of the Brazilian Symposium on Tropical Soils in Engineering. ABMS, 30–41, 1981.

Pinto, C. S. *Basic Soil Mechanics Course in 16 Lessons*. 3ª. ed. Oficina de Textos, São Paulo, SP, 2006.

Stefanou, S.; Papazafeiriou, A. Effects of spatial variability on soil liquefaction: Some design recommendations. *Eurasian Journal of Soil Science*, 2 (2):122–130, 2013.