Proceedings of the 17th Pan-American Conference on Soil Mechanics and Geotechnical Engineering (XVII PCSMGE), and 2nd Latin-American Regional Conference of the International Association for Engineering Geology and the Environment (IAEG), La Serena Chile, 2024.

Complementation of an iron ore tailing SBTn classification using k-means clustering technique

Análise de classificação SBTn aplicada a rejeitos de minério de ferro usando técnica de agrupamento k-means

Pablo Dos Santos Cardoso Coelho
Faculty of Engineering – Department. of Civil and Environmental Engineering, University of Alberta, Canada e-mail:. pdossan1@ualberta.ca

Lucas Deleon Ferreira
Department of Urban Engineering, Federal University of Ouro Preto, Minas Gerais, Brazil. e-mail: lucas@ufop.edu.br

ABSTRACT: The present paper employs multivariate statistical methods to aid the classification of a microstructured partially saturated iron ore tailing through the SBTn classification system, providing means to evaluate the occurrence of groups of readings with similar geotechnical behavior, considering the evaluation of normalized tip resistances, Q_{tn}, normalized friction ratios, F_r, and normalized pore pressure differential, U_2. The dimensionality of variables studied in this research was conducted through the application of principal component analysis, which found that the maintenance of two principal components explained 88.79% of the observed data variability. Usage of the k-means algorithm adopting the Euclidean distance has shown that the studied tailing could be separated, in terms of the defined principal components, into three different groups. The defined groups were compared with the classification charts proposed by the SBTn classification system, after which it was found that the F_r-Q_{tn} chart was more accurate in explaining the behavioral variations of the studied tailing than the U_2-Q_{tn} chart.

KEYWORDS: multivariate statistics; clustering methods; CPTU; geomaterial classification; iron ore tailings.

1 INTRODUCTION.

Mining tailings hydraulically disposed in mining dams are materials classically treated as predominantly soft, contractive, and saturated. However, it is observed that unconventional scenarios, in which there is partial saturation, the presence of iron oxide levels, among other specific occurrences, are responsible for inducing unconventional geotechnical behaviors, as stated by Lo Presti et al. (2016) and Schnaid (2020). The suctions and/or cementations induced from these situations give these materials characteristics in situ indicative of greater strengths and stiffness than those observed for such materials in conditions of total saturation and non-cementation – mainly at small strains, in the case of cementations (Robertson, 2016) – masking potential brittle or low resistance behaviors of these materials.

Robertson (2016) presents the normalized Soil Behavior Type (SBTn) classification system, widely applied in geotechnical engineering practice as an important tool for the classification of geomaterials based on CPTU tests. The system is based on the combined evaluation of normalized tip resistances and normalized friction ratios.

For potentially cemented materials, Robertson (2010) proposes the use of the behavioral classification chart presented by Schneider et al. (2012) and modified by Robertson (2016), which considers the behavior of pore pressure developments along the sampler penetration, which holds strong explanatory potential about the nature of the tested material, according to Schneider et al. (2012). However, there is a restriction on the applicability of this chart due to the saturation of the analyzed material. Lo Presti et al. (2016) also point out that the analysis of partially saturated soils is unconventional, with methodologies of analysis not widespread and of a restrictive character.

In the described scenario, it is possible to perceive the fragility of non-conventional material evaluations through the SBTn system since the effects promoted by partial saturation tend to produce pore pressure values along the CPTU cracks that are not very representative of the perceptible behaviors in saturated materials explained by Schneider et al. (2012), able to elucidate the behavior of the studied materials despite the occurrence of cementations.

Thus, it is proposed in this article to disseminate in geotechnical engineering practice the use of tools that can produce complementary information that can solidly help define and evaluate the SBTn classes of tested materials – from which liquefaction potential analyses are conducted – the use of principal multivariate statistics, pointed out by Carvalho and Ribeiro (2020) as an efficient tool to describe field behaviors. In the present work, the usage of principal component analysis (PCA) and multivariate clustering techniques using the non-hierarchical k-means method are employed on CPTU test readings.

1.1 *Principal component analysis (PCA)*

Formally presented by Hotelling (1933), principal component analysis (PCA) is an interdependence technique used to conduct multivariate statistical analysis that allows the identification of the database structure and the evaluation of its dimensionality reduction through the adoption of statistical variables dependent on the database's original variables (Hair et al., 2009). These statistical variables are called principal components (PC), consisting of linear combinations of the variables in the database.

It is assumed that the adopted PC variables hold the nature and explanatory character of the original variables (Hair et al., 2009). It should also be observed that the use of the PC variables requires a conceptual basis capable of explaining the meaning of the combinations represented by them.

The perception of intercorrelation between the data is also a premise for the use of PCA (Mingoti, 2013), with databases with minimum correlation values of 0.30 between the variables being desirable – and values above 0.70 being indicators of high correlations (Hair et al., 2009). Bartlett's sphericity test is presented as an important tool for the verification of data intercorrelation, consisting of a hypothesis test capable of predicting the occurrence of significant correlations in the correlation matrix of the analyzed data.

The obtainable PC values correspond to the products between eigenvalues of the covariance matrix or correlations - in the case of standardized data - and the random vectors, Z_i, as shown in Equation 1.

$$PC_i = e_i^t Z \tag{1}$$

The proportion of data variability explained by a principal component is given by Equation 2, in which the values of λ_i are equivalent to the variances of the principal components PC_i and equal, therefore, to the eigenvalues of the covariance/correlations matrix of the variables analyzed:

$$Var(PC_i) \big/ Var_{total} = \lambda_i \big/ \sum \lambda_i \tag{2}$$

Hair et al. (2009) state that there are no definitive criteria for the precise selection of the number of PC variables to be maintained in a multivariate statistical analysis, although existing empirical criteria tend to produce good estimates. Scree (Cattell, 1966) and Kaiser (1970) criteria are widely used to select the number of retainable principal components. The scree criterion suggests, through graphical analysis, the adoption of a number of principal components from which the adoption of the subsequent components produces insignificant increases in variance in the sum of variances. Such a number is perceived from a shape similar to an elbow in a PC_i-λ_i graph.

The Kaiser criterion assumes that the maintenance of the principal components should be associated with the eigenvalues of the correlation matrix greater than 1. Hair et al. (2009) point out, however, that for databases with less than 20 original variables, this criterion tends to suggest excessive extractions of PC variables.

Multivariate statistical techniques were used to treat readings from CPTU tests conducted on iron ore tailings, which were hydraulically disposed in a mining dam located in Minas Gerais – Brazil, in the region of Quadrilátero Ferrífero.

In addition to the resistance behaviors conventionally addressed by the SBTn system, the generations of normalized pore pressure differentials in the tested tailings were studied using the statistical methods employed, assuming that the pore pressure generations can retain relevant information about the geomaterials (Lunne et al., 1997).

1.2 *K-means clustering algorithm*

Cluster analysis can be understood as statistical techniques used to group elements by similarity, whose definition is based on a distance function established for each problem. Sakr and Gaber (2014) present the k-means clustering method as an iterative non-hierarchical clustering method formally presented for the first time by MacQueen (1967). The method aims at clustering n variables into k groups, k being defined preliminarily to conducting the clustering.

The k-means clustering algorithm takes as a premise the definition, randomly or not, of the initial centroid of the k groups stipulated before the analysis. After defining the initial centroids, the sample elements of the analyzed database have their distances from the centroid assessed one by one so that each element is inserted in the group with a less distant centroid. From the insertion of new elements in the groups, the centroids are recalculated, and the analysis proceeds iteratively until the convergence of the groups.

The function conventionally used to calculate the distance in cluster analysis by the k-means algorithm, commonly resulting in good results, is the Euclidean distance, as exposed by Bortolossi (2002). This distance can be understood as the geometric distance between two points, defined in a space of any dimensions, for the points p, of coordinates (p_1, p_2, ..., p_n), and x, of coordinates (x_1, x_2, ..., x_n) through Equation 3.

$$d(x,p) = \sqrt{(x_1 - p_1)^2 + \dots + (x_n - p_n)^2} \tag{3}$$

Since the adoption of larger numbers of groups results in shorter average distances between clusters and, therefore, less explanatory potential for differences between those clusters, Thorndike (1953) presents the elbow method as a fundamental tool to define the number of k groups to be adopted in the application of the k-means clustering method.

The elbow method, similar to the scree method, consists of a graphical interpretation of Cartesian space in which values of k are plotted as a function of the average within-cluster distances provided by k. It is observed graphically, from a sudden change in the slope of the curve, that after a certain value of k, little decrease in the value of within-cluster average distance is achieved with the increase of k. Thorndike (1953) defines this k value as the ideal number of groups to be adopted in a cluster analysis by the k-means method.

1.3 *SBTn classification system*

The classification system called Normalized Soil Behavior Type (SBTn), presented by Robertson (2016), is an update of the works previously conducted by Robertson (2010).

The classification system takes as a premise concepts from the theory of critical state of soils to promote the typification of geomaterials and the characterization of their susceptibility to liquefaction through the analysis of CPTU test results (Robertson, 2016).

The methodology to characterize geomaterials from the SBTn system is based on the use of the SBTn classification chart proposed by Robertson (2016). This chart consists of a Cartesian space in which the normalized friction ratio, F_r, is combined with the normalized tip resistance, Q_{tn}, of a determined sample point. Its position in relation to the proposed classification regions is verified based on historical data.

Robertson (2016) points to the relevance of conducting the microstructure analysis of the studied materials through the verification of the K^*_G factor, defined by Equation 4, since microstructured materials tend to present higher Q_{tn} values at small strains, i.e., during the penetration of the CPTU sampler tip, which is not maintained through large strains due to the breakdown of the microstructures. Such behavior tends, therefore, to mask the real nature of the analyzed material, causing the positioning of sample points in inadequate regions and sometimes characterizing contractive materials as dilative - positioned above the CD = 70 curve.

$$K_G^* = \left({G_0}/{q_n}\right) \times Q_{tn}^{0,75} \tag{4}$$

According to Robertson (2016), microstructured materials tend to have K^*_G values greater than 330. The Q_{tn}-F_r chart does not consider the classification of the studied geomaterials considering the behavior of pore pressure generation, being unable to explain large strain behaviors for porous elements located behind the conical tip, and to demonstrate the permeability of those materials and, therefore, the type of response to external loads (Lunne et al., 1997). Robertson (2016) points to the relevance of conducting the study of pore pressures along the penetration through the Q_{tn}-U_2 classification chart, proposed by Schneider et al. (2012) and modified by Robertson (2016).

The aforementioned chart is used as an additional tool to assess the behavior of the studied materials, since phenomena such as microstructuring tend to mask behavior at small strains that become evident at large strains. Carvalho and Ribeiro (2019) point out that systems based on the employment of two charts tend to be inefficient for offshore soundings.

The evaluation of the pore pressure differential, U_2, defined by Equation 5, based on the described methodology, has as the main restriction, as indicated by Robertson (2016), the need for accurate u_2 readings by maintaining the saturation of the porous element along the CPTU penetration.

$$U_2 = {(u_2 - u_0)}/{\sigma'_{v0}} \tag{5}$$

Given the absence of the phreatic level in the studied tailings deposit, despite an observable significant residual degree of saturation in the fine materials, it is assumed that the chart proposed by Schneider et al. (2012) is not applicable. However, it is assumed that the behavior of generating pore pressure along the penetrations may contain relevant information about the nature of the materials.

2 MATERIALS AND METHODS

2.1 *Database construction*

The database used to conduct the studies in this work consists of a set of 9,820 readings obtained over six CPTU tests performed on partially saturated iron ore tailings disposed hydraulically in a mining dam currently in the process of de-characterization.

These tests were performed in accordance with the standards and limits determined by the standard 22476-1 (ISO, 2012), with the propagation speeds of type S waves (v_s) being measured in three boreholes: every meter, until the depth of 18.00 m in one of the boreholes, with two additional tests at depths of 20.00 m and 21.60 m. In the other two boreholes, the tests were carried out, similarly, meter by meter, up to a depth of 22.00 m in the first and between depths of 9.00 m and 14.00 m in the second one.

The multivariate statistical analysis of the data was conducted using the normalized values of tip resistance, friction ratio, and pore pressure differential as variables, which are also used as classificatory variables in the SBTn instrumental charts presented by Robertson (2016).

A multivariate outlier analysis was conducted for the database presented, observing 162 outliers characterized as reading errors or observations that are not representative of the actual condition of the tailings disposed of in the dam. Therefore, for the present study, a sample of 9,658 sample elements remaining from the removal of the described outliers was adopted. Table 1 presents a statistical summary of the data studied.

Table 1. Statistical summary of the data studied.

	Fr (%)	Qtn	U2
Minimum	-4.250	-1.060	-0.740
1º Quartile	0.690	4.692	0.010
Median	1.300	14.030	0.100
average	1.566	14.288	0.679
3º Quartile	2.000	20.950	1.618
Maximum	11.400	108.340	3.210

The database used is illustrated in Table 2, in which ten sample elements belonging to the analyzed sample are presented. Given the disparity between the orders of magnitude analyzed, it was proposed to standardize the data, as recommended by Hair et al. (2009). After data standardization had been processed, it was shown in Table 2.

Table 2. Partial presentation of the data studied.

Data after standardization				Data Processed		
ID	F_r (%)	Q_{tn}	U_2	F_r (%)	Q_{tn}	U_2
1	2.69	16	18	-1.31	-1.44	-0.76
2	3.24	23	27	-1.42	-1.44	-0.76
3	3.00	18	29	-1.67	-1.43	-0.76
4	2.96	16	30	-2.30	-1.42	-0.76
5	2.86	9	26	-4.46	-1.40	-0.76
6	2.91	13	31	4.77	-1.26	-0.75
7	2.69	9	35	1.87	-1.07	-0.75
8	2.97	17	29	1.25	-0.87	-0.73
9	3.19	21	28	1.01	-0.68	-0.73
10	2.82	12	32	0.99	-0.55	-0.73

2.2 *Methodology*

The methodology adopted in the present research consisted of the conduction of tactile-visual characterization of the materials from deformed samples from SPT tests; database construction; classification of materials using the SBTn classification system; conduction of multivariate statistical analysis of PCA clustering using k-means (using the elbow method, developed by Thorndike (1953), to define the number of groups used and the Euclidean distance as the similarity measure); evaluation of the occurrence of microstructures through the verification of K^*_G; and analysis of the generated clusters, defining the relevance of considering pore pressure generation behaviors for the analyzed tailings and discussing the information provided by the multivariate statistical methods.

2.3 *Analysis and results*

The tailings studied in the present work were classified by tactile and visual means through the evaluation of testimonies from ten SPT surveys, with occurrences of predominantly silty materials interspersed with sandy matrix materials with a fraction of fines. The observed material characteristics are illustrated in Figure 1.

Figure 1. Visual characteristics of the tested material.

The sample elements analyzed had their Q_{tn} and F_r readings obtained from the CPTU tests performed plotted on the SBTn system chart shown in Figure 2, for which a quantitative summary of the obtained classifications is presented in Table 3.

It is clear from the analysis of Q_{tn} and F_r that there is a predominance of contractive behaviors, with 4.63% of the analyzed materials suggesting a dilative nature. The classificatory chart is not, however, able to explain the behavior of the analyzed materials in terms of the generation of pore pressures. The classifications obtained through Figure 2 can be compared with clusters that consider the behavior of the additional variable U_2, pointed out by Schneider et al. (2012) as efficiently explaining the behavior of geomaterials tested via CPTU.

Thus, it was proposed to evaluate the effects of the conjugation of U_2 with Q_{tn} and F_r on the generation of clusters using the k-means method, whose analyses were subsequently compared with the classifications obtained on the Q_{tn}-F_r chart. Prior to conducting the clustering analysis, it was verified through PCA the possibility of reducing the dimensionality of the database.

For the application of multivariate statistical techniques, the correlations between the variables Q_{tn}, U_2, and F_r were analyzed, with a p-value of less than 0.05 being observed for Bartlett's sphericity test, rejecting the hypothesis of the absence of significant correlations between the variables and validating the application of the techniques treated in this article.

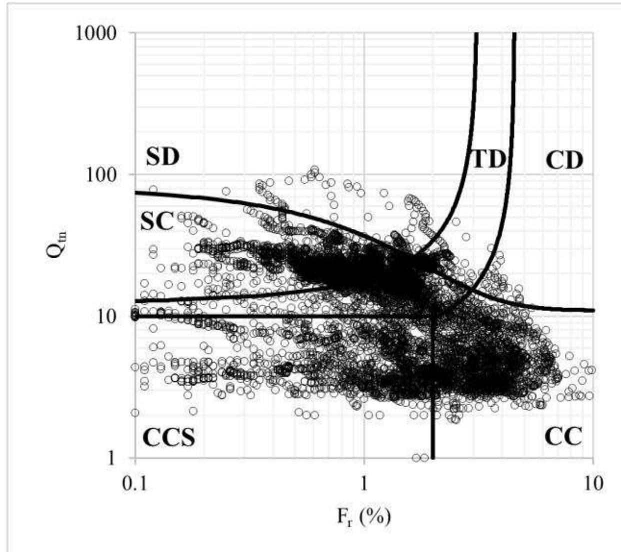

Figure 2. Test readings plotted in the SBTn Q_{tn}-Fr chart.

Table 3. Quantitative summary of the obtained classifications.

CCS	CS	CD	TC	TD	SC	SD
18.99%	23.70%	0.38%	23.61%	2.41%	29.07%	1.84%

In this way, the principal component analysis was performed for the standardized database, obtaining the loadings of the variables presented in Table 4 for each of the PC variables.
The application of the scree and Kaiser (1970) methods indicated, respectively, the maintenance of two and one principal components. Therefore, the conceptual evaluation of the loadings presented in Table 5 was taken as a criterion for retaining the components.

Table 4. Principal components loadings.

PC	F_r (%)	Q_{tn}	U_2
1	0.441	-0.673	0.594
2	0.844	0	-0.529
3	0.305	0.735	0.606

The analysis of the loadings of PC_1 indicates its positive correlation with F_r and U_2 values and negative with Q_{tn} values, thus portraying, for positive values, sample elements with low Q_{tn} values and high F_r and U_2 readings—and opposite behavior for negative PC_1 values. PC_2 has zero correlation with Q_{tn} values, indicating, for positive values, higher F_r values and lower U_2 values. Given the notes made by Lo Presti et al. (2016) of increased resistance observed in CPTU tests as a function of the degree of saturation, it is assumed that the maintenance of PC_2 is relevant to the present study, with potential to explain the degree of saturation of the sampled materials. It was decided not to maintain the PC_3 due to the low variance attributed to it, indicated by the referred methods of assessing the number of principal components, and its absence of relevant geotechnical significance. The adopted variables are able to explain 88.79% of the data variability.

Thus, the cluster analysis was performed using the k-means method for the sample elements studied in terms of PC_1 and PC_2, obtaining the clusters illustrated in Figure 3.

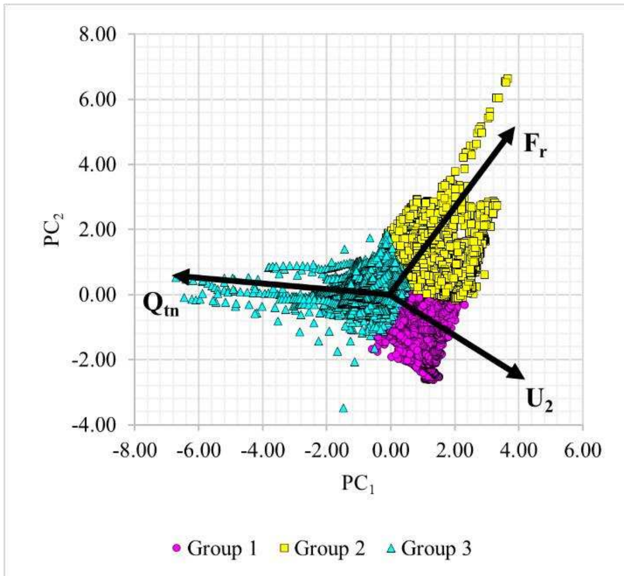

Figure 3. Clusters obtained through the employment of the k-means method.

From the analysis of the resulting clusters, it is observed that the clusters called groups 1, 2, and 3 have the following characteristics:

- In group 1, for which there are predominantly positive PC_1 values associated with predominantly negative PC_2 values, it is perceived the occurrence of materials with significantly low tip resistance, characteristics of clay-like materials, associated with significant generations of pore pressure (Lunne et al., 1997). The occurrence of negative PC_2 values is indicative of the occurrence of less expressive readings of sleeve friction, indicative of possible sensitivity (Robertson and Cabal, 2015).
- In group 2, there are predominantly positive PC_1 values, indicative, in a manner similar to that stated for group 1, of clayey behaviors. However, due to the positive values of PC_2, it is noticed the occurrence of smaller generations of pore pressure than those observed in group 1, accompanied by the most expressive values of F_r among the three groups. The observed behavior corroborates the hypothesis of sensitivity of the clayey behavior materials of group 1, since higher values of sleeve friction for group 2 are expected.
- In group 3, due to the negative values of PC_1 and the lower magnitudes of PC_2—negative and positive—there is an indication of more expressive tip resistances accompanied by low values of F_r and U_2. Such behavior can, according to Lunne et al. (1997), be associated with those expected for materials of a nature transitional to sandy.

The results obtained from the clustering analysis carried out using the k-means method were then compared with the classification of the sampling elements in the Q_{tn}-F_r chart shown in Figure 4. Table 5 presents the comparative summary between the obtained clusters and the classifications resulting from the use of the Q_{tn}-F_r chart.

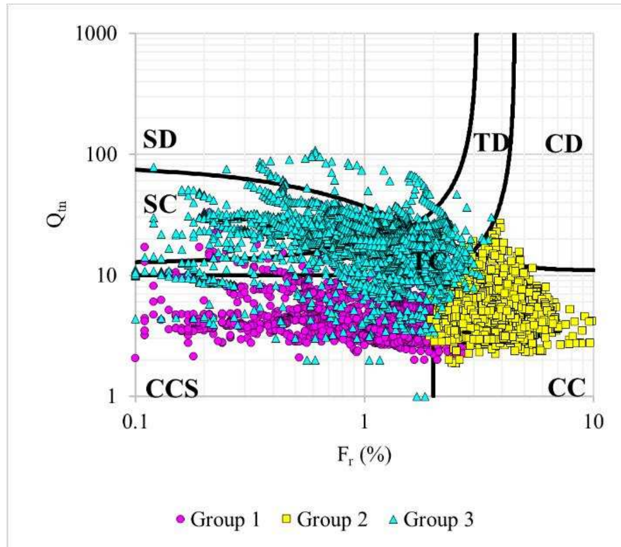

Figure 4. Comparison of the obtained clusters and the SBTn charts.

Table 5. Comparative summary for the obtained clusters and the classifications from the Q_{tn}-F_r chart.

	CCS	CS	CD	TC	TD	SC	SD
1	78.79%	43.10%	0.38%	0.92%	1.05%	2.70%	0.00%
2	0.21%	50.64%	91.51%	0.00%	0.40%	0.00%	0.00%
3	21.00%	6.26%	8.11%	99.08%	98.55%	97.30%	100.00%

The comparative analysis between the clusters shows that the materials with sandy to transitional characteristics are concentrated in group 3, for which the behavior of low values of PC_1 is common. This behavior reflects the predominance of sandy characteristics in these materials (Lunne et al., 1997), i.e., high values of Q_{tn} and low values of F_r. It is observed, however, that a significant part of the materials classified in the Q_{tn}-F_r chart as sensitive clays are classified as components of group 3. This set of sample elements, indicated in Figure 5, is associated with low pore pressure generations and low F_r values, due to low PC_2 values. It is observed that this behavior, associated with low values of tip resistance, is characteristic of clays with expressive values of sensitivity, coherently to the classification of sensitive clays attributed by the SBTn chart.

The other materials classified as sensitive clays, included in group 1, are consistent with the expected behavior for these materials due to the negative values of PC_2 associated with positive values of PC_1, previously described. It is noticeable, also from the evaluation of Table 6, that the dilative clays are mostly represented by group 2, to which characteristically clayey behaviors associated with low generations of pore pressures are attributed. Lo Presti et al. (2016) points out that the occurrence of materials classified as dilative clays in partially saturated deposits, in extreme cases, can occur through the occurrence of suction—the classification not being representative of the nature of the analyzed material.

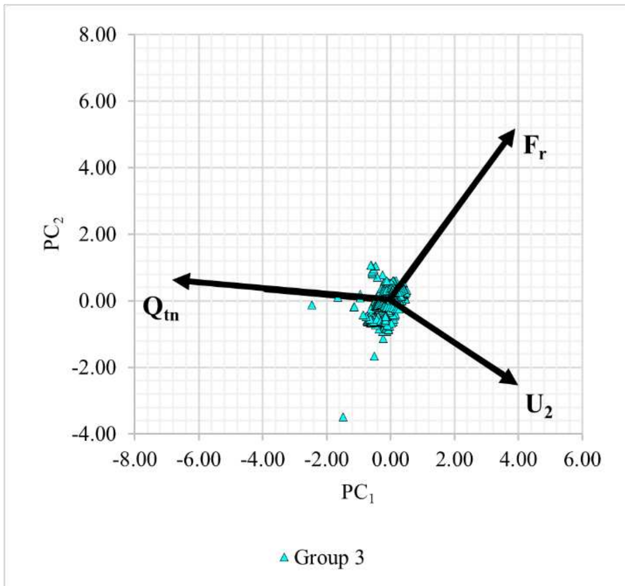

Figure 5. Data from group 3 classified as sensitive clays.

The materials classified as contractive clays, however, are dispersed between groups 1 and 2. This behavior, shown in Figure 6, is attributed to the less expressive generations of pore pressure of some of the materials in group 1, to which can be assigned the occurrence of lower residual saturation, emphasizing the statements of Lo Presti et al. (2016) about partial saturation and possible increments of effective tensions. This assumes the possibility that, to a greater degree of saturation, materials classified as contractive clays will behave as sensitive clays.

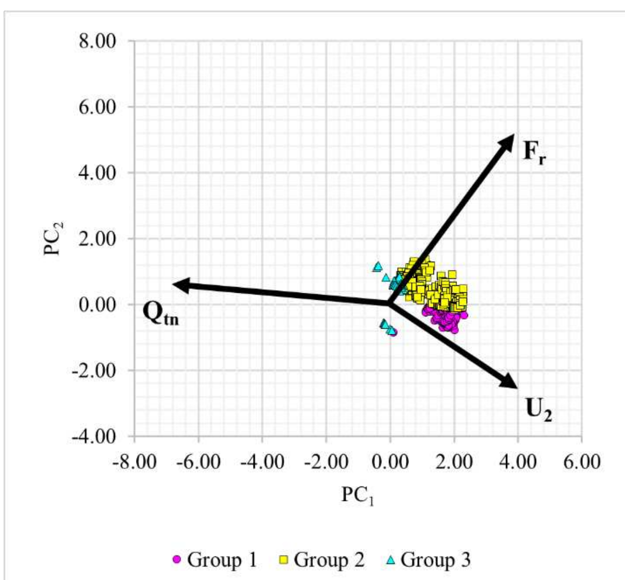

Figure 6. Materials classified as contractive clays.

The microstructure analysis conducted for the sample elements tested via vs measurement, shown in Figure 7, shows a series of complementary information to those previously discussed for groups 1, 2, and 3.

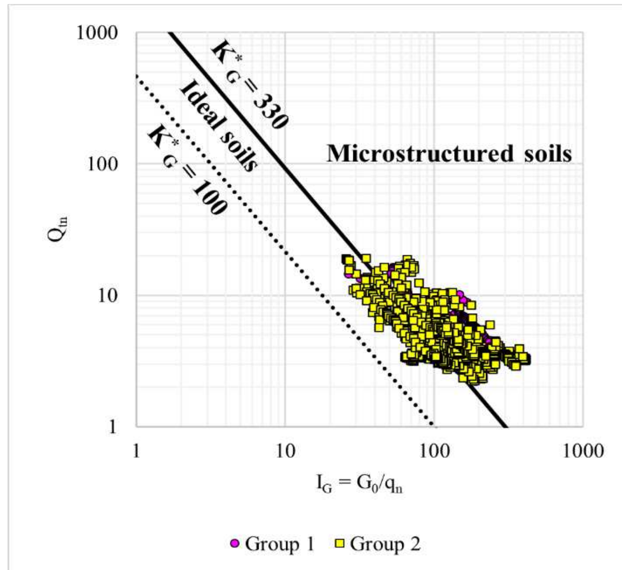

Figure 7. Evaluation of K^*_G for the studied tailing.

Among the K^*_G values obtained, it can be seen, as shown in Table 6, that groups 2 and 3 presents, respectively, 25.73% and 42.36% of their elements with signs of microstructure. It is also noted that the materials from group 3, with predominantly sandy to transitional characteristics, do not show signs of microstructure, according to the limits indicated by Robertson (2016). Thus, the possibility of increases in strength arising from cementation in this group is discarded.

Table 6. Descriptive summary of microstructure materials in relation to the total number of K^*_G measurements.

	CCS	CC	CD	TC	TD	SC	SD	Total
1	22.72 %	2.33 %	0.00%	0.15 %	0.08 %	0.45 %	0.00 %	25.73 %
2	0.00%	4.89 %	37.32 %	0.00 %	0.15 %	0.00 %	0.00 %	42.36 %
3	0.00%	0.00 %	0.00%	0.00 %	0.00 %	0.00 %	0.00 %	0.00%

The analysis of the quantitative presented for group 1 indicates materials with fine behavior are mainly responsible for their occurrences of microstructure, while in group 2, materials with dilative clayey behavior are responsible for the majority of the indications of microstructure. The latter indicates that the sample elements classified as dilative clays possibly show higher Q_{tn} values at small strains than it is observed at large strains—being also possible that these materials may present behavioral characteristics of contractive clayey materials at large strains.

Given the above, the analysis of clusters has led to the indications that:

- Materials classified as sensitive clays are predominantly accompanied by significant generations of pore pressures, although some occurrences of low U2 values are observed, which are associated with localized low residual degree of saturation.

- Materials classified as contractive to dilative clays are predominantly associated with low generations of pore pressure. Such behavior shows, among materials with characteristically fine behavior, the association of low residual degree of saturation with higher values of sleeve friction, if the behavior of contractive clays is compared with that of sensitive clays.
- Materials with transitional to sandy characteristics are mostly associated with the absence of residual saturation levels and, therefore, would probably have a drained response to induced loads.

3 CONCLUSIONS

The cluster analysis conducted by the k-means method efficiently complements explanations about the behavior of partially saturated materials classified by the SBTn system, introducing the analysis of U_2 behaviors into the classification context, which is significant for understanding how materials respond to external loads.

It is noted that regions close to the origin of the PC_2-PC_1 system generate confusion between materials with sandy and sensitive characteristics, as both exhibit low F_r/Q_{tn} ratios. Thus, attention is needed in interpreting the interfaces between these groups.

Additionally, it is pointed out that the transition between materials exhibiting sensitive clay and contractive clay behaviors is not well explained by the proposed clustering. Similar values of U_2 are observed for both materials in regions close to the origins of the PC_2-PC_1 system.

4 ACKNOWLEDGEMENTS

We wish to thank all those who contributed to the execution of this work, particularly CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and UFOP (Universidade Federal de Ouro Preto) for the support.

5 REFERENCES

Bortolossi, H. J. (2002). Cálculo Diferencial a Várias Variáveis. Edições Loyola. (in Portuguese).

Carvalho, L.O., & Ribeiro, D.B. (2019). Soil Classification System from Cone Penetration Test Data Applying Distance-Based Machine Learning Algorithms. Soils and Rocks, 42(2), 167-178. https://www.doi.org/10.28927/SR.422167

Carvalho, L.O., & Ribeiro, D.B. (2020). 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. https://doi.org/10.28927/SR.434607

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

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2009). Análise Multivariada de Dados. Bookman. (in Portuguese)

Hottelling,H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417-441. https://doi.org/10.1037/h0071325

ISO 22476-1. (2012) Geotechnical investigation and testing – Field testing – Part 1: Electrical cone and piezocone penetration test. ISO – International Organization for Standardization. Genebra, CH.

Kaiser, H. F. (1970). A second generation little Jiffy. Psychometrika, 35(4), 401-415.

D. Lo Presti, I. Giusti, B. Cosanti, N. Squeglia and E. Pagani, "Interpretation of CPTu in "unusual" soils", 2016, Rivista Italiana di Geotecnica, 4, 25-44.

Lunne, T., Robertson, P. K., & Powell, J. J. M. (1997). Cone Penetration Testing in Geotechnical Practice. Blackie Academic/Rouledge Publishing.

MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proc. 5th Berkeley Symposium on Mathematical Statistics and Probability. Vol.1, University of California Press, 281-297.

Mingoti, S. A. (2013). Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada. Editora UFMG. (in Portuguese)

Robertson, P. K. (2010). Evaluation of Flow Liquefaction and Liquefied Strenght. Journal of Geothecnical and Geoenvironmental Engineering, 136(6), 842-853. https://doi.org/10.1061/(ASCE)GT.1943-5606.0000286

Robertson, P. K. (2016). Cone penetration test (CPT)-based soil behaviour type (SBT) classification system – an update. Canadian Geotechnical Journal, 53(12), 1910-1927. https://doi.org/10.1139/cgj-2016-0044

Sark, S., & Gaben, M. (2014). Large Scale and Big Data: Processing and Management. CRC Press.

Schnaid, F. (2020). The Ninth James K. Mitchell Lecture: On The Geomechanics and Geocharacterization of Tailings. In 6th International Conference on Geotechnical and Geophysical Site Characterization. ISSMGE. http://isc6.org/index.php/program/keynote-lectures

Schneider, J.A., Hotstream, J.N., Mayne, P.W., & Randolph, M.F. (2012). Comparing CPTU Q-F and Q-Δu2/σvo' soil classification charts. Géotechnique Letters, 2(4), 209-215. https://doi.org/10.1680/geolett.12.00044

Thorndike R L. (1953). Who Belongs in a Family?. Psychometrica, 18, 267-76.