Proceedings of the 8th World Congress on Civil, Structural, and Environmental Engineering (CSEE'23)
Lisbon, Portugal – March 29 – 31, 2023
Paper No. ICGRE 152
DOI: 10.11159/icgre23.152

Interpretation of Cone Penetration Tests to Characterize Tropical Residual Soils Using Machine Learning

Jeniffer Viegas[1], António Gallardo[1], Lucas Bottaro[1], Rodrigo Marinaro[2]
[1]TPF Consultores de Engenharia e Arquitetura S.A
1050-138, Lisbon, Portugal
jeniffer.viegas@tpf.pt
[2]Vale S.A
34006-270, Nova Lima, MG, Brazil

Abstract - Cone penetration test (CPT) has been strongly applied to identify the soil profile and to provide some estimation of soil parameters. Several correlations exist, allowing the geo-characterization of the soil from CPT data. Such correlations must be carefully applied, and whenever possible, corrected with direct measurements of laboratory tests. Tropical residual soils have an inherent variability capable of providing very distinct results from very similar samples. Project designers must deal with this variability and correctly characterize these materials. The present work focuses on a case study where the goal was to distinguish and characterize two soft soils existent on the foundation of a tailings dam in the southwest of Brazil. The construction of the dam is still ongoing, and its foundation belongs to a complex geological environment with soft soils that can reach N_{SPT} blows as low as its own weight. The geological survey identifies two horizons of residual soil of dolomitic phyllite: soft and very soft. However, distinguishing spatially this material regarding its consistence has shown to be a challenging task. Since they differ essentially on the degree of weathering, most parameters for both materials are quite similar, and from laboratory tests, the parameter that helps differentiate these soils is the pore pressure Skempton parameter at failure – A_f. In addition, the groundwater level in the area is not clear, complicating the estimation of the vertical effective stress profile and further parameters from the CPT analysis. To overcome this issue, a sensitive analysis of the influence of groundwater level on the parameters of interest in this work (apparent overconsolidation ratio) was performed. To get as much information as possible from all datasets available, an Exploratory Data Analysis (EDA) followed by the application of an unsupervised learning algorithm was performed. Although an exactly spatial division from these soils were not possible, the EDA and unsupervised learning allow better visualization of the spatial distribution of these soils and grouping by desired characteristics, such as the pore pressure parameter.

***Keywords*:** CPT Analysis; Machine Learning; Geotechnical Characterization; Saprolitic Soil of Dolomitic Phyllite; Tropical Residual Soils.

1. Introduction

Distinguishing soils and characterizing them is one of the first steps in geotechnical design. In large projects, the number of field investigation can easily reach hundreds and how one organizes and extract information from these data can be the key for success. Statistical analysis can be implemented to help dealing with a large dataset. One of the goals of this work is to identify two kinds of soils of same origin and thus very similar: saprolitic soil of dolomitic phyllite and residual soil of dolomitic phyllite.

A common method used to create groups is cluster analysis [1]. The application of this method with Cone Penetration Test (CPT) is common and have been applied for different purposes. In [2], a cluster analysis technique is applied to group CPT data based on the normalized cone resistance, the friction ratio, and the soil behavior index, which led to a layer grouping used later to determine the soil rigidity model. In [3], cluster analysis was used to improve total weight prediction, in [4] to classify different types of phosphogypsum in a stack where the Soil Behavior Type system (SBTn) did not succeed and in [5] to group soil layers, and delineate the lenses and outliers within a sub-layer. Despite the goal: grouping, each of the example cited resorted a specific mathematical approach, such as the K-means, the Gaussian Mixture Model (GMM) or the fuzzy C-means. The present work applies the K-means method to identify two soft soils.

2. Methodology

2.1. Site and Data Description

This case of study refers to materials existent on the offset of a tailings dam located on the *Quadrilátero Ferrífero*, an iron ore rich area located in the state of Minas Gerais, southwest of Brazil. The target soil occurs on the foundation of a tailings dam. The identification and characterization of these soft materials are necessary to accomplish the second phase of the construction, since they were identified on the foundation of the downstream rise of the structure on the left abutment. Although there are hundreds of surveys, the present study focuses on the data localized in the interest area. The available data in this area is presented in Fig. 1, where the red squares represent laboratory samples (triaxial tests and characterization tests – Atterberg limits and granulometric distribution), the yellow circles represent cone penetration tests with measurement of pore pressure, and the black cross represents standard penetration tests. Some of the standard penetration tests (SPT) have reached N_{SPT} blow as low as its own weight, revealing a very soft soil that could not be limited to a specific layer, as shown in Fig. 2. The goal of this work is to identify these soils in CPT results and to understand its spatial distribution. To deal with the large amount of data, all information was treated with an opensource programming language, allowing the standardization and fast treatment of all data. The CPT data were analyzed following the approach in [6]. Fig.3 shows an example of the CPT data with identification of the investigated soils, using SPT investigations existent on the area to confirm the layer zone. For some CPT, the water level was not clear, and a sensitive analysis of the influence of a wrong estimation of the water level in the apparent overconsolidation ratio (OCR) was performed. OCR was defined as:

$$OCR = 0.25 * \frac{q_n}{\sigma'_{v0}} \tag{1}$$

It was possible to see that a wrong estimation of the water level within 3m would impact in a difference in the OCR value about 0.2 to 0.4, as shown in Fig. 4. This figure shows the comparison of residuals for different values of the water table. The residual is determined by the difference between the value of the OCR parameter calculated through CPT correlations that consider the water table level provided by the CPT dissipation test (or its indication in the nearest drilling report) and the OCR value obtained when an increase (shift – always deeper) is made in the value of the water table, of 3m, 5m, 7m or 10m.

2.2. K-Means Method

In this study, the K-means method was applied to group laboratory samples and later, to group CPT data. The K-means algorithm [1],[7],[8] clusters the *n* data points into *K* disjoint clusters *C*, defined as input from the user. The grouping is performed by minimizing the distance from the sample point, x_i to the group center (mean of the samples, μ_j), known as the inertia or within-cluster sum-of-squares criterion.

$$\sum_{i=0}^{n} \min_{\mu_j \in C}\left(\left\|x_i - \mu_j\right\|^2\right) \tag{2}$$

3. Results

2.3. Laboratory Data

Atterberg limits, granulometric distribution, specific weight of particles and undrained triaxial tests were analyzed to distinguish the two materials and to find any parameter that could be chosen as a flag for each material. The selection of the parameters used to perform the cluster analyses was based on the Principal Feature Analysis (PFA) [9] that selects a subset of the original features that contains most of the essential information. Several combinations of features were tested, and the select features were the pore pressure Skempton parameter at failure A_f, the specific weight of solid particles γ_s, and a spatial variable, the X coordinate of the location where the sample was collected. The number of clusters was set to two, since the goal was to separate the sample within the two soft soils. Fig.5 presents the scatter plot for all features in the three planes where the points are colored by the result from the cluster analysis. Cluster 1 (purple color) represent a material with lower specific weight of solid particles and higher pore pressure A Skempton parameter at the failure. From the granulometric curve in Fig.6, it is possible to see that this material is mainly silty. The data from triaxial undrained tests were treated according [10], where the stress paths are presented in the axis:

$$s' = \frac{\sigma'_1 + \sigma'_3}{2} \tag{3}$$

$$t = \frac{\sigma'_1 - \sigma'_3}{2} \tag{4}$$

With σ'_1 and σ'_3 being the major and minor principal effective stresses. The pore pressure A Skempton parameter is defined as:

$$A = \frac{\Delta u}{\sigma'_1 - \sigma'_3} \tag{5}$$

In Fig.7 a), the stress paths are plotted for all samples and in Fig.7 b) the pore pressure A Skempton parameter is displayed as a function of t. In both figures the samples are colored according to the group resulting from the cluster analysis. One can see that the cluster 1 shows a contractile soil with higher pore pressure A Skempton parameter, while cluster 2 shows the opposite.

2.4. CPT Data

The CPT data shown in Fig.1 was treated following [6] and a cluster analysis analog to the one used on the laboratory tests data was also applied to the CPT data. The goal of this section was to find the soft and very soft soil identified on SPT and triaxial samples in the CPT data. Three tests were performed with the same variables as features: pore pressure ratio B_q; soil behavior index I_c; and friction ratio R_f. The difference in the tests was the number of clusters to form, and since it was not possible to know in advance the exact number of soils presented in the CPTs (which included soils other than the 2 of interest), simulations were conducted varying from 3 to 5 clusters (groups of soils). All tests created a group for soil with low tip resistance ($q_c \leq$ 5Mpa), low OCR, and high pore pressure ratio. Fig. 8 shows the result for the test with 4 clusters, in terms of a) cone resistance, qc, b) pore pressure ratio, Bq, c) apparent overconsolidation ratio, OCR, d) soil behavior type index, I_c, e) friction ratio, R_f and f) the 3D distributions of the clusters. The group mentioned above is identified as cluster 1 (cian). Cluster 1 could be the very soft soil identified in the laboratory samples (cluster 1 in Fig. 7). The only issue in this classification would be the I_c parameter, that is in the range of 2.8-3.5, suggesting a material with silty clay to clay behavior, while the granulometric distribution of the soil of interest shows up to 95% of silt. Is important remember, though, that the I_c is a behavior type index, and regardless the soil being almost integrally silty, due to some structure or even the presence of a relatively small percentage of clay it could easily present a response to the probe penetration similar to a clay or silty clay. This group could not be assigned as the soft soil (cluster 0 in Fig.7) due to the elevated pore pressure generated during to the penetration of the probe, since the laboratory tests showed lower pore pressure A Skempton parameter at failure. In addition, when comparing adjacent SPT and CPT, it is possible to understand that the major difference between the soft and very soft horizons is indeed the pore pressure, as shown in Fig.3, where for the very soft soil the excess pore pressure is higher than 500 kPa. From Fig. 8 e), it is possible to note that the existence of cluster 1 cannot be restricted to one layer or specific zone, since this material is spread without a clear pattern all over the study area. To check the matching between soil properties and to verify if cluster 1 is the very soft soil, the data were plotted in 3D. Fig. 9 presents the data colored by a) the cone resistance, q_c, b) the friction ratio, R_f, c) the pore pressure during the cone penetration, u, and d) the pore pressure ratio, B_q. Comparing these figures with Fig. 8 e), it is possible to conclude that cluster 1 has low cone resistance q_c, high pore pressure u (and consequently pore pressure ratio), and high friction ratio, R_f. The high pore pressure levels due to penetration corroborate the triaxial test results, showing that this is a soil that generates elevated excess pore pressure when subject to shear. On the other hand, cluster 0 (dark blue) could be the soft soil, which also present low cone resistance and high friction ratio, but with lower pore pressure generated.

4. Conclusion

The interpretation and cluster analysis of the laboratory tests helped to understand the behavior of these soft and very soft residual soils when subject to shear, showing that the pore pressure parameter at failure (Af) is the main difference between them. Knowing that these materials tend to differ in such manner increased the confidence in establishing Bq as the parameter on CPT that better distinguish them, as clearly observed in the comparison of SPT and CPT surveys. This is a very interesting outcome of this study, since tip resistance and friction ratio are more commonly used for differentiating materials than Bq. Nevertheless, to confirm the results from the CPT cluster analysis, it would be valuable to have samples in-depth, since all laboratory samples were collected at the surface.

Although it was not possible to establish clear layers of very soft soil and the soft soil in CPT soundings, it was possible to see the location where the parameters indicate the possible existence of these soils. This lack of well-defined layers in this study should not be interpreted as an incapacity of the model or inadequacy of the laboratory tests or field surveys, but rather an intrinsic geological characteristic of these tropical residual soils, which show very complex weathering patterns and can not be compared to the layered and better-defined depositions of sedimentary formations, for instance. It is safe to say that there is no guarantee that the very soft soil is restricted to a layer or a specific zone, and treating this problem in such way would not be the best approach, especially for ultimate state evaluation, such as a limit equilibrium analysis. A more realistic methodology could be to understand the spatial variability of the soil parameters and generate random fields, based on the scale of fluctuation observed.

References

[1] MacQueen, I. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings 5th Berkeley Symposium on Mathematical Statistics Problems (pp. 281-297).

[2] Młynarek, Z., Wierzbicki, J. & Wołyński, W. (2018). Use of functional cluster analysis of CPTU data for assessment of a subsoil rigidity. Studia Geotechnica et Mechanica,40(2) 117-124. https://doi.org/10.2478/sgem-2018-0017

[3] Collico, S., Arroyo, M., DeVincenzi, M., Rodriguez, A., & Deu, A. (2022). Clustering analysis to improve total unit weight prediction from CPTu. In Cone Penetration Testing 2022 (pp. 339-344). CRC Press.

[4] Coelho, P, Nogueira, G., Katahira, D., Franco, J. (2022). Interpretação de Ensaios CPTu Realizados em Pilha de Fosfogesso por Meio de Técnicas de Estatística Multivariada. COBRAMSEG 2022. ISBN: 978-65-89463-30-6.

[5] Yin, J., Opoku, L., Miao, Y. H., Zuo, P. P., Yang, Y., & Lu, J. F. (2021). An improved site characterization method based on interval type-2 fuzzy C-means clustering of CPTu data. Arabian Journal of Geosciences, 14(14), 1-11.

[6] Robertson, P. K., & Cabal, K. L. (2015). Guide to cone penetration testing for geotechnical engineering. Gregg Drilling & Testing, Inc, 6.

[7] Clustering. Scikit-learn.org. (Accessed in December 2022). https://scikit-learn.org/stable/modules/clustering.html

[8] Vassilvitskii, S., & Arthur, D. (2006, June). k-means++: The advantages of careful seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (pp. 1027-1035).

[9] Lu, Y., Cohen, I., Zhou, X. S., & Tian, Q. (2007, September). Feature selection using principal feature analysis. In Proceedings of the 15th ACM international conference on Multimedia (pp. 301-304).

[10] ASTM. (2011). Standard test method for consolidated undrained triaxial compression test for cohesive soils. ASTM International - Committee D-18 on Soil and Rock.

Appendix A – Figures

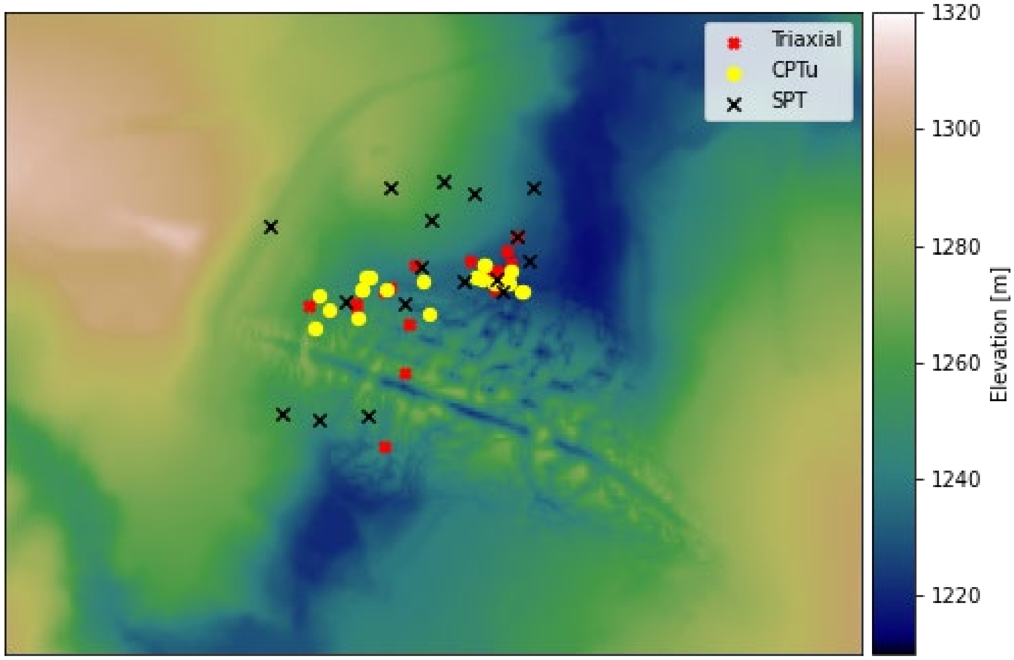

Fig. 1- Location of the investigations: Cone Penetration Test (CPTu), Standard Penetration Test (SPT) and samples collected for Triaxial tests.

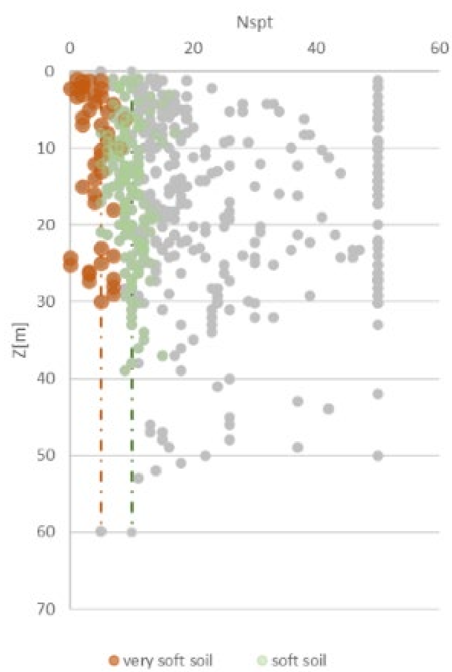

Fig. 2- Blow counting, N_{SPT}, in depth for all SPT interpreted. It is possible to notice that the very soft soil and soft soil are not restricted to a specific depth, suggesting the absence of well-defined layers. Very soft soil: $N_{SPT} \leq 5$. Soft soil: $5 \leq N_{SPT} \leq 10$. The data which appear out of the specified ranges correspond to short points of better or worst consistence enclosed by larger passages, not representative of a significant change in behavior.

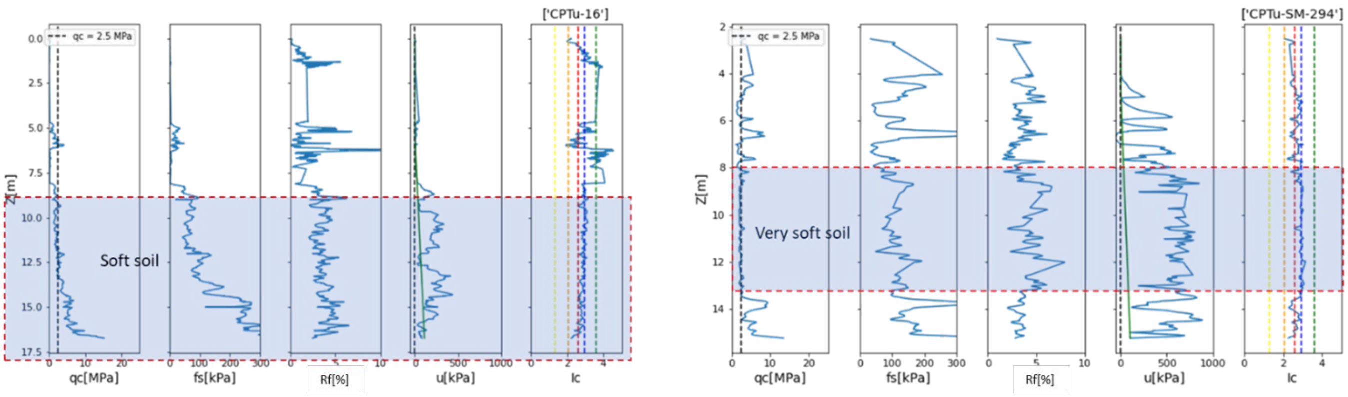

Fig. 3- Example of CPT data for the soft soil and the very soft soil: cone tip resistance q_c, sleeve friction f_s, friction ratio R_f, pore pressure generated during the cone penetration u, and the soil behaviour type index I_c. The main difference between the two materials is the pore pressure, u. The position of these layers was defined in accordance with SPT test.

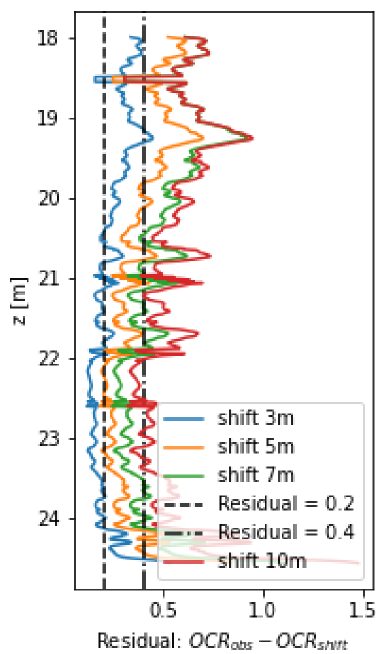

Fig. 4- Influence of the water table on the apparent OCR: comparison of residuals for different values of the water table.

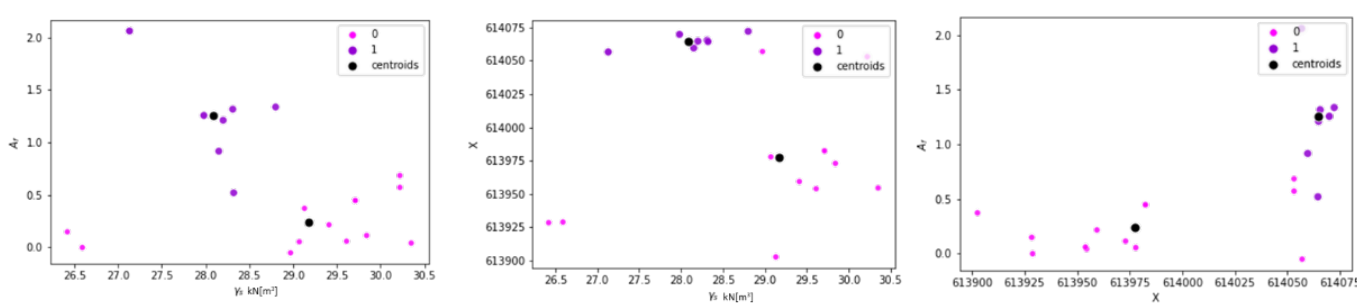

Fig. 5- Scatter plot of the cluster features from the laboratory tests at the three planes.

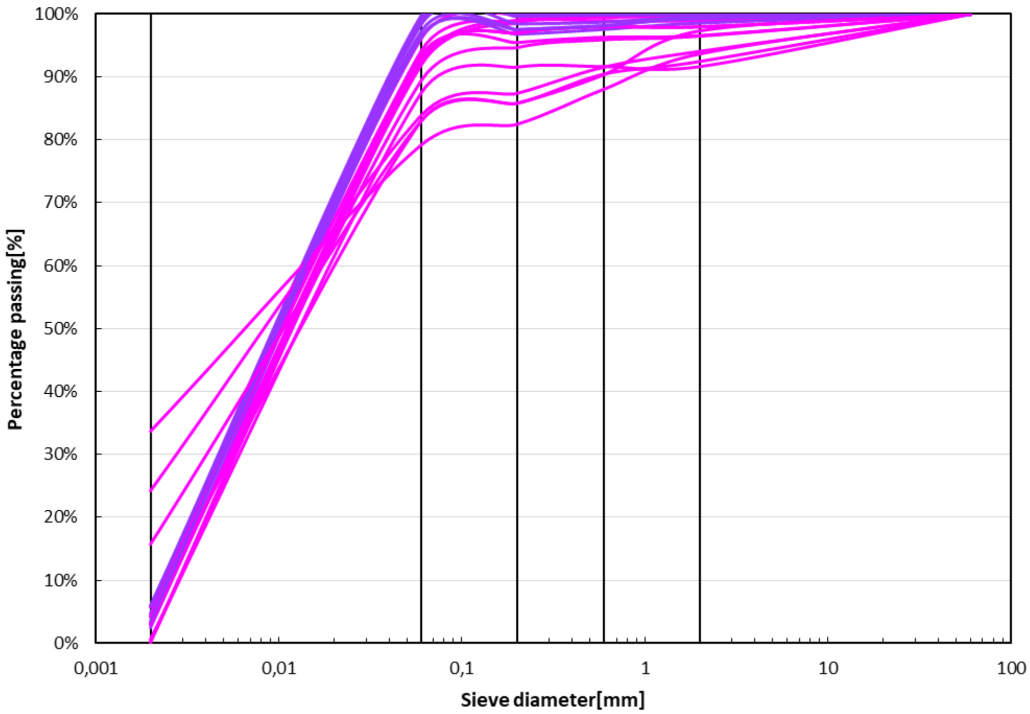

Fig. 6- Granulometric distribution for laboratory samples coloured by the clusters (Purple – cluster 1, Pink – cluster 0). Cluster 1 presents a material with a slightly higher percentage of silt.

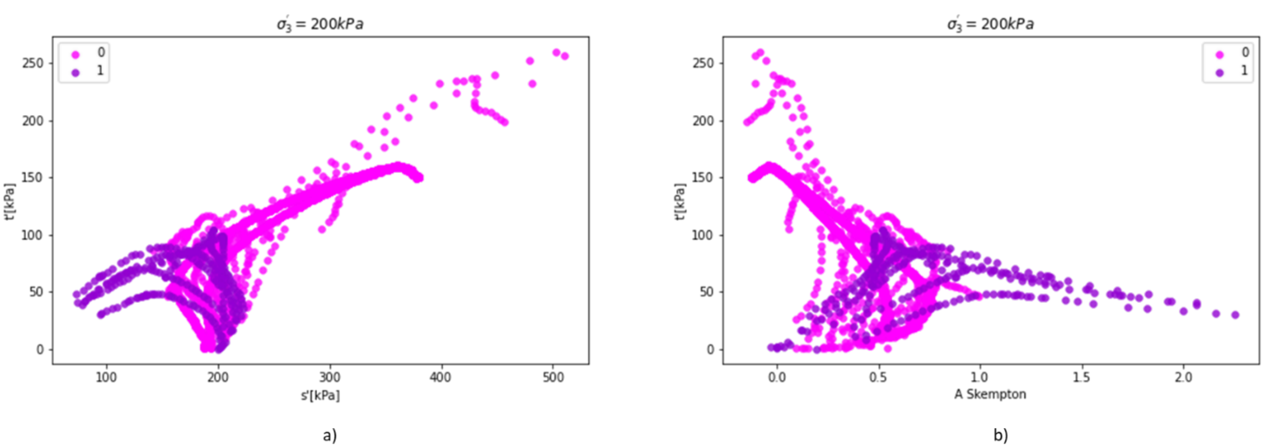

Fig. 7- Triaxial results for the confining stress σ’$_3$ = 200kPa: a) stress path in the s’ – t space, and b) pore pressure parameter in the A – t space, colored by the clusters. Cluster 1 shows a contractile soil with higher pore pressure A Skempton parameter at failure.

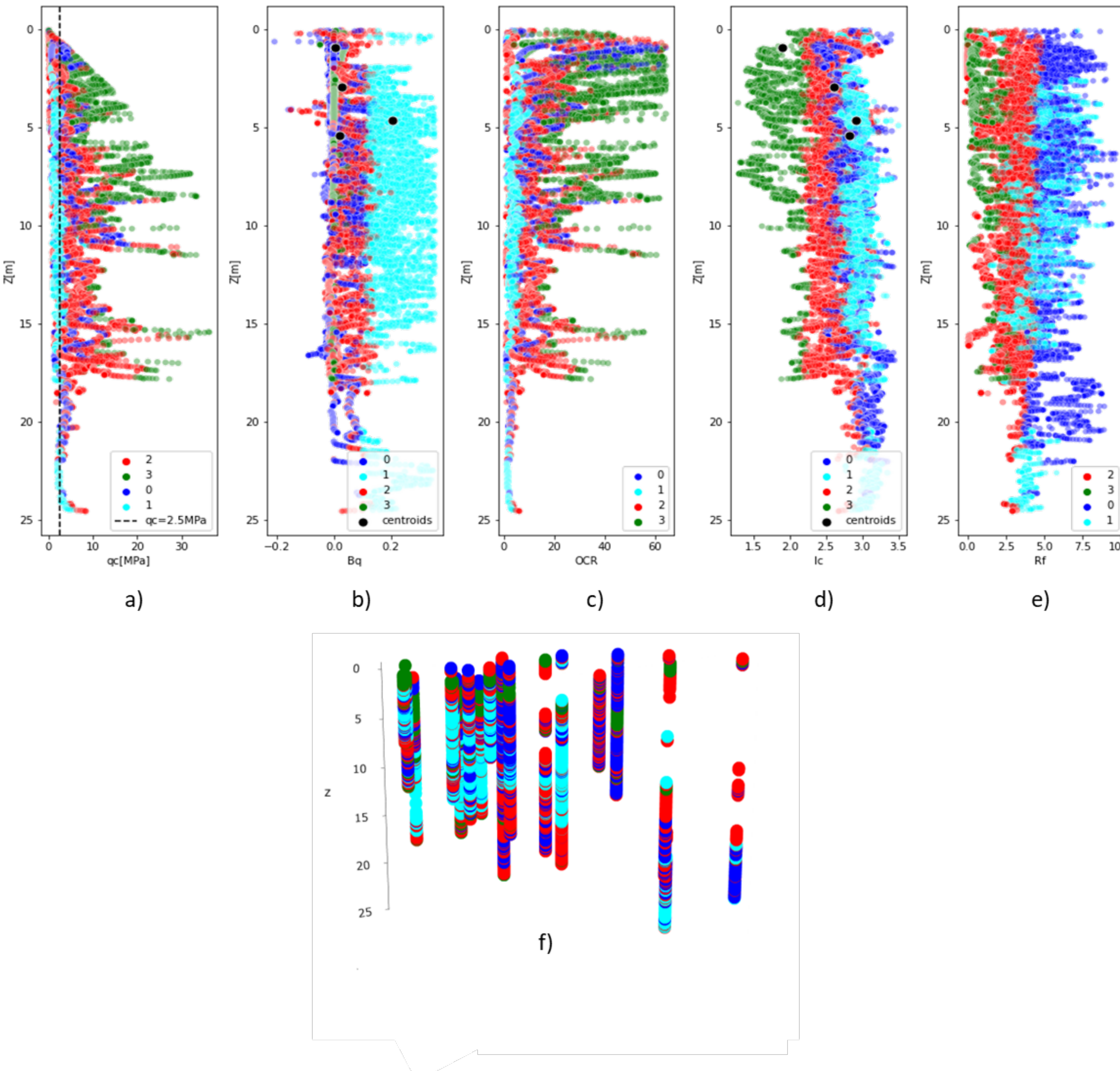

Fig. 8- CPT cluster analysis results: a) Cone tip resistance, q_c, b) Pore pressure ratio Bq, c) Apparent overconsolidation OCR, d) Soil behavior type index I_c, e) Friction ratio R_f, and f) 3D plot of the CPT, colored by the clusters.

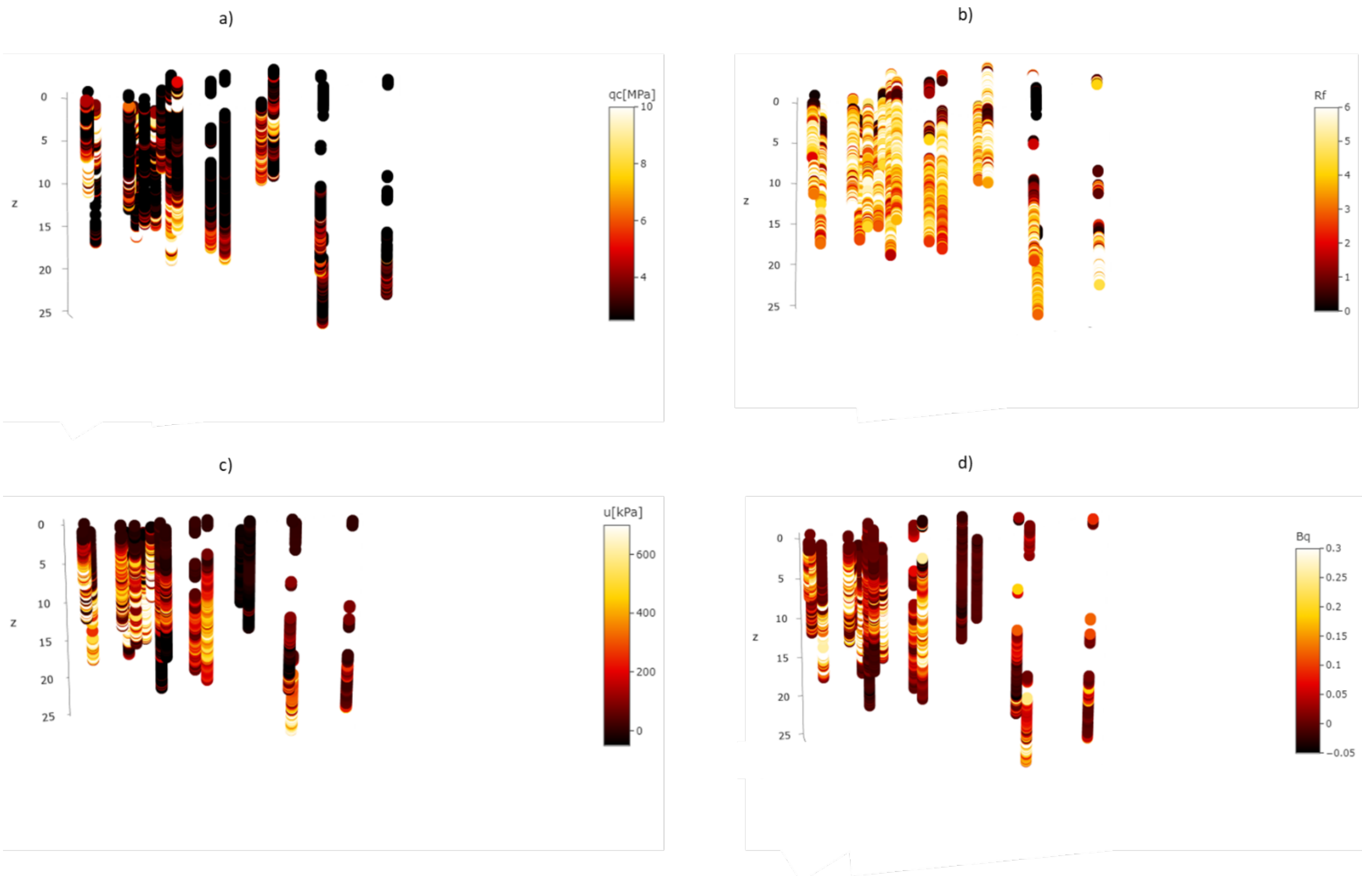

Fig. 9- Spatial distribution of CPT tests, colored by a) Cone resistance q_c, b) Friction ratio R_f, c) Pore pressure generated during cone penetration u, and d) Pore pressure ratio Bq.