Soil stratigraphy from seismic piezocone data and multivariate clustering in alluvial soil deposits: Experience in the Lower Tagus Valley region

F. Molina-Gómez, D. Cordeiro, C. Ferreira & A. Viana da Fonseca
CONSTRUCT-GEO, Faculdade de Engenharia da Universidade do Porto (FEUP)
University of Porto, Portugal

ABSTRACT: The identification of soil stratigraphy at a given site is crucial for geotechnical analysis and design. This study addresses the delineation of soil stratigraphy from SCPTu measurements. For this purpose, SCPTu data from an experimental site in the Lower Tagus Valley region (close to Lisbon) were compiled. Soil stratigraphy was obtained by applying a multivariate clustering approach of seismic piezocone penetration test (SCPTu) data. The hierarchical clustering results were compared against the soil behaviour index profile. Moreover, the differences/similitudes obtained from multivariate clustering were validated by contrasting statistical results against the visual description of samples collected in the experimental site at Lower Tagus Valley using advanced sampling techniques; namely, Gel-Push sampler. The main findings showed that the clustering approach implemented herein detects and groups soil layers with similar soil behaviour types, allowing delineating the soil stratigraphy in multilayer alluvial soil deposits.

1 INTRODUCTION

The identification of soil stratigraphy is one of the main concerns for the characterisation of soils deposits. A realistic identification of the soil variability in depth reduces the failure probability of geotechnical structures. The cone penetration test (CPT) and its enhanced versions provide reliable identification of soil multi-layering due to its simultaneous measurements. Hence, soil stratigraphy can be assessed using CPT data by correlating in situ measurements with parameters inferred by applying a strong theoretical background.

Although no samples are recovered in CPT sounding, the large number of data from this type of testing allows identifying the soil stratigraphy. Therefore, such data can be interpreted by statistical methods. Hegazy & Mayne (2002); Liao & Mayne (2007); Molina-Gómez, Viana da Fonseca, Ferreira, Sousa, & Bulla-Cruz (2021) applied cluster analysis approaches to delineate the stratigraphy in different types of soil deposits. In these studies, the authors compared their statistical analyses against soil behaviour profiles, originally proposed by (Robertson 1990). Such a comparison showed a good fitting between results, validating the applicability of clustering for delineating the soil stratigraphy.

This study addresses the delineation of soil stratigraphy from seismic piezocone (SCPTu) measurements by applying the clustering approach proposed by Molina-Gómez, Viana da Fonseca, Ferreira, Sousa, & Bulla-Cruz (2021). The SCPTu measurements were carried out in an experimental site located at Lower Tagus Valley, south of Portugal. Embedded layers of sandy and clayey soils with an alluvial origin from the Tagus River compose such an experimental site (Viana da Fonseca, Ferreira, Ramos, & Molina-Gómez 2019). Besides, in such a site, high-quality undisturbed samples were collected using an advanced sampling technique —the Gel-Push (GP) sampler. The visual observation of these samples confirmed the detected stratigraphy from the multivariate clustering.

2 DESCRIPTION OF THE CLUSTERING APPROACH

In this study, the agglomerative hierarchical clustering approach proposed by Molina-Gómez, Viana da Fonseca, Ferreira, Sousa, & Bulla-Cruz (2021) was applied. Such an approach comprises the following phases: (*i*) selection of the data and identification of variables; (*ii*) standardisation of the variables; (*iii*) computation of the distance (or similarity) matrix; (*iv*) application of clustering technique; (*v*) selection of the cut level of the dendrogram, associated with the best partition, for defining the number of clusters; (*vi*) interpretation of cluster results; and (*vii*) validation by comparison against soil behaviour type profiles. Besides, it includes a comparison against visual examination of high-quality undisturbed samples collected using the GP sampler, as a novel validation in such a clustering approach. The clustering procedure adopted for this study is described following.

For the data standarisation, the Zscore method was adopted (Hegazy & Mayne 2002):

$$Z_{ij} = \frac{x_{ij} - \bar{x}(X_j)}{s(X_j)} \tag{1}$$

where Z_{ij} is the standardised value, x_{ij} is the SCPTu measurement in certain depth of the SCPTu variable, X_j is the SCPTu variable of each profile, $\bar{x}(X_j)$ is the average of the SCPTu variable and $s(X_j)$ is the standard deviation of the SCPTu variable.

The distance among pairs of objects is represented by a matrix $D_{(n\times n)}$ (Härdle & Simar 2015). Equation 2 describes the matrix distance, D, that contains measures of similarity among the n objects, where $d_{i,j}$ are the distances between the measurements of SCPTu variables.

$$D = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & \cdots & \cdots & d_{1,n} \\ \vdots & a_{2,2} & \ddots & & & \vdots \\ \vdots & a_{3,2} & & \ddots & & \vdots \\ \vdots & a_{4,2} & & & \ddots & \vdots \\ d_{n,1} & d_{n,2} & \cdots & \cdots & \cdots & d_{n,n} \end{bmatrix} \tag{2}$$

Nowadays, there are different similarity measurements, including Euclidean distance, Manhattan distance, Pearson, cosine and power (Johnson & Wichern 2007). Härdle & Simar (2015) claimed that no specific rule or theory is available on which to base the choice of similarity measurements. In this study, Euclidean distance was used to compute the matrix D of piezocone data, as recommended by (Liao & Mayne 2007). Such a method estimates the geometric distance—based on the Pythagorean theorem—of two observations in a multidimensional space. Equation 3 describes the Euclidean distance of two observations, where k is the number of variables and i is the variable index.

$$d(P,Q) = d(Q,P) = \sqrt{\sum_{i=1}^{k} (q_i - p_i)^2} \tag{3}$$

Starting with the matrix D (see Equation 2) an iterative procedure takes place to join/grouping the most similar classes and update matrix D, which ends when all data is grouped in a particular cluster (Bulla-Cruz, Lyons, & Darghan 2021). The iterations are carried out according to a specific aggregation criterion. In this study, the complete linkage method was applied. The algorithm of complete linkage distance is:

$$d(R, P+Q) = \max\{d(R,P), d(R,Q)\} \tag{4}$$

where $P+Q$ are clustered measurements of SCPTu variables and R is a possible new group.

The best-known result of a Hierarchical Classification is a graphical representation—dendrogram. The quality of SCPTu data grouping (i.e. the number of clusters) can be assessed by comparing the number of soil classes identified by the SBT chart (Robertson 1990). Afterwards, such a number is validated by the Average Silhouette Method (Rousseeuw 1987). The Average Silhouette Method computes the average silhouette of observations for different numbers of clusters (K), indicating K_{opt} value as the maximum average silhouette over a range of possible values for a different number of clusters.

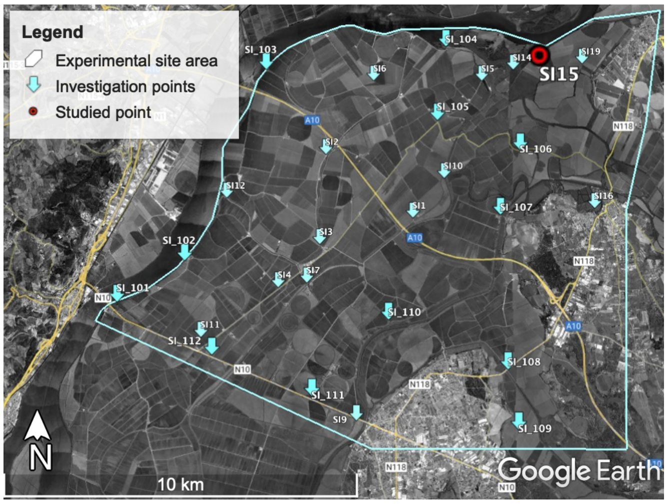

Figure 1: Experimental site map and location of the investigation points

As before noticed, the underlying mathematics of cluster analysis is simple to implement. However, it requires extensive calculations because of the large amount of data. Therefore, all computations were conducted in **RStudio**, which is a free integrated environment for statistical programming and graph plotting (Molina-Gómez, Bulla-Cruz, & Darghan 2019).

3 EXPERIMENTAL SITE

Within the scope of two research projects on soil liquefaction developed in the CONSTRUCT-GEO research centre of FEUP, a vast site characterisation campaign was carried in an experimental site in the Lower Tagus Valley (south of Portugal). The site characterisation comprised the definition of liquefaction risk by microzonation maps (Saldanha, Viana da Fonseca, & Ferreira 2018). This geotechnical test site is located in the municipalities of Vila Franca de Xira and Benavente, covering a zone with 14683 ha area and 50.8 km perimeter. In situ tests for geotechnical characterisation were performed at specific locations, referenced as site investigation points (SI). Figure 1 shows the map and SI locations within the experimental site.

For this study, SI15 was selected since there advanced sampling techniques were implemented for collecting high-quality samples for the first time in Portugal (Molina-Gómez, Viana da Fonseca, Ferreira, Ramos, & Cordeiro 2021). The groundwater level in SI15 was identified at about 2.5 m depth. Moreover, in this point, an SCPTu was carried out. Shear wave velocities (Vs) were measured at each 0.50 m depth. The values of cone resistance (q_c), sleeve friction (f_s) and pore-water pressure generated during cone penetration (u_2) were averaged each ± 0.25 m depth (ranges of 0.5 m), considering the depth of the V_S measurements, to compare the SCPTu data profiles. Figure 2 presents the SCPTu profiles and the averaged values in the specific depths of V_S measurements at SI15.

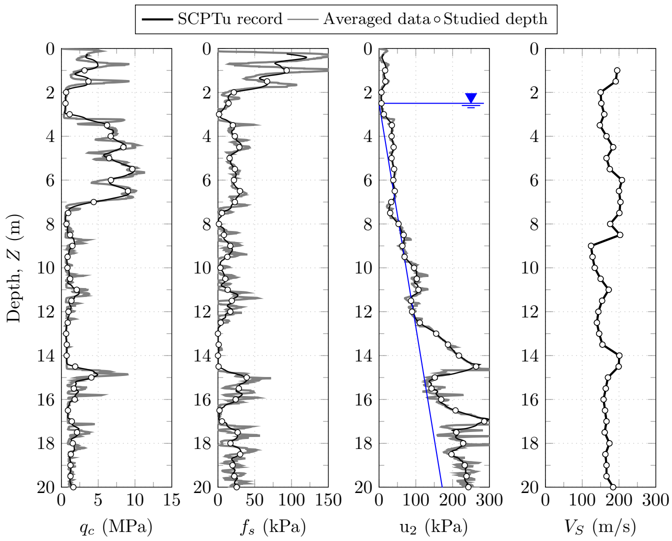

Figure 2: SCPTu profiles at SI15

The SCPTu data were analysed using the software Cliq v.3.0. The analyses were carried out based on the unified approach proposed by Robertson (1990), which includes the classification of soil behaviour type (SBT) zones according to the soil behaviour type index (Ic). Ic computation considers the following procedure:

$$Q_t = \frac{q_t - \sigma_{v0}}{\sigma'_{v0}} \tag{5}$$

$$Fr = \left(\frac{f_s}{q_t - \sigma'_{v0}}\right) 100\% \tag{6}$$

$$Ic = \left[(3.47 - \log Q_t)^2 + (\log Fr + 1.22)^2\right]^{0.5} \tag{7}$$

where q_t is the cone resistance corrected for pore water effects ($q_t = q_c + u_2(1-a)$); a is the cone area ratio; σ_{v0} is the current in situ total vertical stress; σ'_{v0} is the current in situ effective vertical stress.

Figure 3 presents the interpretation of the soil profiles in terms of soil type chart. Classification results indicate that the soil profile at SI-15 has five different soil types, which are characteristic of soil deposits in the Lower Tagus Valley from the Quaternary (Ferreira, Viana da Fonseca, Ramos, Saldanha, Amoroso, & Rodrigues 2020).

The soils of the experimental site appear in zones **3**, **4**, **5** and **6**, thus typically behave as ‘clay to silty clay’, ‘clayey silt to silty clay’, ‘silty sand to sandy silt’ and ‘clean sand to sand to silty sand’, respectively. Points located in Zone **1** were not considered for defining the number of clusters because there is not a well-defined characterisation for these soil types; that is, out-of-range zones with $Ic =$ N/A. Based on the above consideration, the clustering of SI15 can be performed considering four groups.

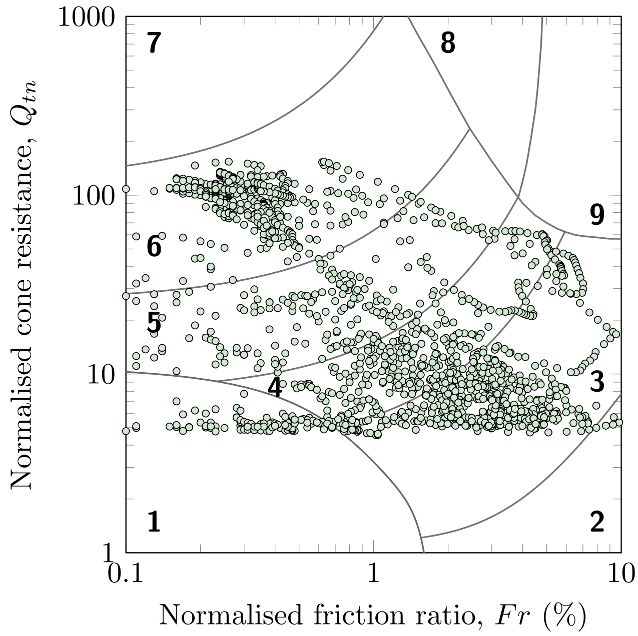

Figure 3: Results of soil classification type

4 ADVANCED SAMPLING

The collection of high-quality samples is critical to obtain representative test results for soil characterisation in the laboratory (Ferreira, Viana da Fonseca, & Nash 2011). A typical sampling campaign involves the following stages: drilling, sampler insertion, sampler retrieval, tube sealing, transport, soil extrusion, sample storage and preparation for element testing (Hight 2000). Nowadays, sampling processes are conducted using diverse techniques, such as hydraulically activated push tubes, rotary devices and freeze sampling, where the freezing method is the most expensive (Viana da Fonseca & Pineda 2017).

In this research, an advanced sampling technique developed by the Japanese geotechnical company Kiso-Jiban Consultants, to develop a method able to replace freezing techniques. There are four different variations of GP: GP-Rotary, GP-Drilling, GP-Triple, and GP-Static. In this study, the GP-Static (GP-S) was implemented. This sampling device uses a hydraulic activated fixed-piston and follows the same principle of the Osterberg-type sampler.

On the other hand, the GP is an advanced sampling technique, which uses a viscous polymer gel to collect high-quality undisturbed samples. The gel use, as the main innovation for soil sampling, afforded the name for this technique. The gel must be prepared at a 1 vol% concentration ratio of the viscous polymer in clean water (Mori & Sakai 2016). The purpose of using the viscous gel is to significantly reduce the friction between the sample and walls liner during both the insertion of the sampler into the ground and during the sample extrusion in the laboratory (Viana da Fonseca, Ferreira, Molina-Gómez, & Ramos 2019). Besides, the rheological properties of the polymer gel allow preserving the soil structure of the collected samples.

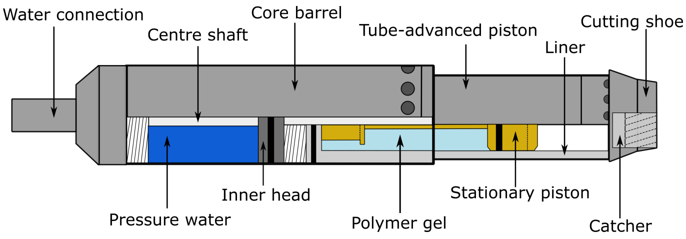

Figure 4: Schematic of GP sampler components (Molina-Gómez, Ferreira, Ramos, & Viana da Fonseca 2020)

Unlike conventional hydraulic activated fixed-piston, the GP-static device includes three pistons: (i) the stationary piston, (ii) the sampling tube-advancing piston, and (iii) the core-catcher activating piston. At the maximum performance, the GP sampler can collect samples with 71 mm diameter and 1 m length. However, according to the soil type to be sampled, the device can suffer damages in the cutting shoe (Molina-Gómez, Viana da Fonseca, Ferreira, Ramos, & Cordeiro 2021). Details about the performance and quality sampling using the GP sampler are provided in Molina-Gómez, Ferreira, & Ramos (2020). Figure 4 schematises the components of the GP-Static sampler.

The GP sampling was carried out in four boreholes located at about 5 m distance to the SCPTu sounding. During the experimental campaign, a total of 14 GP samples were collected at the SI15 location. The sampling depths were selected considering suitable layers for liquefaction assessment by element testing in the laboratory. The liquefiable layers were defined by detecting zones with $Ic < 2.6$. Figure 5 illustrates the sampling depths for each technique. Moreover, in Figure 5, the sampling depths are contrasted against the averaged soil behaviour index profile of SI15.

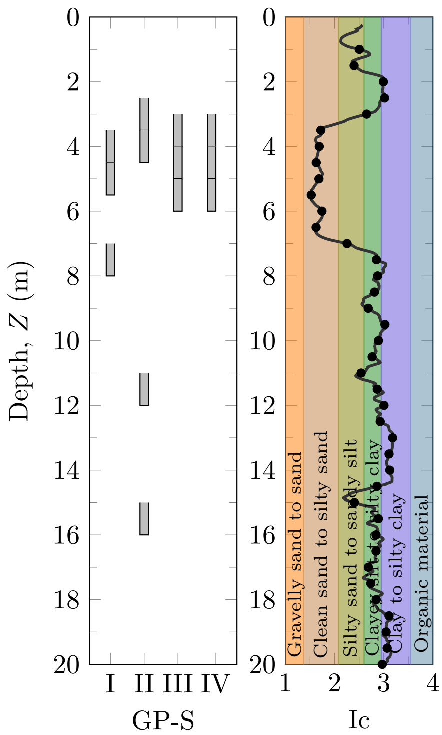

Figure 5: Sampling depths for each advanced technique and averaged soil behaviour index profile at SI15

5 ANALYSIS OF RESULTS

A cluster analysis in terms of the depth (Z) and using the standardised data recorded at SI15 by the four channels of SCPTu (e.g. q_c, f_s, u_2 V_S) was carried out. Liao & Mayne (2007) used normalised data to define the soil stratigraphy by applying a multivariate approach (e.g. Q_t and Fr parameters). However, Molina-Gómez, Viana da Fonseca, Ferreira, Sousa, & Bulla-Cruz (2021) obtained a better groping using standardised raw data instead of standardised normalised data. These authors showed that the differences between clustering were because of the uncertainties associated with the in situ stress-state assessment, which depends on the unit weight (γ). Nevertheless, γ is a physical property that differs according to the CPT correlation applied, although the literature reports several correlations to estimate such a physical parameter, such as Mayne, Peuchen, & Bouwmeester (2010, Robertson & Cabal (2010, Lengkeek, de Greef, & Joosten (2018).

In this study, only the clusters with at least two continuous measurements are considered soil layers, as suggested by Liao & Mayne (2007). Hence, a minimum layer thickness of 1.0 m was defined. Clusters with points without a continuous group of data (single points) were considered as lenses or transition zones (Hegazy & Mayne 2002). The cluster analysis was conducted for obtaining four groups, which correspond to different soil behaviour types or layers previously identified in Figure 3. Such a number of groups was compared against the results of the Average Silhouette Method, which allows estimating the optimum number of clusters, K_{opt}. Figure 6 shows $K_{opt} = 4$, validating the selection of the number of clusters based on the SBT criterion.

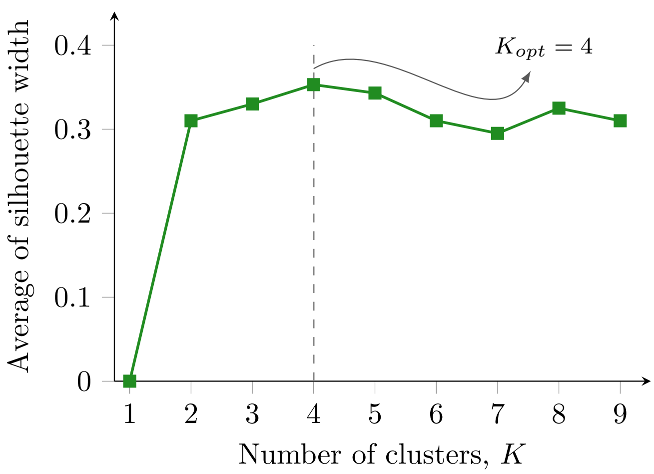

Figure 6: Estimation of the optimum number of clusters

Figure 7 presents the dendrogram obtained from the cluster analysis of SCPTu data. This dendrogram shows the clustering of all depths and the distances between the aggregate clusters. A distance of about 5 defined four clusters, which correspond to the layers composing the soil stratigraphy of the studied zone.

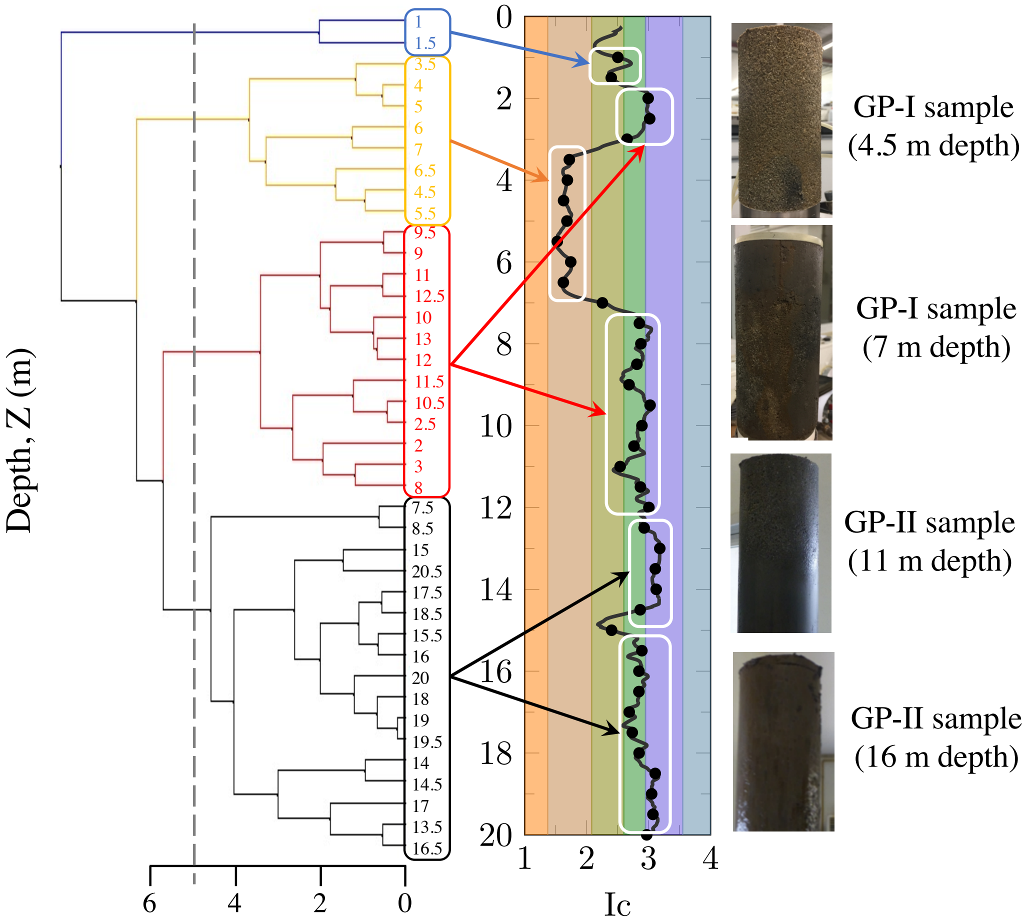

Figure 7: Soil stratrography of SI15 inferred from cluster analysis and comparisons between dendrogram against Ic profile and high-quality samples

In addition, Figure 7 indicate the SBT zone of each cluster through a comparison against the Ic profile SI15 sounding. This grouping represents the soil layers that behave as 'silty sand to sandy silt' (1 m to 1.5 m depth), 'clean sand to sand to silty sand' (3 m to 7 m depth), 'clayey silt to silty clay' (2 m to 3 m depth and 8 m 14.5 m depth) and 'clay to silty clay' (12.5 m to 14.5 m depth and 15.5 m to 20 m depth). Besides, two transitional layers composed of soils that behave as 'silty sand to sandy silt' were identified at 7.5 and 15 m depth.

On the other hand, Figure 7 shows photographs of GP samples collected at representative depths of the layers identified in the cluster analysis (e.g. 4.5 m, 7 m, 11 m and 16 m depth). A visual examination in the laboratory of GP samples supported cluster results; that is, the sample collected at 4.5 m depth is a sand, the sample collected at 7 m depth corresponds to a silt mixture, and the samples collected at 11 m and 16 m depth are clay with different content of granular particles.

The simultaneous comparisons between dendrogram, Ic profile and high-quality GP validated the implementation of the clustering approach to define the soil stratigraphy and detect interlayers from SCPTu data. Therefore, the clustering approach proposed herein provided a reliable data grouping for defining soil stratigraphy of alluvial soil deposits of the Lower Tagus Valley—a region with high interlayering. Moreover, this clustering approach can be applied to other datasets from other experimental sites using raw data from the four parameters of the SCPTu. This information provides a method that allows defining soil layers with similar properties and behaviour, useful in the design of geotechnical structures.

6 CONCLUSION

In this paper, a hierarchical clustering approach for defining the soil stratigraphy from SCPTu data at a site investigation point located in the Lower Tagus Valley (south of Portugal) has been applied. The approach comprises a complete-linkage algorithm, formulated in **RStudio**, which used the Euclidean distance of standardised SCPTu measurements (q_c, fs, u_2 and Vs). The method is objectively capable of classifying similar groups of data in the soil profile, delineating different layer boundaries and soil transitions. The statistical results were validated by comparing the dendrogram against the Ic profile and visual examination of high-quality samples collected using the GP sampler. These comparisons are the main contribution of this study since they confirmed the correspondence between the defined stratigraphy (clusters) with the soil behaviour type and soil composing the soil layers. Therefore, this clustering approach can be applied to other datasets from other experimental sites using raw data from the four parameters of the SCPTu.

ACKNOWLEDGEMENTS

This work was also financially supported by UIDB/04708/2020 and UIDP/04708/2020 of CONSTRUCT – Institute of R&D in Structures and Construction funded by the national funds through the FCT/MCTES (PIDDAC). The first author acknowledges the Portuguese Foundation for Science and Technology (FCT) for the support through SFRH/BD/146265/2019 grant.

REFERENCES

Bulla-Cruz, L. A., L. Lyons, & E. Darghan (2021). Complete-Linkage Clustering Analysis of Surrogate Measures for Road Safety Assessment in Roundabouts. *Revista Colombiana de Estadística 44*(1), 91–121.

Ferreira, C., A. Viana da Fonseca, & D. F. T. Nash (2011). Shear wave velocities for sample quality assessment on a residual soil. *SOILS AND FOUNDATIONS 51*(4), 683–692.

Ferreira, C., A. Viana da Fonseca, C. Ramos, A. S. Saldanha, S. Amoroso, & C. Rodrigues (2020). Comparative analysis of liquefaction susceptibility assessment methods based on the investigation on a pilot site in the greater Lisbon area. *Bulletin of Earthquake Engineering 18*, 109–138.

Härdle, W. K. & L. Simar (2015). *Applied Multivariate Statistical Analysis* (4 ed.). Springer Berlin Heidelberg.

Hegazy, Y. & P. W. Mayne (2002). Objective site characterization using clustering of piezocone data. *Journal of Geotechnical and Geoenvironmental Engineering 128*, 986–996.

Hight, D. W. (2000). Sampling Methods: Evaluation of Disturbance and New Practical Techniques for High Quality Sampling in Soils. In *7º Congresso Nacional de Geotecnia*, Porto.

Johnson, R. A. & D. W. Wichern (2007). *Applied Multivariate Statistical Analysis* (6 ed.). Pearson Education International.

Lengkeek, A., J. de Greef, & S. Joosten (2018). CPT based unit weight estimation extended to soft organic soils and peat. In *Cone Penetration Testing IV: Proceedings of the 4th International Symposium on Cone Penetration Testing (CPT'18)*, Delft, The Netherlands, pp. 389–394. Taylor & Francis Group.

Liao, T. & P. W. Mayne (2007). Stratigraphic delineation by three-dimensional clustering of piezocone data. *Georisk 1*, 102–119.

Mayne, P., J. Peuchen, & D. Bouwmeester (2010). Soil unit weight estimation from CPTs. In *2nd International Symposium on Cone Penetration Testing*.

Molina-Gómez, F., L. Bulla-Cruz, & E. Darghan (2019). Profiles analysis as a modality of repeated measures for comparing grain size distributions in granular bases. *Measurement: Journal of the International Measurement Confederation 146*, 930–937.

Molina-Gómez, F., A. Viana da Fonseca, C. Ferreira, C. Ramos, & D. Cordeiro (2021). Novel sampling techniques for collecting high-quality samples: Portuguese experience in liquefiable soils. In *6th International Conference on Geotechnical and Geophysical Site Characterization, ISC'6*, Budapest, pp. paper 115.

Molina-Gómez, F., C. Ferreira, & A. Ramos, C. Viana da Fonseca (2020). Performance of gel-push sampling in liquefiable soils. *Géotechnique Letters 10*, 256–261.

Molina-Gómez, F., A. Viana da Fonseca, C. Ferreira, F. Sousa, & L. A. Bulla-Cruz (2021). Defining the soil stratigraphy from seismic piezocone data: A clustering approach. *Engineering Geology 287*, 106111.

Mori, K. & K. Sakai (2016). The gp sampler: a new innovation in core sampling. pp. 99–124. Australian Geomechanics Society.

Robertson, P. K. (1990). Soil classification using the cone penetration test. *Canadian Geotechnical Journal 27*, 151–158.

Robertson, P. K. & K. L. Cabal (2010). Estimating soil unit weight from CPT. In *2nd International Symposium on Cone Penetration Testing*, Huntington Beach, CA, USA.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics 20*, 53–65.

Saldanha, A. S., A. Viana da Fonseca, & C. Ferreira (2018). Microzonation of the liquefaction susceptibility: case study in the lower Tagus valley. *Geotecnia 142*, 07–34.

Viana da Fonseca, A., C. Ferreira, F. Molina-Gómez, & C. Ramos (2019). Collection of high-quality samples in liquefiable soils using new sampling techniques. pp. paper 014.

Viana da Fonseca, A., C. Ferreira, C. Ramos, & F. Molina-Gómez (2019). The geotechnical test site in the greater lisbon area for liquefaction characterisation and sample quality control of cohesionless soils. *AIMS Geosciences 5*, 325–343.

Viana da Fonseca, A. & J. Pineda (2017). Getting high-quality samples in 'sensitive' soils for advanced laboratory tests. *Innovative Infrastructure Solutions 2*, 34.