

# Semi-automated probabilistic soil profiling using CPTu

Stefano Collico

*Igeotest s.l. Figueres, Spain, [stefano.collico@upc.edu](mailto:stefano.collico@upc.edu)*

Marcos Arroyo

*Department of Civil and Environmental Engineering (DECA), Universidad Politecnica de Cataluña, Barcelona, Spain, [Marcos.Arroyo@upc.edu](mailto:Marcos.Arroyo@upc.edu)*

Marcelo Devincenzi

*Igeotest s.l. Figueres, Spain, [marcelo@igeotest.com](mailto:marcelo@igeotest.com)*

Alejandro Rodriguez

*Igeotest s.l. Figueres, Spain, [alejandror@igeotest.com](mailto:alejandror@igeotest.com)*

Amadeu Deu

*Igeotest s.l. Figueres, Spain, [amadeu@igeotest.com](mailto:amadeu@igeotest.com)*

**ABSTRACT:** Geotechnical soil profiling of a site is a key step in geotechnical design. Cone Penetration Test (i.e. CPTu) has become a widely used in-situ method for soil classification based on characteristic responses to the cone. Nevertheless, CPTu-based profile delineations inevitably require engineering interpretation, as the CPTu record is affected both by inherent variability and measurements errors. A semi-automated CPTu data interpretation tool is here presented aiming to simplify the soil delineation task while eliciting the assumptions and heuristics that designers apply in this process. The tool fits bivariate normal distributions to interpreted CPTu data. Classification makes use of conventional class boundaries –here taken from Robertson 1990 chart–, which are applied in sequence with user-specified refinement. A first level soil delineation is based on sand-like and clay-like behaviour. A layer zooming is then provided by an intermediate soil delineation, which differentiates purely sand-like and clay-like classes from soil mixtures. Finally, a maximum layer refinement is given by a second level soil delineation, which considers different classes within sand-like and clay-like ones. Depending on the delineation level selected, thin layers are consolidated to adjacent ones through different approaches. The proposed algorithm is tested on CPTu sounding records performed at Barcelona harbour (Spain). The simplified approach proposed can be easily adapted to different parametric charts and/or for multiple CPTu records.

**Keywords:** Cone Penetration Tests, probabilistic soil delineation, bivariate normal distribution, inherent variability.

## 1. Introduction

A fundamental step in geotechnical design is soil profiling, which essentially involves ground layer identification. In the last decades Cone Penetration tests (i.e. CPT, CPTu) have become the method of choice for soil profiling due to their reasonable costs, speed and repeatability. CPTu based profiling uses measured test responses to classify soils. This is most frequently done using charts in which each CPTu measurement may be represented in one (or various) two-dimensional diagrams where empirically pre-established boundaries separate different soil types [1, 2, 3, 4, 5, 6]. Inspecting the profile of soil class assignment, the analyst establishes layer boundaries and separates ground units.

Given the uncertainties inherent to soil deposits and unavoidable measurement errors, this is a difficult process. For instance, it is frequent that very thin layers of alternating classes result from the test output. Simplifying that raw profile is necessary, but prone to subjective decisions that may cause misunderstandings and/or disputes between the different parties intervening in a design. To minimize those problems the process of CPTu based soil delineation should aim for transparent, easy-to-understand and robust algorithms.

In the geotechnical literature several probabilistic-based automated delineation approaches have been developed, rigorously addressing measurement uncertainties and inherent variability of soil properties [7, 8, 9, 10]. These works question the validity of pre-established classification charts and include the classification boundaries and/or the classifying variables themselves as part of the characterization problem. Although these methods have been shown effective, their use in engineering practice is still limited, partly because they rely on relatively abstract probability and statistics concepts and partly because they are difficult to relate with established practices.

The work presented here is based on a more modest approach, suitable for practical use, in which pre-existing CPTu result classification charts are accepted as a given. Efforts are aimed instead at simplifying and clarifying the use of those pre-established charts in soil profiling for engineering design. The methodology applied has three core elements: a probabilistic description of CPTu classification data; a sequential approach to classification in which the level of class refinement is selected by the user, and an option for automated thin layer consolidation – similar to that given in [11]. In what follows the new methodology is briefly described and then showcased with some real examples.

## 2. Methodology

### 2.1. Robertson 1990 SBT classification

A frequently used Soil Behaviour Type (SBT) chart for soils is that proposed in [1], Fig. 1. It uses two normalized CPTu parameters (normalised cone tip resistance  $Q_{t1} = (q_t - \sigma_{v0})/\sigma'_{v0}$  and friction ratio,  $FR = 100 \cdot f_s/(q_t - \sigma_{v0})$ ) as classifiers. A total of 9 different classes were defined in the original chart. The correspondence between SBT chart CPTu data assignment and similar soil textural descriptions of nearby samples is not perfect. This may be interpreted as uncertainty in class boundary locations in the chart and treated as adding measurement error to the CPTu classification data [12, 13, 14]. However, it is questionable if textural classifications are better adapted to engineering design than classifications based purely on mechanical responses to the test and, in newer classification charts [3] textural class descriptions are abandoned. Robertson 1990 chart is thus taken here as a suitable example of conventional tool for soil classification and its class boundaries are accepted as given.

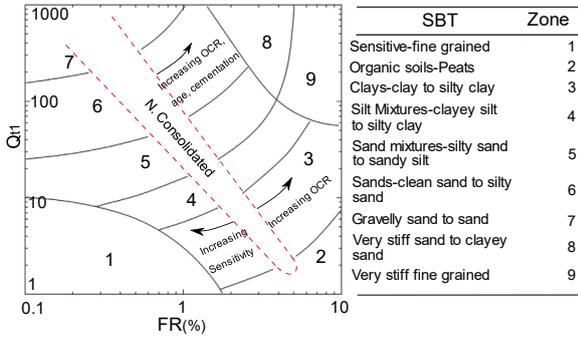


Figure 1. Original Robertson 1990 chart.

### 2.2. Probabilistic description of classing data

When a set of CPTu datapoints are plotted in a classification chart, such as Robertson 1990, a certain scatter is always observed. As an example, Fig. 2 reports cone tip resistance and sleeve friction profiles for a 1 m stretch of a CPTu, along with the corresponding data plot in the Robertson chart. These data clouds can be used to fit different statistical models. Previous work [17, 9] has shown that bivariate normal distributions are useful to fit this kind of datasets and this is the approach followed here.

Because the classification chart is double-logarithmic the relevant expression to fit the data (e.g.  $\ln(Q_{t1})$ ,  $\ln(FR)$ ) may be written as:

$$f_{(\ln(Q_{t1}), \ln(FR_i))} = \frac{1}{2\pi\sigma_{\ln(Q_{t1})}\sigma_{\ln(FR)}\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left[\frac{(\ln(Q_{t1}) - \mu_{\ln(Q_{t1})})^2}{\sigma_{\ln(Q_{t1})}^2} + \frac{(FR_i - \mu_{FR})^2}{\sigma_{\ln(FR)}^2} - \frac{2\rho(\ln(Q_{t1}) - \mu_{\ln(Q_{t1})))(\ln(FR_i) - \mu_{\ln(FR)})}{\sigma_{\ln(Q_{t1})}\sigma_{\ln(FR)}}\right]\right) \quad (1)$$

with the following parameters:

$\mu_{\ln(Q_{t1})}$ ,  $\mu_{\ln(FR)}$  mean value of the logarithm of  $Q_{t1}$  and  $FR$  data respectively;  
 $\sigma_{\ln(Q_{t1})}^2$ ,  $\sigma_{\ln(FR)}^2$  = variance of the logarithm of  $Q_{t1}$  and  $FR$  data respectively;  
 $\rho$  = correlation coefficient for  $\ln(Q_{t1})$  and  $\ln(FR)$ .

An example of the fitted bivariate normal distribution is shown in (Fig. 2b). Contours are also shown corresponding to different probability percentiles (0.1, 0.2, ...). The relation of these percentile contours and the class boundary limits that divide the classification chart will be now exploited to identify soil layers.

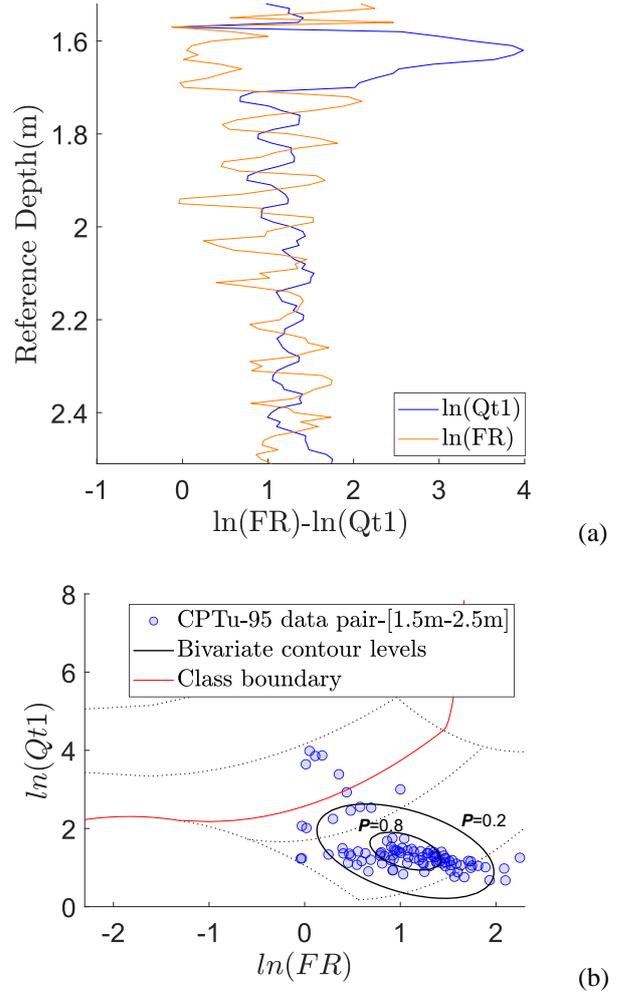


Figure 2. a) CPTu sounding record profiles for 1m stretch. b) Corresponding bivariate distribution contour levels within Robertson chart.

### 2.3. Class assignment criteria

Once a classification dataset has been fitted by a bivariate distribution the decision about which class is assigned to all its underlying data is based only on how the distribution relates to the class boundaries.

A bivariate distribution will be assigned to a certain soil class if it is contained within its boundaries. However, for practical reasons, that containment is not absolute, but partial. Class assignment is nuanced by two criteria expressed through pre-established acceptance thresholds. These thresholds are described by two user-defined parameters, namely:

- noise-threshold parameter,  $P$
- class-tolerance parameter,  $m$ .

The numerical value of  $P$  specifies the complement of the cut-off probability value for the adjusted layer bivariate normal density function. For instance, if  $P = 0.1$  only the data within the level curve of the fitted bivariate corresponding to a probability density of 90% has to be contained within the class boundaries.

The noise threshold  $P$  parameter is introduced to account for the presence of extreme values of classifying parameters within the dataset. Those extreme values might be considered as originating from data noise (e.g. shells within a fine layer; instrument malfunction). The analyst should therefore judge what is the likely level of noise in the data.

The class-tolerance threshold parameter  $m$  is introduced to account for the presence of layers straddling various classes. The value of  $m$  is defined as the area proportion of the level curve corresponding to  $P$  that is accepted to lie across the chart boundaries. The value selected for  $m$  reflects the degree of adherence to the pre-established class boundaries by the analyst.

The two thresholds introduced are inspired by common interpretation heuristics applied by CPTu data analysts. The selection of such threshold provides the first step within the proposed semi-automated approach.

## 2.4. Staged classification refinement

The original Robertson (1990) chart is subdivided in nine soil classes. In many applications such refinement may not be required. Therefore the method introduces a three-staged classification procedure in which the level of refinement applied in classification is left as an option to the user.

The coarser level of classification operates with only two soil classes:

1. Clay-like soil behavior (C-L);
2. Sand-like soil behavior (S-L).

C-L class represents an undrained soil response [3] merging the original Robertson SBT zones 2-3-4-9. S-L is representative of a drained response to CPTu and merges the original SBT 5-6-7 and 8. The boundary line that identifies these two classes is reported in Fig. 3. Soil layers identified by these classes are here onwards named as primary layers.

The second level of classification operates with three classes, aiming to differentiate units into soil mixtures, purely sand-like and purely clay-like. This is inspired by work [3, 18] highlighting how soil mixtures (e.g. silts with low plasticity, sometimes referred to as transitional soils) are characterized by a partially drained response. The original SBT classes assigned to each of this second classes are described in Table 1 and indicated in Fig. 3. Finally, the highest level of classification refinement is obtained by including all the Robertson 's SBT lines. Such classes are referred here as tertiary classes.

Table 1 Soil groups for second level class refinement

Soil group	SBT Robertson (1990)
Clay	1-2-3 part of zone 9
Soil-Mixtures	4-5 part of zone 8 and 9
Sand	6-7- part of zone 8

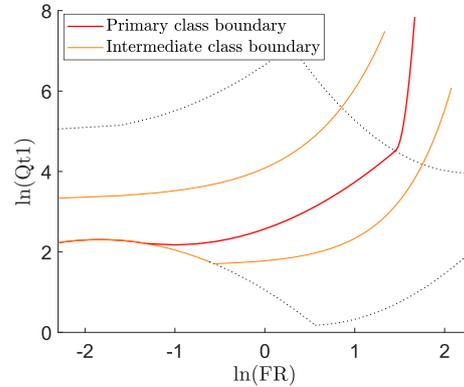


Figure 3. Class boundary at first and intermediate staged classification on Robertson 1990 chart.

Even if a highly refined classification level is finally desired, it seems advantageous to initially perform classification at lower refinement levels to gain understanding of the profile. Subsequently, a more detailed layer delineation can be obtained adopting the second and/or a third level classification.

For the second and third level delineation, in case that the bivariate distribution simultaneously crosses two different SBT boundaries, two area ratios are computed and compared with the  $m$  threshold selected.

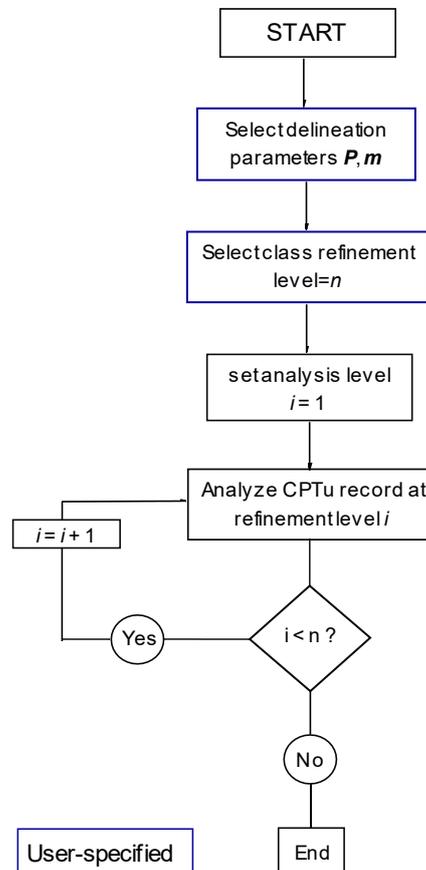


Figure 4. General code workflow

## 2.5. Code workflow

The method proposed can be summarized as follow (Fig. 4). First, the analyst specifies the level of classification refinement desired and values for the two thresholds  $P$  and  $m$ . The classification refinement level to apply is then specified and the code proceeds to analyze the CPTu input record.

When analysing a CPTu record (Fig. 5) the code starts by selecting a segment of CPT data pairs long enough (e.g. 15 points) to fit a bivariate normal density distribution. Once initialized, the bivariate density distribution is updated with each subsequent CPTu data pair, moving downwards through the record. As new data is added the bivariate distribution of the layer moves on the classification plane. This motion goes on until class boundaries are exceeded by the percentile contour identified by  $P$  with tolerance given by threshold  $m$ . At that moment the data pairs represented by the bivariate are all assigned to a layer. The layer class is given by the location of the mean value of the bivariate distribution. The classification resumes with the next segment in the CPTu record, until this is fully analysed (Fig. 5).

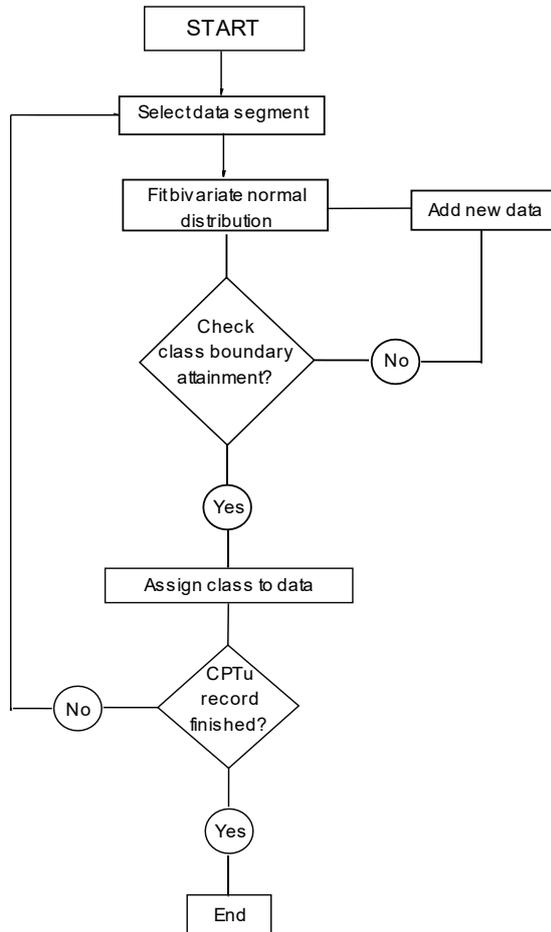


Figure 5. Analysis of CPTu record at a given refinement level

To clarify the procedure a synthetic example is illustrated in Fig. 6. The analysis takes place at the first refinement level and the threshold values are  $P = 0.1$  and  $m=0.1$ .

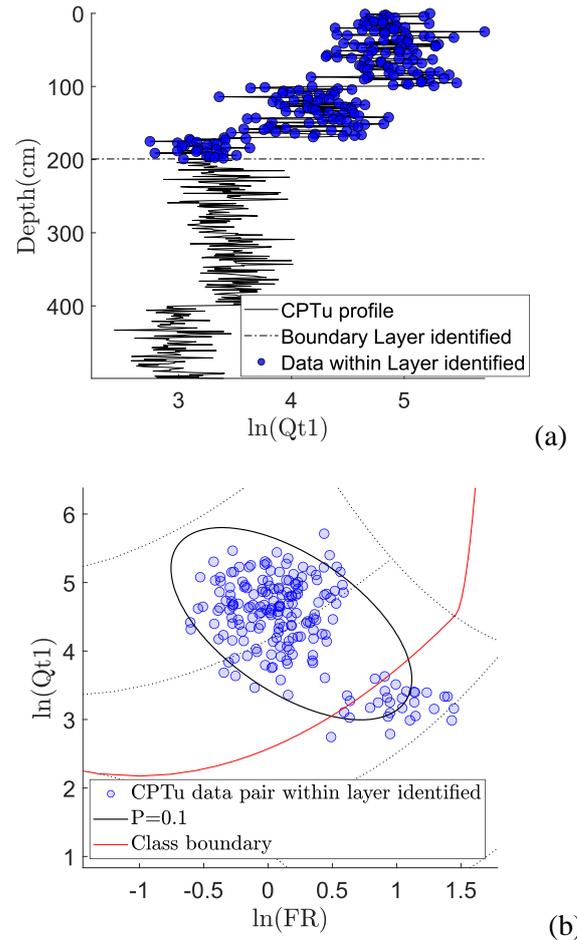


Figure 6. a) Random generate soil unit CPTu sounding record and identified boundary layer for  $P=0.1$  and  $m=0.1$ . b) Corresponding code frame for the identified layer.

## 3. Thin layer treatment

Whatever the level of classification refinement specified, once the CPTu record has been analysed, thin layers (e.g. thinner than 20 cm) might be present.

It is well accepted that CPTu data acquisition is faster than mechanical layer thickness resolution, which is a function of the cone diameter employed [15] – recall that the standard dimension is 35.7mm. The measured cone tip resistance is affected by the so-called development and sensing distances. The cone has to penetrate up to a certain depth within a soil unit to fully reflect its response: that is the development distance [16]. The sensing distance intervenes as the cone approaches a soil layer of different stiffness. In a such a case, the cone start senses its presence a few cone diameters ahead of the interface between layers [16]. Due to those two limits, the measured soil resistance does not coincide with the “true” one, which is referred to the measured soil resistance in absence of both development and sensing distances.

Therefore, the cone is only able to unambiguously identify layers above a minimum layer resolution of 150-200mm. A simplified practical approach to deal with this difficulty is to merge or “consolidate” layers below this minimum with adjacent ones of similar soil behaviour

type. The procedure proposed by Ganju et al. [11] has been applied here.

After the record is analysed layers identified as thin layers are merged. The procedure depends slightly on the classification level selected. At the first and second refinement levels thin layers are merged based on the closeness on mean value of  $\ln(Q_{t1})$ . At the third level, when using the full Robertson 1990 classification chart, the situation is identical to that of Ganju et al. [11] and the same auxiliary criteria based on class groupings are applied.

#### 4. Illustrative example

##### 4.1. Case study

As an illustration the procedure described is tested on CPTu records obtained at Barcelona harbour (Spain). Two adjacent CPTu soundings records (CPTu95, CPTu94), each up to 20m depth from seabed surface Fig. 7 are considered.

An independently established soil layer profile was available for the site, derived from cores retrieved at the same location where CPTu95 was performed. Laboratory samples were retrieved each 5m depth and core description followed UNE-EN ISO 14688-1.

This core-based layer delineation also took place at two levels. An initial simplified delineation was first assumed Fig. 8a. After the laboratory results were available a more detailed classification was attempted Fig. 8b. Layer delineation at this site was difficult, particularly up to 11.5 m depth, with samples recovered suggesting large heterogeneity. The limits chosen carried significant uncertainty. In several sections, e.g. within 4-6.5m and 7.95-11m depth the complex structure did not allow a clear separation of layers.

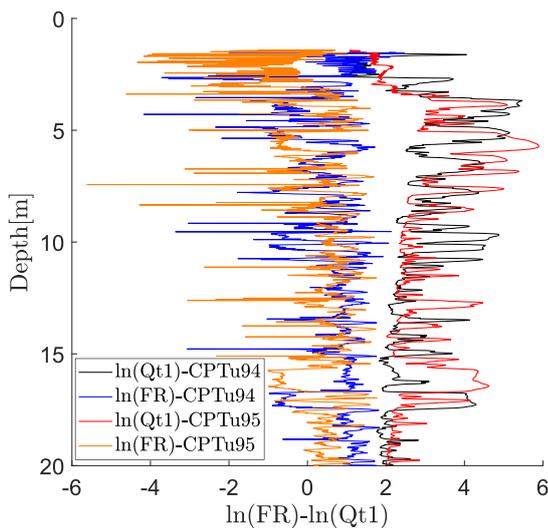


Figure 7. CPTu record for the two adjacent locations considered.

The layering of Fig. 8a was directly checked against cone results. Each of the CPTu95 data pairs was assigned into the Sand-Like or the Clay-Like classes. The data representative of S-L behaviour and C-L behaviour was then plotted on the  $Q_{t1}$  profile providing a general overview of the SBT at the site (Fig. 7). Percentages of S-L and C-

L data within each layer identified in the profile of Fig. 8a are reported in Table 2. From this analysis the primary layers identified by core inspection appear plausible. To go beyond this level of analysis the new tool was necessary.

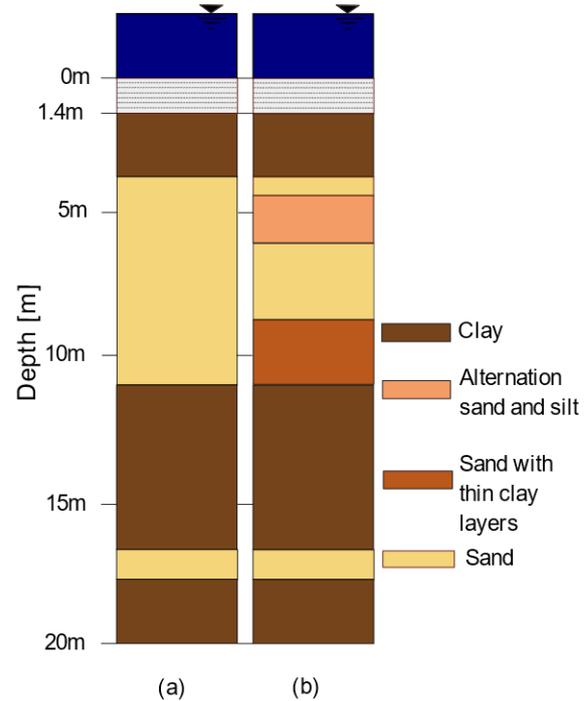


Figure 8. Soil delineation at the site of CPTu95 derived from core inspection and lab results.

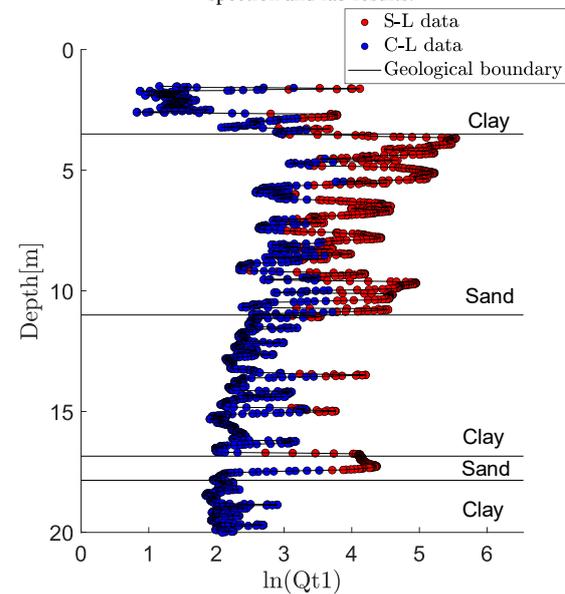


Figure 9. S-L, C-L behaviour within each layer in Figure 8a.

Table 2. % of Sand-Like and Clay-Like data within each geological unit.

Geological primary layers depth [m]	% S-Like data	% C-Like data
1.4-3.5	17	83
3.5-11	65.4	34.60
11-16.85	8	92
16.85-17.85	68.8	31.2
17.85-20	0	100
Total	48.75%	51.25

## 4.2. CPTu-aided soil delineation

The semi-automated procedure previously described was applied to the CPTu 95 record. The  $P$  and  $m$  model parameters were varied until a good match was obtained in primary layer delineation with that available from the core. Threshold values of  $P = 0.8$  and  $m=0$  were able to select boundaries (Fig. 10a) at the same locations of the core-based layering.

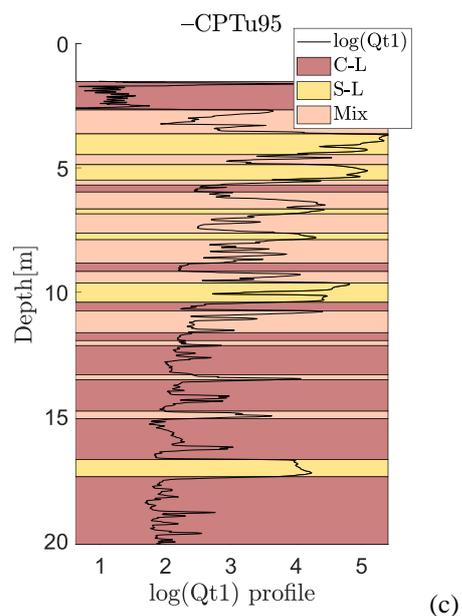
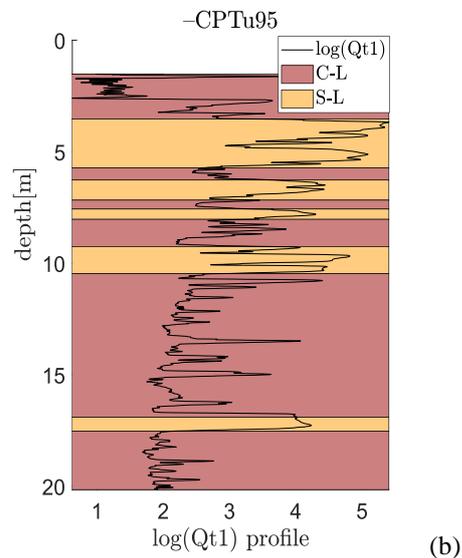
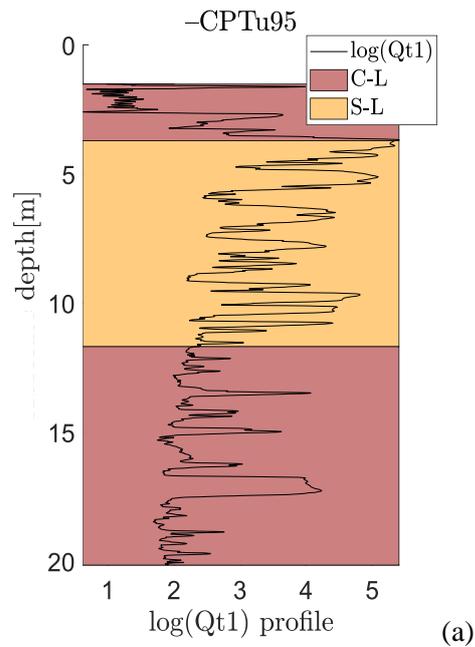
However, such a large  $P$  value indicates a noise cut-off close to the mean of the bivariate distribution. This was considered too high, suggesting –correctly– that the reference soil delineation in Fig. 8a was too blunt. Reducing the threshold value to  $P = 0.4$  and maintaining  $m=0$  was enough to closely approximate the more refined core-based soil delineation of Fig 8b (Fig. 10b). Several primary layers are now identified within 4-11m depth.

The same  $P$  and  $m$  values (i.e.  $P = 0.4$  and  $m = 0$ ) were then employed in more refined classifications. Results at the second level (Fig. 10c) show how transitional soils are dominant up to 10.5m depth, whereas after 10.5m first and second level delineations agree well.

Finally, the highest-class refinement, incorporating all the Robertson SBT lines is presented in (Fig 10d). The profile obtained at this third level of refinement is similar to that obtained with a representative commercial CPTu data treatment software (Datgel) (Fig. 10e). Results in Fig. 10d appear simpler because –as those of the coarser classification stages– they have been already corrected for thin layers. Indeed, in the interlayered zone (e.g. 4-11m) at each stage classification, several thin layers (thickness lower than 200 mm) were identified.

Examining the results of the three levels together allows the analyst to select the level of detail that is more meaningful for engineering design process. In this case this would be at level 1 or 2, depending on the application (preloading or pile design, say).

The previous analysis may be interpreted as eliciting the implicit analyst criteria (i.e. values of  $P$  and  $m$ ). Once these implicit thresholds have been identified they may be applied systematically to other investigation points at the site. As an example, the same  $P$  and  $m$  model parameters (i.e.  $P=0.4$ ,  $m=0$ ) are then employed for the second CPTu sounding record (CPTu94). Results in terms of first, second and third level soil delineation are reported in Fig 11a, Fig. 11b and Fig. 11c.



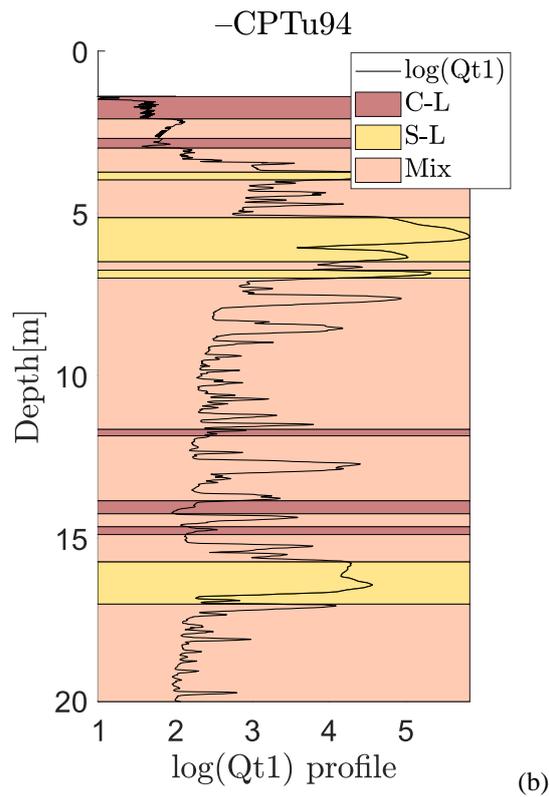
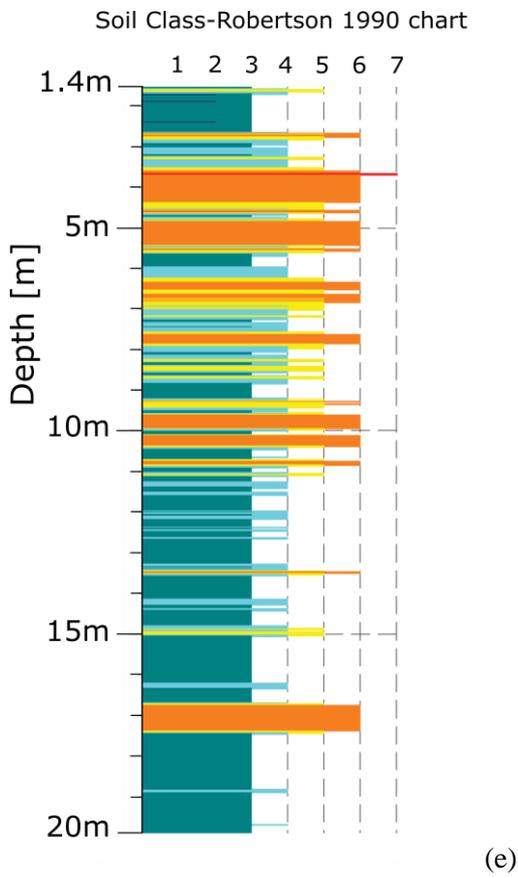
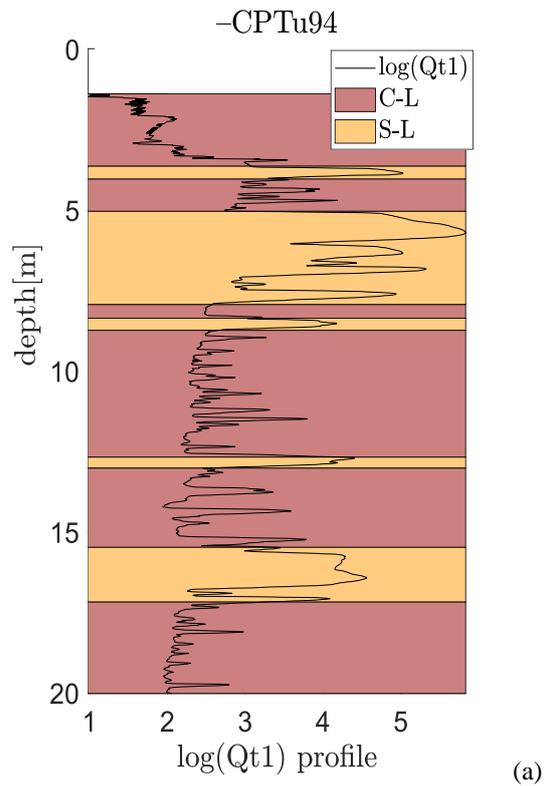
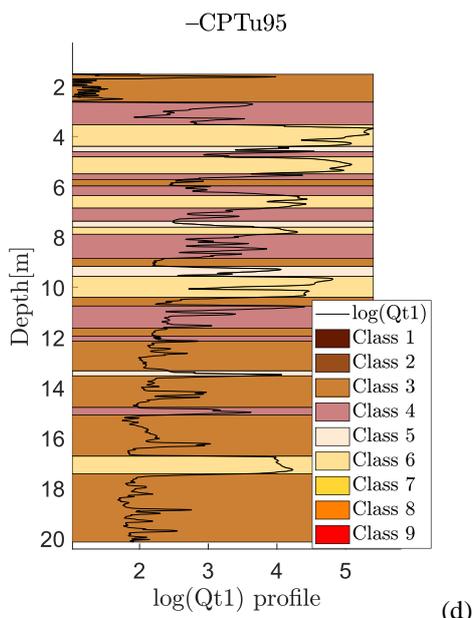


Figure 10. Results of CPTu95-based soil delineation for  $P=0.4$  and  $m=0$  at different scale: a) First level soil delineation. b) Second level soil delineation. c) Third level soil delineation.

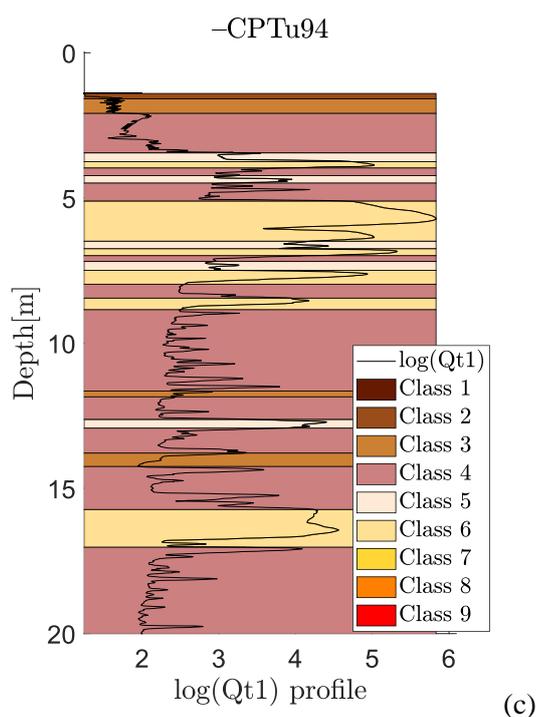


Figure 11. Results of CPTu94-based soil delineation for  $P=0.4$  and  $m=0$  at different scale: a) First level soil delineation. b) Second level soil delineation. c) Third level soil delineation.

## 5. Conclusion

This work describes a novel procedure for soil delineation based on CPTu sounding records aiming to reduce uncertainties and facilitate communication in the process of layer delineation. The procedure assumes that analysts apply individual heuristics in CPTu record interpretation and offers a tool to elicit those assumptions and thus facilitate consensus.

Apart from that benefit, the methodology proposed has several development possibilities worth mentioning. For instance, it provides a measure of soil variability at every identified layer (through the corresponding bivariate parameters). It can be easily extended to different data representation (i.e. univariate approach, based on Soil Type Behaviour Index  $I_C$ ) and to different SBT charts. Furthermore, the same procedure can be easily extended to simultaneous analyses of multiple CPTu records, for instance testing the assumption of same soil class appearing within a particular depth interval.

## Acknowledgement

The first author's work has been supported by the funding from the European Union's Horizon 2020 research and innovation programme under the Marie-Sklodowska-Curie grant agreement No 721403.

## References

[1] Robertson, P. K. "Soil classification using the cone penetration test". Canadian Geotechnical Journal, 27(1), pp. 151-158, 1990. <https://doi.org/10.1139/t90-014>  
[2] Robertson P. K. "Interpretation of cone penetration tests—a unified approach", Canadian Geotechnical Journal (46), pp. 1337-1355, 2009.

<https://doi.org/10.1139/T09-065>

[3] Robertson P. K. "Cone penetration test (CPT)-based soil behaviour type (SBT) classification system — an update", Canadian Geotechnical Journal (00), pp. 1–18 (2016). <https://doi.org/10.1139/cgj-2016-0044>  
[4] Schneider, J.A., Randolph, M.F., Mayne, P.W., and Ramsey, N.R. "Analysis of factors influencing soil classification using normalized piezocone tip resistance and pore pressure parameters". Journal of Geotechnical and Geoenvironmental Engineering, ASCE, 134(11): pp. 1569–1586, 2008. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2008\)134:11\(1569\)](https://doi.org/10.1061/(ASCE)1090-0241(2008)134:11(1569))  
[5] Schneider, J.A., Hotstream, J.N., Mayne, P.W., and Randolph, M.F. "Comparing CPTU Q-F and  $Q-u_2/\sigma'_{v0}$  soil classification charts", Geotechnique Letters, 2 (4), pp. 209–215, 2012. <https://doi.org/10.1680/geolett.12.00044>  
[6] Cetin K., O., Ozan C. "CPT-Based Probabilistic Soil Characterization and Classification", Journal of Geotechnical Geoenvironmental Engineering, 135(1), pp. 84-107, 2009. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2009\)135:1\(84\)](https://doi.org/10.1061/(ASCE)1090-0241(2009)135:1(84))  
[7] Cao, Z., & Wang, Y. "Bayesian approach for Probabilistic Site characterization Using Cone Penetration Testing", Journal of Geotechnical and Geoenvironmental Engineering, 139 (2), pp. 267-276, 2013. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0000765](https://doi.org/10.1061/(ASCE)GT.1943-5606.0000765)  
[8] Wang, Y., Huang, K., & Cao, Z. "Probabilistic identification of underground soil stratification using cone penetration tests". Canadian Geotechnical Journal, 50 (7), pp. 766-776, 2013. <https://doi.org/10.1139/cgj-2013-0004>  
[9] Depina, I., Le, T. M., Eiksund, G., & Strøm, P. "Cone penetration data classification with Bayesian Mixture Analysis", GEORISK, 10, pp. 27-41, 2016. <https://doi.org/10.1080/17499518.2015.1072637>  
[10] Ching J., Wang J. "Application of the transitional Markov chain Monte Carlo algorithm to probabilistic site characterization, Engineering Geology 203, pp. 151–167, 2016. <https://doi.org/10.1016/j.enggeo.2015.10.015>  
[11] Ganju, E., Prezzi, M., & Salgado, R. "Algorithm for generation of stratigraphic profiles using cone penetration", Computers and Geotechnics, 90, pp. 73-84, 2017. <https://doi.org/10.1016/j.compgeo.2017.04.010>  
[12] Douglas B. J. & Olsen R. S. 'Soil classification using Electric cone penetrometer', in Symposium on Cone Penetration Testing and Experience, Geotechnical Engineering Division, ASCE. St. Louis, Missouri, USA, 1981, pp. 209–227.  
[13] Uzielli, M., Vannucchi, G., & Phoon, K. K. "Random field characterisation of stress normalised cone penetration testing parameters". Geotechnique, 55(1), pp. 3-20, 2005. <https://doi.org/10.1680/geot.55.1.3.58591>  
[14] Jung, B., Gardoni, P., and Biscontin, G. "Probabilistic soil identification based on cone penetration test", Geotechnique 58(7), pp- 591-603, 2008. <https://doi.org/10.1680/geot.2008.58.7.591>  
[15] Lunne, T., Powell, J., & Robertson, P. K. "Cone Penetration Testing in Geotechnical practice", Blackie Academic & Professional, London, England, 1997.  
[16] Boulanger, R. W., & DeJong, J. T. "Inverse filtering procedure to correct cone penetration data for thin layer and transition effects", CPT'18 4th International symposium on Cone Penetration Testing, Delft, The Netherlands, 2018, pp. 25-44, 2018.  
[17] Wang, X., Wang, H., & Liang, R. Y. "A hidden Markov random field model-based approach for probabilistic site characterization using multiple cone penetration test data". Structural Safety 70, pp. 128–138., 2018. <https://doi.org/10.1016/j.strusafe.2017.10.011>