

INTERNATIONAL SOCIETY FOR SOIL MECHANICS AND GEOTECHNICAL ENGINEERING



This paper was downloaded from the Online Library of the International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE). The library is available here:

<https://www.issmge.org/publications/online-library>

This is an open-access database that archives thousands of papers published under the Auspices of the ISSMGE and maintained by the Innovation and Development Committee of ISSMGE.

The paper was published in the proceedings of the 10th European Conference on Numerical Methods in Geotechnical Engineering and was edited by Lidija Zdravkovic, Stavroula Kontoe, Aikaterini Tsiampousi and David Taborda. The conference was held from June 26th to June 28th 2023 at the Imperial College London, United Kingdom.

To see the complete list of papers in the proceedings visit the link below:

<https://issmge.org/files/NUMGE2023-Preface.pdf>

Automated CPT interpretation and modelling in a BIM/Digital Twin environment

R.B.J. Brinkgreve¹, F. Tschuchnigg², A. Laera³, S. Brasile³

¹*Department of Civil Engineering & GeoSciences, Delft University of Technology, Delft, NL*

²*Graz University of Technology, Graz, Austria*

³*Seequent – The Bentley Subsurface Company, Italy / Netherlands*

ABSTRACT: Following up on previous research on Automated Parameter Determination (APD), in which the soil stratification and numerical model parameters are automatically derived from individual CPTs, this article describes ongoing research in which the geotechnical modelling workflow is further automated in a BIM / Digital Twin environment. Especially in a preliminary project phase, when limited soil data are available, a workflow in which CPT data are used to automatically create a 3D geological model from which 2D or 3D numerical models can be extracted, may be very helpful in exploring different design alternatives. For existing (infrastructural) projects, such an automated system in a Digital Twin environment could also help responsible authorities to check the infrastructure's safety under changing conditions. In addition to the description of technical solutions used for automatic layer detection and clustering (based on Machine Learning) across different CPTs, the article touches upon the discussion on transparency and accessibility of the automated system in view of the expertise and responsibilities of the operating geotechnical engineer.

Keywords: CPT; automation; geological modelling; clustering; machine learning (ML)

1 INTRODUCTION

The Cone Penetration Test (CPT) is a very useful and cheap in-situ test to investigate ground properties for geotechnical engineering in medium to soft soils. Some countries maintain accessible databases with CPT and other ground data. Especially in an early project phase, when limited soil data are available, CPTs can provide a valuable source of information from which geotechnical design alternatives may be analysed and evaluated. No doubt, for detailed geotechnical design more accurate soil investigations based on high quality lab testing are needed. In any case, CPTs (in addition to borings) are very useful in characterising the (3D) sub-soil stratigraphy to a depth of about 30 m (and even beyond), which is relevant for any infrastructural construction project.

Although the processing of CPT data has been automated to a certain extent, further automation steps can be taken to make geotechnical modelling for engineering and design purposes more efficient and reliable, especially in digital cloud-based modelling concepts such as Building Information Modelling (BIM) and Digital Twins.

In previous publications by the authors and co-workers it was described how CPT data can be used to automatically determine soil properties as well as model parameters of constitutive models based on paths of

correlations (Brinkgreve, 2019; Van Berkomp et al., 2022; Brinkgreve & Brasile, 2022). The so-called Automated Parameter Determination (APD) concept has been demonstrated in conjunction with the interpretation of a single CPT, but it is not limited to that. In fact, APD can be used more generally on averaged data from corresponding soil layers across different CPTs at a project site; in other words: from all CPT readings contributing to the same 3D soil layer in the underground. Therefore, our research has branched into automatic layer determination, which is the topic of this publication.

As a next step following the identification of soil layers in a *single* CPT, automatic identification of *3D soil layers* composing a 3D sub-soil model is more challenging. After a short review of previous work on automatic layer determination, Section 2 first focuses on an alternative method for detecting layers from single CPTs. Section 3 then continues with the automatic clustering of sub-layers across multiple CPTs at a project location to compose 3D soil layers. Examples are included to demonstrate the working of automatic layer detection and clustering. The clustering is used as a pre-processor to create borehole input data as used in geological modelling software to create a 3D sub-soil model; this is described in Section 4. The 3D sub-soil model can be used subsequently in a BIM / Digital Twin environment to extract 2D or 3D models for numerical analysis.

Although automation improves the workflow efficiency and reduces errors due to re-entering of data from one application to another, there is some concern that too much automation may harm the geotechnical engineering profession. Therefore, Section 5 discusses the issue of transparency and accessibility of automated systems in view of the expertise and responsibilities of the geotechnical engineer. The article ends with some conclusions and references for further reading.

2 LAYER DETECTION IN SINGLE CPT

Most CPT interpretations are based on Robertson's method, translating combinations of cone resistance q_c and sleeve friction f_s (or friction ratio R_f) into Soil Behaviour Type *SBT* (Robertson 1990, 2009, 2010, 2016) for every CPT reading (typically every 1, 2 or 5 cm in depth). In geotechnical engineering, the decomposition of individual CPTs into a limited number of soil layers with averaged properties, is common practice. Since this process is generally based on 'engineering judgement', interpretations may vary from one engineer to another. Methods have been proposed to automate this process and obtain more objective and consistent results. Most methods use the *SBT* as a starting point to compose layers and impose a minimum layer thickness d_{min} or desired number of layers as 'a priori' condition. The more recent methods are based on a probabilistic or Bayesian approach addressing the uncertainty and spatial distribution of CPT data (Wang et al., 2013; De Zeeuw, 2021). The disadvantage of probabilistic methods is that these are relatively time-consuming and may not be efficient in practical projects when multiple CPTs need to be processed.

In this article an alternative approach is followed. The idea is that the transition from one layer to another results in an increase of the combined standard deviations in q_c (or the corrected cone resistance q_t or normalised cone resistance Q_m) and R_f (or normalised friction ratio F_r) considering the moving average and moving standard deviation over a certain number of readings $n_{average}$ (e.g. 7) around each data point, to reduce fluctuations within each layer. The result is a plot of combined standard deviations as a function of depth, in which the 'peaks' indicate potential layer boundaries.

After the maximum peak has been identified, all values within a distance d_{min} are removed to ensure that subsequent layer boundaries respect the minimum layer thickness. Then, the next maximum peak is identified and treated, etcetera.

It can happen that the two layers at both sides of a peak are very similar. In that case the peak can be ignored (deleted) from the list. To check if adjacent layers are similar, average values of $\log(q_t/p_a)$ and $\log(R_f)$ are calculated for these layers (p_a = atmospheric pressure) and the distance of their respective positions in Robertson's diagram is measured. Note that the averaging and

distance calculation involves log-values rather than actual values of q_t/p_a and R_f . If the distance is less than a certain *threshold* value (typically in the range [0.1, 0.3]) the peak does not represent a true layer boundary and can be deleted from the list. The higher the *threshold* value, the more peaks are deleted resulting in a smaller number of layers with a larger layer thickness. However, this last step may not be necessary as explained in Section 3.

The procedure is pointwise listed below:

- Store all CPT readings and derived CPT parameters with depth (*depth*, q_c , q_t , f_s , R_f , etc.)
- Calculate moving averages and moving standard deviations of $\log(q_t/p_a)$ and $\log(R_f)$ over a range of $n_{average}$ readings.
- For the desired maximum number of layers:
 - Search for the highest standard deviation ('peak')
 - Store the corresponding depth in a list
 - Remove all standard deviations within a distance d_{min} around this level
 - Search the next peak in standard deviation
- Optionally: For all peaks (depths) in the list:
 - Calculate average values of $\log(q_t/p_a)$ and $\log(R_f)$ of adjacent layers
 - If the distance of the respective points ($\log(q_t/p_a)$, $\log(R_f)$) in Robertson's diagram of two adjacent layers is less than a certain *threshold* value, delete this peak and depth.
- Decompose the CPT into (sub-)layers based on the remaining peaks (depths) in the list.
- Calculate averaged CPT parameters for all (sub-)layers based on the contributing CPT readings.

Instead of the term 'layer', it is preferred to use the term 'sub-layer' when considering the decomposition of individual CPTs. This is because a sub-layer is part of a bigger entity, namely a 3D soil layer, as demonstrated in Section 3.

2.1 Layer detection example

CPT000000149495 as obtained in GEF format from the Dutch database Dinoloket (<https://www.dinoloket.nl>) is used to demonstrate the (sub-)layer detection procedure. Figure 1 (left) shows q_t and R_f as a function of depth. The ground surface is located at NAP -4.78 m, where NAP is the Dutch reference elevation level. The groundwater table is assumed 1.0 m below ground surface.

The standard deviation with depth, calculated as the sum of the moving standard deviation of $\log(q_t/p_a)$ and $\log(R_f)$ over a range of 7 readings (from 3 above until 3 below each data point) is plotted next to the CPT diagram in Figure 1 (middle).

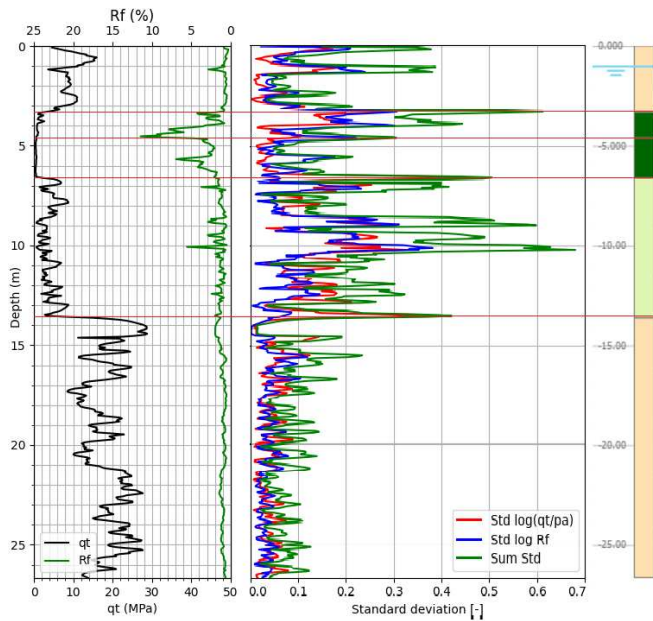


Figure 1. Left: q_t and R_f as function of depth for CPT00000149495 from a location in Rotterdam, NL. Source: Dinoloket.nl. Middle: Moving standard deviation of $\log(q_t/p_a)$ and $\log(R_f)$ and their sum as function of depth. The 'peaks' show potential layer boundaries. Right: Decomposition into limited number of soil layers (5).

Starting with the maximum peak, a total of 20 local peaks are identified. The minimum layer thickness d_{min} is taken as 1.0 m. Although multiple peaks and hence, multiple potential soil layer boundaries are initially identified, most of them are deleted (including the maximum peak at 10 m depth, which turns out to be no real layer boundary). This was achieved by using a rather high *threshold* value of 0.3. The result is only five different soil layers as shown in Figure 1 (right) consisting of a top sand layer ($SBT = 6$), followed by two subsequent clay layers ($SBT = 3$), a silty sand layer ($SBT = 5$) and a deep sand layer ($SBT = 6$). In fact, the upper one of the two adjacent 'clay' layers is actually peat, but organic soils are often not properly predicted using Robertson's method. Meanwhile, Lengkeek (2022) proposed improvements for the identification of organic soils, which will be included in future applications.

3 CLUSTERING OF LAYERS ACROSS MULTIPLE CPTS

At many locations the subsoil is not composed of fully horizontal soil layers, so the interpretation and decomposition of a single CPT or borehole into (sub-)layers is insufficient to compose a 2D or 3D sub-soil model for geotechnical engineering and analysis purposes. In general, the connection of corresponding soils from one borehole to another requires the expertise of a geotechnical engineer. The automation of this process is a challenge that several researchers have investigated, but as

to date there is no robust method that accurately determines 3D soil layers across boreholes.

The composition of 3D soil layers resembles the 3D stratification of geological formations in the deeper underground. Geological software such as Leapfrog can help building and visualizing a 3D sub-soil model based on borehole data and surfaces (triangulated or point clouds). Leapfrog was originally developed for deep mining applications but can also be used to create a 3D sub-soil model for geotechnical engineering applications. CPT data can be transformed into 'boreholes', optionally combined with ground surface elevation data to build a 3D sub-soil model. Automation of this process still requires some pre-processing to link similar soil (sub-)layers across boreholes.

Pre-processing of borehole data in terms of clustering of corresponding sub-layers, can be done by means of Machine Learning (ML). Using q_t/p_a and R_f only is not enough to identify corresponding soil layers across boreholes, since, for example, shallow sands could be incorrectly linked to deeper sands with similar CPT parameters. The solution is to add *depth* as an additional parameter.

After the interpretation and decomposition of a single CPT into sub-layers, the averaged $\log(q_t/p_a)$ and $\log(R_f)$ together with the averaged depth per sub-layer can be regarded as data points in a 3D version of Robertson's diagram, where the third dimension is *depth*. In this way, sub-layer data points from multiple CPTs at a project location can be visualized all together as 'dots' in a 3D diagram (Figure 2).

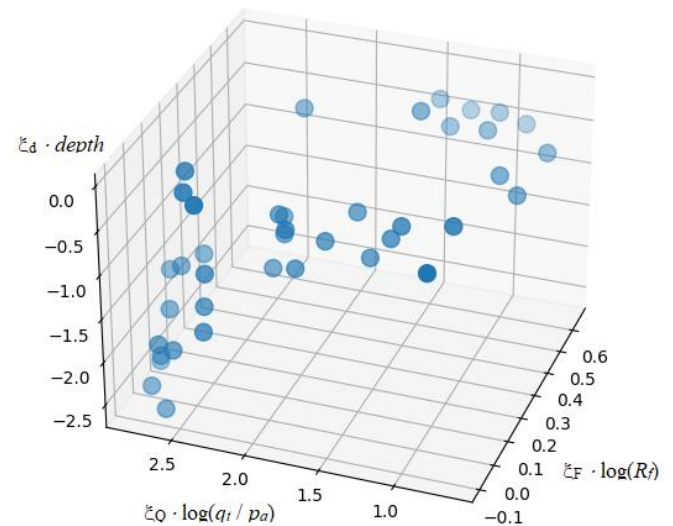


Figure 2. 3D visualization of data points

The 3D data points are then sent to a clustering algorithm to identify clusters of similar points. Thereby, scaling factors (ξ_Q , ξ_F , ξ_d) are used for all three quantities to ensure that all of them have a similar influence in the clustering process.

In the world of ML, clustering algorithms are found within the branch of unsupervised learning methods. This means, there are no ‘a priori’ categories defined to which a data point can belong, and there is no separate set of data points needed for training of the algorithm. Clustering algorithms search for clusters of points or point densities in a 2D or 3D environment. The result is a list with respective cluster number for each data point. Data points for which no cluster can be found are given the cluster number -1 : the ‘left-overs’.

There are different types of clustering algorithms, of which *K-Means* and *DBSCAN* are probably the ones mostly used. For our application we have chosen the *DBSCAN* algorithm (Ester et al., 1996), as available in the Python library *Scikit-learn*. Input to *DBSCAN* are the 3D data points composed of averaged and scaled CPT parameters ($\xi_Q \cdot \log(q_t / p_a)$, $\xi_F \cdot \log(R_f)$, $\xi_d \cdot \text{depth}$) for all sub-layers identified from a series of CPTs from a project site. Besides the scaling parameters, the algorithm requires a minimum number of data points in each cluster (n_{min}) and an accuracy parameter (eps). The lower eps , the stricter the similarity should be for data points to be in the same cluster. Hence, a smaller value of eps generally gives more clusters.

We have investigated the influence of these parameters and obtained good results using $\xi_Q = 1.0$, $\xi_F = 1.0$, $\xi_d = 0.1$, $n_{min} = 3$ and eps in the range $[0.15, 0.3]$. Figure 3 shows the clustering of the same data points as in Figure 2 using $eps = 0.26$.

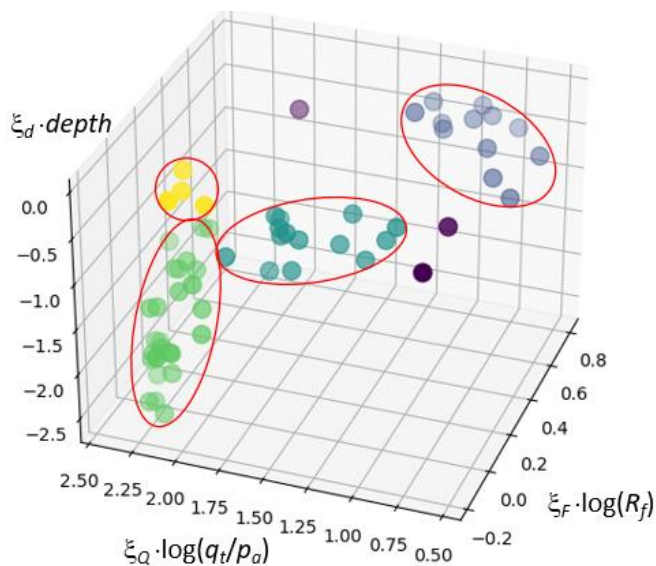


Figure 3. 3D visualization of data points after clustering

After clustering there are generally some data points for which no cluster can be found (the ‘left-overs’), indicated by the cluster number -1 . These are the purple dots in Figure 3. To ensure that the left-overs are somehow added to existing clusters, the distance of these points to the centre (average) of each cluster is calculated, and the points are added to their closest cluster.

In the end, a file is created with information of each sub-layer, including their corresponding cluster number, which can be used to compose a 3D sub-soil model.

Some remarks about the clustering procedure:

- Although the use of ML techniques may seem highly sophisticated, it only takes 3 lines of code to implement the clustering algorithm in Python (when importing the *Scikit-learn* library).
- The clustering algorithm does not only combine sub-layers from different boreholes, it also combines similar sub-layers from the same borehole. This makes the deletion of ‘peaks’ between similar sub-layers in the layer detection procedure (Section 2) redundant, as they will all be treated by the clustering procedure.
- Although *SBT* is determined for all sub-layers and for the clusters as a whole, based on averaged CPT parameters, the clustering procedure does not take *SBT* into account, but just considers the density of data points in the 3D Robertson diagram, irrespective of *SBT* zones.

3.1 Clustering example

In addition to CPT000000149495, as used in Section 3.1, three other CPTs from the same site in Rotterdam were obtained from Dinoloket, i.c. CPT000000149485, CPT000000149481 and CPT000000149488.

Figure 4 shows all CPT graphs in terms of q_t and R_f as a function of depth, together with their decomposition into sub-layers based on the layer detection procedure described in Section 2 using a *threshold* of 0.1. Note that in this case CPT000000149495 is decomposed in a larger number of sub-layers compared to Figure 1, due to the smaller *threshold* value. In total, there are 60 sub-layers in the four CPTs, numbered 0 to 59, as indicated in Figure 4. The typical numbering in Python is adopted here, in which the first item is numbered 0. Corresponding soil layers can be identified across CPTs as well as within the same CPT.

The clustering was performed using the previously mentioned parameters ($\xi_Q = 1.0$, $\xi_F = 1.0$, $\xi_d = 0.1$, $n_{min} = 3$ and $eps = 0.26$). This gives a total of four clusters with the following sub-layers:

- Cluster 0 = [1,2,3,13,15,16,17,18,28,29,30,31,46,47]
 Cluster 1 = [4,5,6,7,19,20,32,33,34,48,49,50,51,52]
 Cluster 2 = [8,9,10,11,12,21,22,23,24,25,26,35,36,37,38,39,40,41,42,43,53,54,55,56,57,58,59]
 Cluster 3 = [0,14,27,44,45]

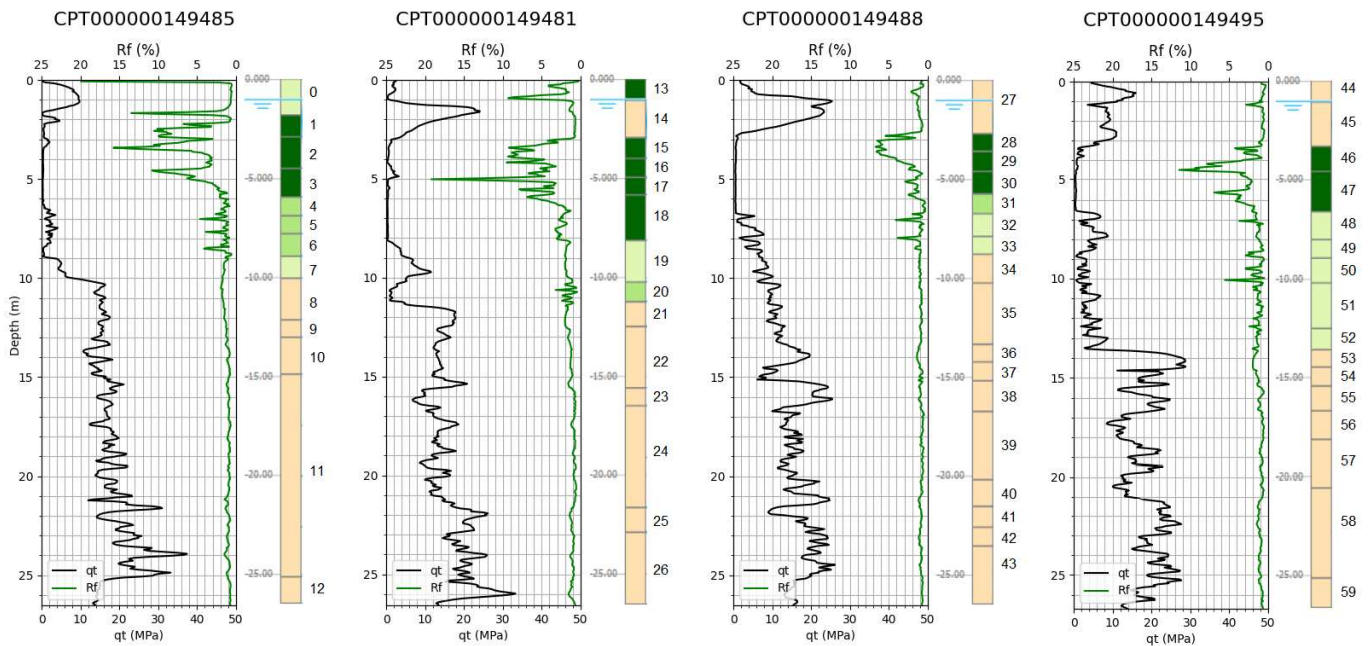


Figure 4. CPT graphs and decomposition into sub-layers

It can be seen from Figure 4 that the clusters contain similar sub-layers; not only across CPTs, but also within the same CPT. In this way, the deep sand sub-layers in CPT000000149495 are still combined into the same 3D soil layer, similar as Figure 1, but here a result of the 3D clustering (part of cluster 2). It can also be seen that the clustering is independent of *SBT*, since sub-layer 0 (*SBT* = 5) is in the same cluster (3) as sub-layers 14, 27, 44, and 45 (*SBT* = 6). Similarly, sub-layer 31 (*SBT* = 4) is in the same cluster (0) as the ‘clay’ sub-layers (*SBT* = 3). Looking at the CPT diagrams and given the relatively coarse layer distribution, this makes sense.

After the clustering, average values of the CPT parameters are calculated for the entire 3D soil layers. These can be used to calculate the layer’s overall *SBT* as well as additional soil and model parameters using the APD concept (Van Berkomp et al., 2022).

4 ELABORATION OF 3D SUB-SOIL MODEL

After the clustering procedure, two comma separated value (CSV) files are created containing ‘borehole’ data that can be imported in Leapfrog to build a 3D sub-soil model. The first file contains the locations and depths of the CPTs; the second file contains the elaborated data, consisting of separate lines for each of the sub-layers. Each line includes the top and bottom of the sub-layer, the cluster number and average values of the CPT parameters. The individual ‘boreholes’ can be visualised in 3D as columns composed of different colours, according to the cluster numbers, so each colour represents a 3D soil layer.

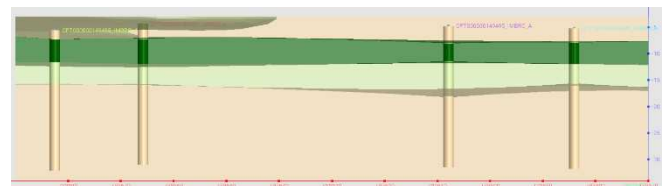


Figure 5. 3D visualization of ‘boreholes’ in a 3D model based on CPT data and clustering

The cluster numbers and depth data are used to create a geological model by running a stratification algorithm, which results in the creation of 3D layer boundaries. Leapfrog uses advanced numerical techniques, such as kriging and radial basis functions, to interpolate and extrapolate around available data, thereby creating 3D surfaces. The stratification can be influenced by providing mesh densities, orientation trends and aspect ratios. We have used a horizontal / vertical aspect ratio of 10:1, which is typical for soil layers. From the layer boundaries and model boundaries, 3D layer volumes are generated, which completes the 3D sub-soil model.

Figure 6 shows the elaborated subsoil model for the entire project location based on a total of 25 CPTs.

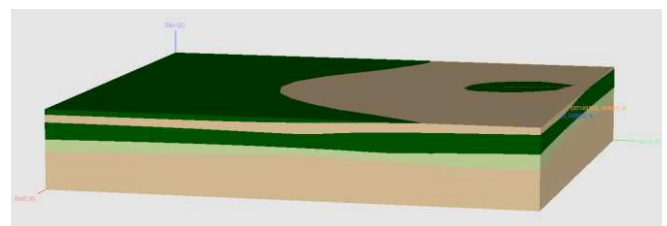


Figure 6. 3D sub-soil model created from ‘boreholes’

5 AUTOMATION VS. RESPONSIBILITY

As automatic processing of geotechnical data highly improves the efficiency of the geotechnical workflow and ‘simplifies’ the work of geotechnical engineers, a discussion has started how far automation should go. No doubt, the geotechnical engineer remains responsible for a geotechnical design, even if this has been highly automated. Therefore, the engineer needs to possess the necessary skills to judge the outcome of automated processes, just as how they would use their own skills to perform the job ‘manually’. To enable engineers to take their responsibility imposes the following requirements on automated systems:

- *Transparency* to the extent that essential ‘decisions’ made in the system are clear and can be overruled.
- (Default) parameters, controlling main functions in the system, can be set and tuned.
- Intermediate and end results from the system can be *verified* and *adapted*.
- Optionally: *Extendibility* of the system to incorporate the expert’s knowledge and experience.

Engineers who feel that these requirements are insufficiently fulfilled by a system, or engineers who are insufficiently skilled to complete a job *without* the system, should not use such an automated system. However, the authors believe that an automated system that is properly used by qualified engineers, will greatly enhance the workflow and reduce possible errors. It will make designs more economic and the engineering job more attractive and joyful. This is not only beneficial for the geotechnical engineering profession, but for all stakeholders in a project.

6 SUMMARY

This paper describes methods for automatic interpretation and processing of CPTs into soil layers. An alternative way of layer detection for single CPTs is presented based on the identification of ‘peaks’ when plotting the moving standard deviation of CPT data as a function of depth. Based on the decomposition of a series of individual CPTs into soil sub-layers, a Machine Learning clustering algorithm is then used to find corresponding layers across multiple CPTs (3D soil layers). Input to the clustering algorithm is the averaged CPT data together with the average depth per sub-layer.

The clustering can be regarded as a search for local densities (clusters) of ‘dots’ in a 3D version of Robertson’s diagram. The result from the clustering is used to create ‘boreholes’ that can be read and processed by geological modelling software to create a 3D subsoil model, which can be used to extract 2D or 3D geometric models for numerical analysis. Averaged CPT data from all contributing sub-layers can be used in conjunction with the Automatic Parameter Determination concept, as published earlier by the authors and co-workers,

to determine soil and model parameters for the 3D layers.

The working of these algorithms has been demonstrated by means of an elaborated example based on a series of CPTs obtained from the Dutch subsoil database Dinoloket.nl.

At the end of the paper it is discussed that automation of the geotechnical workflow can be beneficial for the geotechnical engineering profession as well as for all stakeholders in a project, provided the system is *transparent* and *extendible*, results can be *verified* and *adapted* and the system is used by qualified engineers.

7 REFERENCES

- Brinkgreve, R.B.J. 2019. Automated Model and Parameter Selection: Incorporating expert input into geotechnical analyses. *Geostrata* **23**(1), 38-45.
- Brinkgreve, R.B.J., Brasile, S. 2022. Automatic Finite Element Modelling and Parameter Determination for Geotechnical Design. Workshop *Numerische Methoden in der Geotechnik: Proceedings* (ed: Grabe, J.). Institute of Geotechnical Engineering and Construction Management, TU Hamburg-Harburg, 83-97.
- De Zeeuw, G. 2021. Probabilistic identification of soil stratigraphy using CPT data. Additional Thesis. TU Delft.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X. 1996. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. *Proceedings, 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press, 226-231.
- Lengkeek, H.J. 2022. Testing and Modelling of Sheet Pile Reinforced Dikes on Organic Soils – Insights from the Eemdijk Full-Scale Failure Test. PhD thesis. TU Delft.
- Marzouk, I., Tschuchnigg, F., Paduli F.J., Lengkeek, H.J., Brinkgreve, R.B.J. 2022. Determination of fine-grained soil parameters using an automated system. *Proceedings, CPT '22*, Bologna (eds: Gottardi, G., Tonni L.). CRC Press, 540-545.
- Robertson, P.K. 1990. Soil classification using the cone penetration test. *Canadian Geotechnical Journal*, **27**(1), 151-158.
- Robertson, P.K. 2009. Interpretation of cone penetration tests – a unified approach. *Canadian Geotechnical Journal* **46**(11), 1337-1355.
- Robertson, P.K. 2010. Soil Behaviour Type from the CPT: an update. *Proceedings: 2nd Int. Symposium on Cone Penetration Testing*, Huntington Beach, USA, Vol. 2. 575-583.
- Robertson, P.K. 2016. Cone penetration test (CPT)-based soil behaviour type (SBT) classification system - an update. *Canadian Geotechnical Journal* **53**(12), 1910-1927.
- Van Berkom, I.E., Brinkgreve, R.B.J., Lengkeek, H.J., De Jong, A.K. 2022. An automated system to determine constitutive model parameters from in-situ tests. *Proceedings, 20th Int. Conf. on Soil Mechanics and Geotechnical Engineering*, Sydney (eds. Rahman & Jaksa). Australian Geomechanics Society, 329-334.
- Wang, Y., Huang, K., Cao, Z. 2013. Probabilistic identification of underground soil stratification using cone penetration tests. *Canadian Geotechnical Journal* **50**, 766-776.