Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-796

Evaluating an Unsupervised Clustering-Based Approach for Stratigraphic Interpretation of Piezocone Penetration Testing (PCPT) Data with Statistical Validation

Ivan Cazarez[1], Yu Hin Wong[1,2], Muhamad S. Haji Kasri[1], Frederick Linley[1], Toby Masters[1], Nivethan Kanagarasa[1], George McElroy[1]
[1]*Geoquip Marine, United Kingdom, ivan.cazarez@geoquip-marine.com*
[2]*School of Architecture, Built Environment, Computing and Engineering, Birmingham City University, Birmingham, UK, yuhin.wong@mail.bcu.ac.uk*

ABSTRACT: Offshore geotechnical engineering relies heavily on accurate soil behaviour predictions, with Piezocone Penetration Testing (PCPT) serving as a fundamental method for assessing subsurface conditions. However, the large volumes of data generated, and the inherent complexity of the ground conditions revealed by PCPT present significant challenges for traditional analysis techniques, which often require substantial engineering judgement. As datasets continue to grow in size and complexity, the need for more efficient and reliable methods for data processing and interpretation becomes increasingly critical. This paper presents an evaluation of a machine learning (ML) approach to the stratigraphic interpretation of PCPT data using unsupervised clustering. The approach incorporates a workflow comprising of data preprocessing, anomaly handling, agglomerative clustering with a connectivity matrix to identify continuous soil layers, and a dual validation framework comprising external statistical testing and internal consistency analysis using piecewise linear regression with RANSAC (Random Sample Consensus). External validation assesses alignment between ML-predicted and reported stratigraphy, while internal validation quantifies the coherence of each predicted layer. The results indicate that the ML approach produces stratigraphic models with strong agreement to reported profiles and measurable internal consistency, offering a reproducible, data-driven alternative to manual interpretation. While certain challenges remain, including the refinement of boundary detection in complex sequences and the optimisation of proximal tolerance thresholds, the approach improves the efficiency, transparency, and consistency of PCPT data interpretation, and provides parameters that can support future Bayesian uncertainty estimation.

KEYWORDS: Piezocone Penetration Testing (PCPT), Machine Learning, Automated Clustering, Offshore Geotechnical Engineering

1 INTRODUCTION

In recent years, machine learning (ML) techniques have gained traction within the geotechnical industry because of their ability to efficiently process large datasets and reveal complex, non-linear patterns that may not be apparent through traditional analysis. Some of the ML methods explored have been related to seismic inversion (Huang et al., 2021), ground modelling (Zhang et al., 2020) and Piezocone Penetration Testing (PCPT) interpretation (DeJong et al., 2022) for offshore infrastructure, among others.

The potential applications of PCPT processing are extensive, as it is a widely adopted technique for the rapid characterisation of offshore ground conditions. It is relatively cost-effective, performed in situ, and capable of producing high-resolution, continuous datasets across significant depths. These advantages make it particularly valuable in the context of offshore renewables, where timely and reliable ground characterisation is essential for design and risk management.

One of the primary outcomes expected from PCPT processing is the determination of a site's stratigraphy, as previously explored by Cazarez et al. (2025) and Hudson et al. (2023), through the application of agglomerative clustering, a form of hierarchical clustering, to identify soil layers (clusters) from PCPT data. It is also worth noting that other relevant studies have utilised agglomerative clustering to determine soil type from onshore laboratory testing data (Brosse et al., 2025).

In this paper, the agglomerative clustering workflow is evaluated using a two-stage framework for assessing the quality of unsupervised PCPT-based soil layering models:

1. External validation: the layer depths generated from the agglomerative clustering are compared against those from the reported stratigraphic schematisation (defined in Section 2.3). Hypothesis testing is then carried out to quantify the level of agreement between the two sets of layer depths.
2. Internal validation: a piecewise linear regression (PLR) method using Random Sample Consensus (RANSAC) algorithm is applied to each layer (cluster) to evaluate its internal consistency, quantified by the linear regression fit quality. Layers exhibiting poor conformity to a linear trend may indicate heterogeneity or transitional soil behaviour, prompting refinement of cluster boundaries during post-processing.

These two perspectives together evaluate both the external plausibility and internal coherence of the derived layering model to be evaluated, providing a more comprehensive assessment of its suitability for geotechnical interpretation. Moreover, the outputs from the internal validation, including slope and intercept estimates, residual statistics, inlier-outlier composition, and measures of scatter, combined with the results from the hypothesis testing, provide the necessary statistical parameters to support a Bayesian uncertainty estimation framework for each ML-generated cluster. While the present study does not implement Bayesian modelling, the parameters derived here can form the likelihood inputs for such an approach, enabling probabilistic assessments of stratigraphic confidence in future work.

The proposed model facilitates rapid identification of soil stratigraphy from PCPT data, which is crucial for enabling the timely scheduling of subsequent onshore laboratory testing and for supporting early-stage engineering decisions. By automating and refining the process, this study contributes towards a more efficient and consistent workflow in offshore geotechnical site investigations.

2 DATASETS

2.1 *Ground conditions*

To evaluate the ML-based clustering and validation framework across a range of geological contexts, five (5) offshore investigation sites were selected from distinct geographic regions, each with contrasting soil conditions and depositional histories. The selection aimed to test the applicability of the approach across variable sediment types, consolidation states and layering complexities.

Testing methodologies included: seabed (SB) PCPT, downhole (DH) PCPT, and alternating sampling with downhole (ALT) PCPT, as summarised in Table 1. Geological profiles, interpreted from site investigation reports, were as follows:

- Site 1, sand unconformably overlying overconsolidated clay.
- Site 2, medium dense sand overlying interbedded silt, sand and clay.
- Site 3, clay overlying dense silty sand interbedded with silt and clay.
- Site 4, clay interbedded with sand, increasing in strength with depth from extremely soft to firm.
- Site 5, normally consolidated clay overlying very dense sand.

Table 1. Investigative techniques and scope

Site	No. SB	No. DH	No. ALT	Max Depth	Mean Depth
Site 1	8	0	11	55.7m	26.1m
Site 2	0	10	0	108.9m	93.9m
Site 3	8	0	0	27.9m	22.7m
Site 4	0	6	0	35.8m	24.8m
Site 5	22	7	0	71.0m	24.1m

The soil behaviour type (SBT) chart from Robertson (2010) was used to summarise the PCPT responses across the datasets (Figure 1). As shown in the SBT chart, each dataset contains a large number of datapoints, indicating that the majority of the boreholes provided high-resolution data suitable for ML clustering and statistical validation.

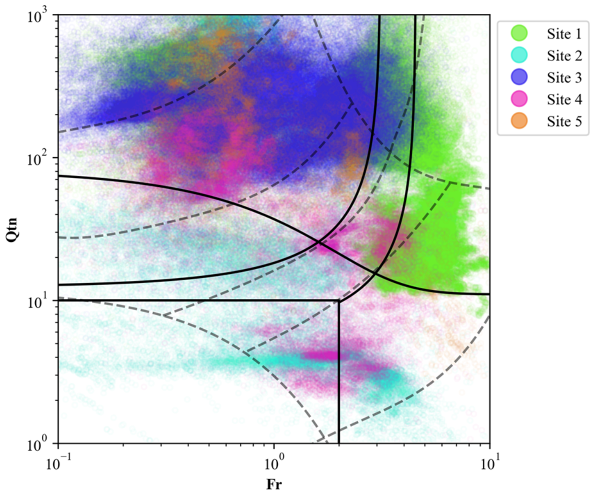

Figure 1. A summary of the soil type of the different sites used in this study, using the soil behaviour type chart (SBT) from Robertson, 2010.

2.2 *PCPT*

The PCPT used in this study were conducted using both 5 cm^2 and 10 cm^2 cones, in accordance with ISO 22476-1, with measurements including tip resistance (q_c), sleeve friction (fs) and pore pressure (u_2). The largest datasets correspond to Sites 2 and 5, including seabed and downhole tests, with maximum penetration depths reaching 108.9 m below seabed (BSB). Sites 3 and 4 had shallower maximum depths, from 27.9 to 35.8 m BSB.

2.3 *Stratigraphic Schematisation*

The soil layer boundaries for each site were determined using two-phase stratigraphic schematisation process:

- Preliminary interpretation from in-situ PCPT profiles, supported by visual inspection of proximal samples in the case of ALT datasets.
- Refinement following results from onshore laboratory testing phase in the case of ALT datasets.

The result of this process is referred to as *reported stratigraphy*, which serves as the baseline stratigraphy for validation, whereas the *ML-predicted stratigraphy* refers to the stratigraphy derived from the layer depths generated by the agglomerative clustering approach. For each location, a ML-predicted stratigraphy was generated using the established methodology. The outputs were then calibrated by visually inspecting the number of layers and adjusting the selected distance threshold if necessary (Figure 2).

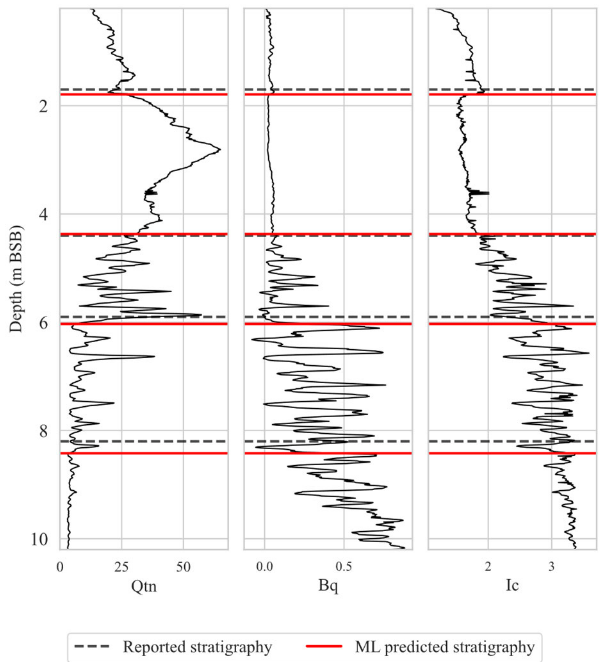

Figure 2. Output generated by hierarchical clustering (Cazarez et al., 2025) against reported stratigraphy.

3 METHODOLOGY

3.1 *Applied clustering algorithm*

An agglomerative clustering ML technique with connectivity matrices was employed to identify continuous ground layers using Ward's linkage method, along processes of standardisation and normalisation (Cazarez et al., 2025). This approach is particularly effective where soil types are distinct, and layer boundaries are relatively sharp.

The workflow included a two-stage pre-processing procedure to clean the raw PCPT data:

- Removal of the embedment stage of the PCPT cone.
- Correction of the tip-sleeve separation.

For downhole PCPT locations, an additional pre-processing step was required to address the gaps in the recorded data. This was achieved through linear interpolation, ensuring the necessary continuity in the dataset for clustering.

The number of clusters generated was controlled by a distance threshold parameter. The greater the distance threshold, the larger the size of the cluster, resulting in fewer number of soil layers generated. However, determining the optimal number of clusters remains a significant challenge, as results are sensitive to both the clustering parameters and inherent variability of the PCPT measurements. The distribution of the selected distance threshold for the ML-predicted stratigraphy across the different sites is presented in Figure 3, whereas distance threshold average values, ranging from 4.1 to 6.0 across the sites, are presented in Figure 3. Variations in both the mean values and the spread of the distance threshold indicate the heterogeneity in ground stratigraphy across all sites.

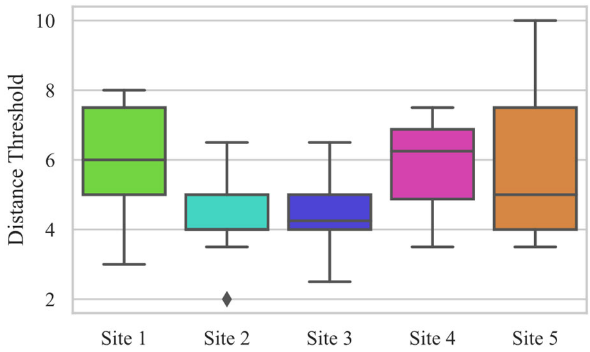

Figure 3. Box and whisker plot of best fit distance threshold by site selected for analysis and training.

3.2 *External validation*

To generate the dataset for external validation, each soil layer boundary from the ML-predicted stratigraphy was paired with the closest reported soil layer boundary in depth. The depth difference between each pair was then calculated. This nearest-boundary matching ensures that comparisons are made between equivalent boundaries, while preserving large differences where mismatches occur. In cases where the number of ML-predicted and reported boundaries differed, unmatched boundaries were retained with their respective difference in depths calculated by taking the closest boundary from the ML-predicted model to reflect potential over or under segmentation. An example of this is shown in Table 2.

Table 2. Example of depth difference between ML-predicted and Reported stratigraphy

ML-predicted Stratigraphy (m)	Reported Stratigraphy (m)	Depth Difference (m)
0.0	0.0	0.0
1.0	/	1.0
2.0	2.2	0.2
3.0	3.3	0.3

In Table 2, the ML model predicted a boundary at 1.0m but the this was not present in reported stratigraphy with the closest boundary being at 0.0m. Therefore, the difference in depth is calculated as 1.0m for validation purposes.

The set of depth differences was then tested using a one-sample, two-tailed *t*-test as shown in Equation (1), to assess whether the mean difference between ML-predicted and reported boundaries differs significantly from zero. The hypotheses were:

- Null hypothesis (H_0): The mean depth difference is 0 m (no bias in predicted boundaries).
- Alternative hypothesis (H_1): The mean depth difference is not 0 m.

A one-sample *t*-test is a statistical test used to determine if a sample mean significantly differs from a known or hypothesised population mean (Walpole et al., 2006). The sample data is composed of the difference between the ML-predicted and reported stratigraphic depths in meters BSB. The *t*-value is calculated as:

$$t = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \quad (1)$$

Where $\bar{x}$ is the mean difference between ML-predicted and reported boundary depths (m BSB); μ_0 is the hypothesised mean (0 m) under H_0; σ is the sample standard deviation of differences; n is the number of paired boundaries

This approach provides a statistical measure of agreement between the ML-predicted and reported boundaries, while ensuring consistency with the paired-boundary methodology used throughout this study. These boundary differences are not independent random samples from a continuous population but are algorithmically derived outputs of a hierarchical clustering procedure with connectivity matrices. Unlike centroid-based clustering, hierarchical clustering does not optimise boundaries with respect to a statistical population mean. Therefore, the one-sample *t*-test applied here should be interpreted primarily as a diagnostic for detecting systematic bias in ML-predicted boundary depths relative to the reported stratigraphy, rather than a strict test of a population parameter under classical statistical assumptions.

This approach provides a clear, quantitative measure of whether the ML-predicted stratigraphy tends to over or under-predict boundary depths when compared to the reported stratigraphy.

3.3 *Metrics of efficacy*

To complement the bias assessment, the accuracy and completeness of the ML-predicted stratigraphy were evaluated using several site-level metrics:

- Failure rate: proportion of reported stratigraphic boundaries for which the ML model produced no proximal match within a specified tolerance.
- Sr: structural recall metric; is the percentage of ML-predicted boundaries successfully generated in comparison to the reported stratigraphy within the proximal tolerance (captures how well structural detail is reproduced).
- R^2: coefficient of determination between the reported stratigraphy and ML-predicted depths.
- MAE: mean absolute error (m) between matched boundary depths.

3.3.1 *Proximal tolerance*

PCPT is conducted at a standard penetration rate of 20 mm/s per ISO 22476-1, with data typically recorded at depth intervals of at least every 10 mm (application- and system-dependent). This logging rate comfortably satisfies sampling needs such as the Shannon-Nyquist theorem for the bandwidth of depth-varying CPT signals.

However, the matching tolerance must cover more than instrument resolution, it needs to accommodate (i) vertical variability between nearby soundings, (ii) uncertainty in stratigraphic schematisation, (iii) heave/vertical-datum corrections offshore, (iv) gap-filling/interpolation (for DH tests), and (v) clustering segmentation error. Accordingly, this study adopts a proximal tolerance of 4.5m as a conservative window derived from cm-scale logging resolution plus the above interpretation uncertainties. Thorough estimation of the proximal tolerance is beyond the scope of the present work.

3.4 *Internal validation*

Internal validation was performed using a PLR approach to assess the internal consistency of each proposed layer. This method is a modelling approach in which a dataset is divided into contiguous sections, with each section approximated by its own linear relationship, allowing for abrupt changes in slope at predefined or detected breakpoints; it provides an interpretable, data-driven measure of quality, and its regression metrics can be used in subsequent algorithms to guide the selection of the optimal number of layers in each borehole.

Although PCPT profiles are not perfectly piecewise linear, linear trends provide a practical approximation of the internal consistency within each layer. In this context, "linearity" does not imply an idealised geological form but rather provides a simple benchmark for assessing whether data within a layer are sufficiently homogenous to be approximated by a straight-line fit.

The PLR method was applied to four PCPT-derived parameters (*Qtn, Fs, Bq,* and *Ic*) selected for their sensitivity to soil type variation and their amenability to representation through linear segments. These parameters reflect the physical processes governing soil behaviour, and their trends often approximate linearity within individual strata formed by depositional processes.

Within each layer, RANSAC linear regression was applied independently to each parameter, treating the PCPT parameter as the dependent variable and depth as the independent variable. RANSAC iteratively selects random subsets of data, fits a model, and evaluates its quality based on the number of inliers. The iteration with the maximum number of inliers was selected, and the model was then refitted to those inliers to produce the final regression. This approach is robust to outliers, which are common in PCPT datasets due to localised anomalies. RANSAC was implemented using scikit learn (Pedregosa et al., 2011) with the following configuration:

- min_samples = 2, allowing the algorithm to explore candidate models without constraining the eventual inlier proportion.
- $max\_trials = 1000$, ensuring sufficient iterations to avoid suboptimal fits, with early termination if convergence was achieved.
- Residual threshold, set adaptively using the mean absolute deviation, MAD, Equation (2) of the data:

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}| \tag{2}$$

Where MAD is the Mean Absolute Deviation; n is the number of values in the dataset; x_i is the individual value; $\bar{x}$ is the mean of the dataset

For each layer-level regression, the root mean squared error (RMSE) was computed and normalised by the standard deviation of the parameter within the respective layer to obtain the normalised RMSE (NRMSE). While NRMSE provides a convenient scale-independent measure of fit quality, it is not a direct output of the RANSAC algorithm and does not in itself validate the robustness of the RANSAC fitting process. Instead, NRMSE here serves as a post-hoc diagnostic, quantifying the degree to which the consensus inlier set identified by RANSAC confirms to a linear trend relative to the layer's natural variability (Figure 4).

These regression metrics are derived from algorithmically clustered layers and therefore not independent statistical samples in the traditional sense. As with the external validation *t*-test, the results should be interpreted primarily as diagnostic indicators of internal coherence rather than as a formal inferential statistics describing a population. Layers with high inlier percentage and low NRMSE are considered internally consistent and well-approximated by a linear trend, whereas layers with low inlier percentage and/or high NRMSE are flagged for potential subdivision or boundary refinement in subsequent processing.

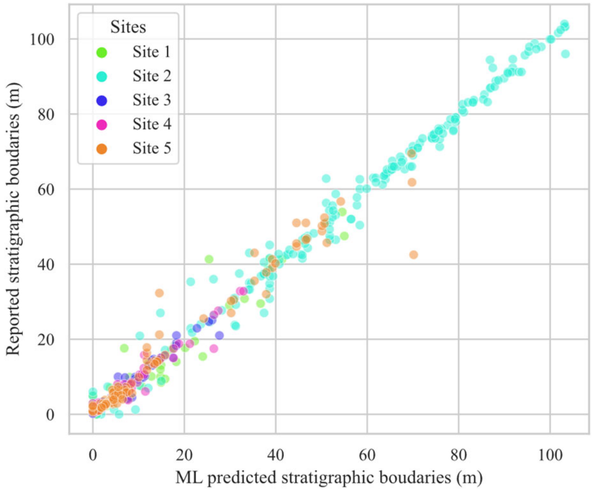

Figure 4. The correlation between ML-predicted and reported stratigraphy.

4 RESULTS

4.1 *Statistical Analysis*

Following the procedure described in Section 3.2, each ML-predicted stratigraphy boundary was paired with the nearest reported stratigraphic boundary, and the depth difference between each pair was calculated. Figure 5 shows the histogram of the depth difference between the ML-predicted and reported stratigraphy, showing that the mean of depth difference is 0.03m, with MAE of 0.63m for all the sites.

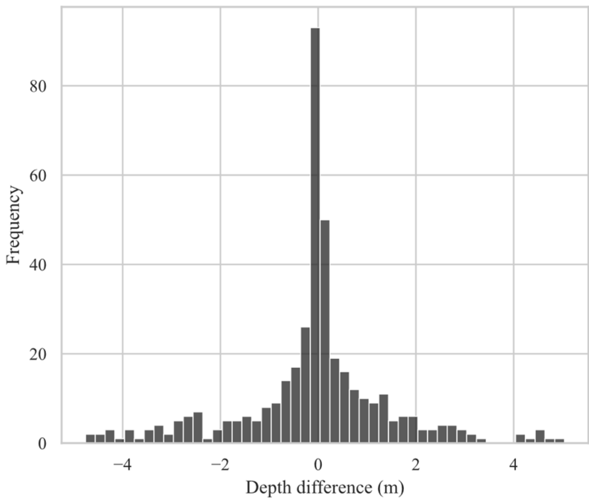

Figure 5. Histogram of depth difference between ML-predicted stratigraphy and reported stratigraphy (only data within 5.0m difference presented for clarity).

Figure 4 shows the distribution of the boundary depths from ML-predicted stratigraphy and the reported stratigraphy for each specific site. This figure shows a strong overall correlation between the two stratigraphy, with minimal scatter observed at shallower depths <20m BSB, slightly higher variability between 20 and 60m BSB, and minimal variability at greater depths >60m below seabed.

This analysis was performed on signed depth differences to maintain consistency with the bias-detection focus of the on-sample *t*-test defined. Absolute values were not used in the *t*-

test, as doing so would test a different null hypothesis (mean absolute error), rather than the presence of systematic bias.

The resulting dataset yielded a t-value of 1.23 and an associated two-tailed p-value of 0.22. With a significance level of $\alpha = 0.05$, corresponding to 95% confidence threshold (as also applied by Edmonds et al., 2025), this result does not allow us to reject the null hypothesis that the mean depth difference is zero. In other words, there is no statistically significant evidence of systemic over- or under-prediction of boundary depths across the dataset by the ML approach.

Regarding the metrics of efficacy, a congruous boundary was defined in this research as any ML-predicted boundary that (1) is not a boundary beyond the proximal tolerance, and (2) is not a reported boundary missed entirely by the ML model. The congruous dataset is the set of all matched congruous boundaries within the proximal tolerance. Table 3 summarises the site-by-site results.

Table 3. Metrics of efficiency obtained from external validation analyses.

Site	Failure rate (%)	Sr (%)	R^2	MAE (m)
Site 1	6	71	0.93	0.99
Site 2	0	64	0.89	0.61
Site 3	14	62	0.95	0.79
Site 4	0	51	0.94	0.32
Site 5	0	73	0.96	0.50

Across all sites, the global R^2 for congruous boundaries exceeded 0.89. At the site level, 55% of individual locations achieved R^2>0.95, indicating very high correlation between predicted and reported boundaries.

For the congruous dataset, 44% of boundaries had a difference of less than 0.15 m between predicted and reported depths. Overall, 12% of all boundaries were not proximal matches, with fewer than 2% of congruous boundaries falling outside the 4.5 m threshold. The largest observed absolute errors for congruous boundaries were 13.9 m and 12.2 m in two isolated cases corresponding to boundaries in stratigraphically complex zones.

4.2 *PLR- RANSAC analysis*

For each soil layer profile generated following the procedure outlined in Section 3.1, the PLR-RANSAC procedure was applied to each layer and for each of the four PCPT-derived parameters (*Qtn, Fr, Bq, Ic*). An example of a layered profile with fitted regression lines is shown in Figure 6.

The outputs of this analysis (inlier percentage and NRMSE) were computed for each layer-parameter combination:

- Inlier percentage measures the proportion of points in the layer that conformed to the RANSAC-identified linear model within the adaptative MAD-based residual threshold.
- NRMSE quantifies the magnitude of residuals relative to the standard deviation of the parameter within that layer, providing a scale-independent indicator of fit quality.

Mean NRMSE values across parameters were then used as input features in the optimisation and ranking stage (described in section 3.5) to evaluate and prioritise candidate layering configurations. This approach emphasises layers with low NRMSE and high inlier percentage, which are considered internally consistent and well-approximated by linear trends.

Additionally, NRMSE is not a direct measure of RANSAC's robustness or internal validation; it is a post-hoc fit quality metric applied to the inlier set identified by RANSAC. The regression metrics reported here are descriptive diagnostics of internal coherence and should not be interpreted as inferential statistics in the traditional sense. As in external validation, the layers themselves are the product of an algorithmic segmentation, not independent samples from a statistical population.

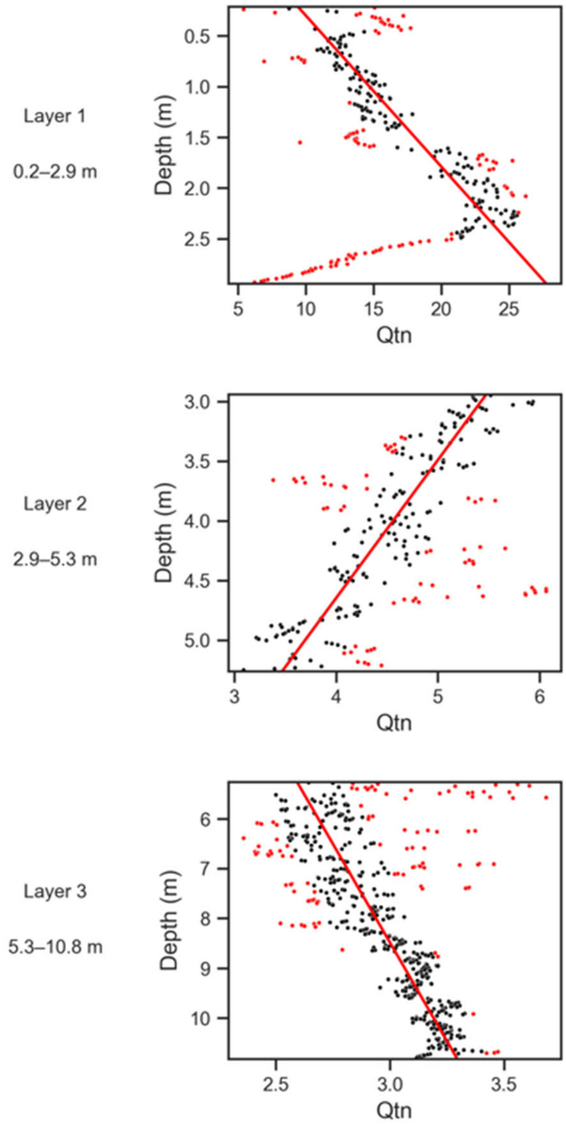

Figure 6. The result of RANSAC regression on a PCPT profile of the parameter Qtn. The red line represents the RANSAC regression line on each layer. The black and red points are inliers and outliers, respectively.

5 DISCUSSION

Offshore geotechnical investigations generate large volumes of high-resolution data, but turning those numbers into a coherent understanding of the subsurface remains a task that often depends on experienced interpreters working layer by layer. The starting point for this research was the recognition that PCPT produces continuous, centimetre-scale records of soil behaviour, a richness of information that can be difficult to interpret consistently across sites, depths, and geological settings when done entirely by hand. The core idea was simple: if a ML approach could reliably identify the same key stratigraphic boundaries that an experienced geotechnical engineer would see, it could accelerate project timelines, improve reproducibility, and lay the ground for more quantitative assessments of uncertainty.

To make that possible, the work first assembled a dataset broad enough to capture the complexity and variability that offshore projects encounter. The five selected sites were intentionally different, in geography, depositional environment, and soil composition, so that any proposed method would have to perform not just in clean, well-layered sands, but also in mixed sequences, transitional layers, and clays with varying consolidation states. The PCPT data came from both seabed and downhole tests, ensuring that the workflow would be tested against the practical realities of offshore acquisition modes. This diversity meant that the ML model would be exposed to the full spectrum of PCPT responses, from smooth trends to abrupt shifts, and from noise-free sections to spikes caused by shells, gravels, or other anomalies.

The workflow combined agglomerative clustering, guided by a connectivity matrix to ensure depth continuity, with careful pre-processing to make profiles from different equipment and modes directly comparable. Validation was approached from two complementary angles: external checks for plausibility against reported stratigraphy, and internal checks for coherence within each predicted layer. External validation shows that, on average, there is no statistically significant evidence of systematic over- or under-prediction of boundary depths by the ML workflow. Nonetheless, this reflects only the overall trend; in a few instances at specific depths and locations, the ML-predicted stratigraphy failed to reproduce boundary layers, with discrepancies exceeding 15 m. This highlights that while the algorithm performs well globally, its layer-merging nature limits its ability to recover missed boundaries once clusters are combined. Addressing this would require additional pre-and post-processing steps to refine segmentation in problematic zones.

Another challenge identified as a result of this research lies in selecting the distance threshold that controls the number of clusters. In its current form, the method relies on selecting the distance threshold parameter to determine the optimal number of clusters, often requiring iterative user input to capture the necessary soil layer boundaries. This introduces subjectivity and limits automation. One possible route forward is to integrate a feedback mechanism that evaluates clustering outcomes against the efficacy metrics presented here, failure rate, structural recall, R^2, MAE, inlier percentage, and NRMSE, and iteratively adjusts the distance threshold to converge on an optimal configuration. Such an adaptative loop could reduce manual intervention, but it would require sufficient training data to calibrate the plausibility checks. By combining domain knowledge, robust validation metrics, and self-iterative clustering, future refinements of this workflow could further enhance both its accuracy and its interdependence from manual tuning, while keeping its outputs aligned with geological reality.

6 CONCLUSIONS

This study presents the statistical validation of a structured workflow of the stratigraphic interpretation of PCPT data using agglomerative hierarchical clustering with connectivity matrices, supported by a dual statistical validation framework. The approach demonstrates that clustering-derived layers can be evaluated both for external plausibility against expert-derived baselines and for internal consistency through piecewise linear regression and RANSAC. This combination establishes a sound statistical rationale for applying hypothesis testing to PCPT profiles segmented by hierarchical clustering, despite the algorithm's non-centroid nature, and ensures that model quality is assessed from complementary perspectives.

The methodology produces parameters, including depth offsets, inlier proportions and normalised regression errors, that form a practical basis for future Bayesian uncertainty estimation. The adoption of PLR and RANSAC provides robustness to outliers, aligns with geotechnical interpretations of gradual property change within layers, and yields interpretable quality metrics. While a provisional 4.5m tolerance was applied to match predicted and reported boundaries, the study recognises that this value is based on assumptions; its optimisation warrants dedicated research to balance instrument resolution, interpretation uncertainty, and site variability.

Overall, the work contributes a reproducible, data-driven framework for PCPT stratigraphic modelling that maintains alignment with domain practice while introducing statistical rigour. It lays the foundation for extending the analysis into probabilistic uncertainty modelling and for refining key parameters, such as tolerance thresholds, through targeted future studies.

7 REFERENCE

Brosse, A. M., Christopher, N, Barwise, A. Ghasemi, P. 2025. Using machine learning clustering techniques as a support for optimizing the laboratory testing campaign of an offshore wind farm. Proc, 5th Internation Symposium on Frontiers in offshore geotechnics, Nantes, 895-857

Cazarez, I. Kanagarasa, N. Masters, T. Linley, F. Wong, Y.H. 2025. Advanced Machine Learning PCPT Interpretation in Offshore Geotechnical Investigations. Proc, 5th Internation Symposium on Frontiers in offshore geotechnics, Nantes, 927-932

DeJong, J.T., et al. 2022. Geotechnical Applications of Machine Learning: Current Trends and Future Directions. Computers and Geotechnics.

Huang, Y., et al., 2021. Deep learning for seismic-to-geotechnical property prediction. Soil Dynamics and Earthquake Engineering.

Hudson, K.S., Ulmer, K.J., Zimmaro, P., Kramer, S.L., Stewart, J.P. and Brandenberg, S.J., 2023. Unsupervised machine learning for detecting soil layer boundaries from cone penetration test data. Earthquake Engineering & Structural Dynamics, 53(4), pp.1027–1046. https://doi.org/10.1002/eqe.3961

Kamp, F. Gunawan, H, Hamdan, N, Thumann, V,. 2025. Machine learning approach for predicting plug heave for suction caissons in cohesionless soil. Proc, 5th Internation Symposium on Frontiers in offshore geotechnics, Nantes, 908-914

Robertson, P.K. (2010). Evaluation of cone penetration tests (CPT) for geotechnical site characterization. Soils and Rocks, 33(2), pp. 121–132

Robertson, D. G. (2009). Interpretation of cone penetration tests – a unified approach. Canadian Geotechnical Journal, 46(11), 1337-1355. https://doi.org/10.1139/T09-073

Shi Y, Ke G, Soukhavong D, Lamb J, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T, Titov N, Cortes D 2025. lightgbm: Light Gradient Boosting Machine. R package version 4.6.0.99, https://github.com/Microsoft/LightGBM.

Walpole, R. Myers, R. Myers, S. Ye, K. 2011. Probability & Statistics for Engineers & Scientists.

Zhang, J., et al., 2020. Application of Machine Learning in Subsurface Modelling. Engineering Geology.

Bishop, 2006, Pattern Recognition and Machine Learning