

Geo-risk assessment using GroundIQ® solution: Case histories on artificial intelligence-based method for 3D ground modelling

Simon Feller

Geoconsultancy Land, Fugro, Belgium, s.feller@fugro.com

Jean-François Vanden Berghe

Geoconsultancy Land - Europe, Fugro, Belgium, jf.vandenbergh@fugro.com

Alexandre Bolève

Geophysics, Fugro, Netherlands, a.boleve@fugro.com

Malek Allani

Geoconsultancy Land, Fugro, Belgium, m.allani@fugro.com

ABSTRACT: This paper presents a series of case studies demonstrating the application of artificial intelligence (AI) techniques specifically clustering and Machine Learning (ML) within Fugro's GroundIQ® solution for 3D ground modelling. The integration of AI enables the automated construction of subsurface models by correlating geophysical data with geological and geotechnical inputs, significantly reducing human bias and improving efficiency. Clustering, an unsupervised method, is used to classify soil units based on geophysical parameters, while ML algorithms such as CatBoost and Random Forest are employed to predict subsurface characteristics using labelled datasets. Case histories from the Netherlands, Belgium, and Germany illustrate the versatility of these methods across varying geological contexts. The results show that AI-driven modelling enhances the accuracy and reliability of soil stratigraphy interpretation, especially when supported by high-quality and extensive in-situ testing. This approach offers a robust framework for subsurface risk assessment and foundation design, particularly in complex urban environments.

KEYWORDS: GroundIQ®, Clustering, Machine learning, matrix, correlation, 3D model.

1 INTRODUCTION

Processing large and diverse amounts of data often poses significant challenges, especially when the processing must be done manually through various calculations and analyses. This typically requires substantial human effort, considerable IT resources, and a lot of time.

The geosciences are no exception to this trend, particularly given the increasing capacity of acquisition systems that collect ever-growing amounts of data in the field. Faced with these challenges—complexity of projects, available personnel resources, and the quantity and variability of data—the use of artificial intelligence emerges as a promising and relevant solution for effectively managing and analyzing the data collected at research sites (Sauvin, et al., 2019) (Sauvin, et al., 2022).

Artificial intelligence, a branch of computer science, focuses on developing intelligent machines capable of performing tasks typically associated with human intelligence (Hastie, et al., 2001). AI distinguishes between unsupervised techniques, known as clustering, and supervised learning techniques, which include machine learning and deep learning (Samuel, 1959).

The integration of AI into the GroundIQ® solution aims to automatically construct a 3D soil model by incorporating geological and geotechnical data into the geophysical analysis. Traditionally, this task was performed by a geotechnical engineer or geologist. However, with AI, it can now be executed automatically and swiftly, eliminating the subjective bias of the operator.

This article delves into the specifics of clustering and machine learning methods and presents case studies from several projects across Europe. By leveraging AI, the process of building accurate and reliable soil models becomes more efficient, ensuring that geotechnical investigations are based on comprehensive and precise data. This advancement not only streamlines the workflow but also enhances the overall quality

of subsurface analysis, making it a valuable tool for modern geoscience applications.

2 SUPERVISED AND UNSUPERVISED TECHNIQUES

2.1 Clustering

Clustering (unsupervised method) involves grouping geodata based on similarities. This process helps to create more interpretable data clusters that correspond to geological or geotechnical units (Figure 1). The goal is to identify general trends from the input data, simplifying the subsurface picture for easier interpretation. In geosciences, Clustering method uses geophysical data such as shear wave velocity, compressional wave velocity and resistivity values.

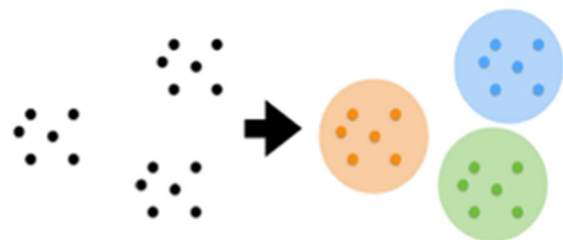


Figure 1 : Clustering method principle

The clustering technique can be directly applied to CPT data using the Robertson point cloud distribution (Wang, et al., 2018) (Xiangrong & Hui, 2021), where the friction ratio (Fr) is plotted on the x-axis and the normalized cone resistance (Q_{tn}) on the y-axis. This method results in an automated classification of data points, representing possible different geological units (Figure 2).

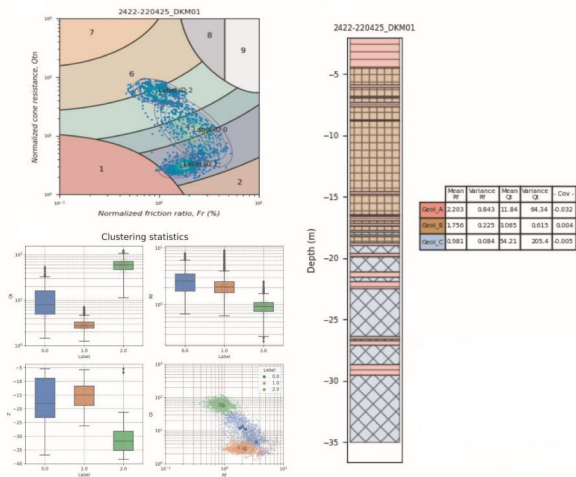


Figure 2: Clustering technique on CPT data. Top left: clustering result on Roberston graph. Bottom left: output clustering statistics. Right: synthetic geological log resulting from the clustering.

The advantages of this type of grouping can be summarized as follows:

- Simplifying a large amount of information (combining data to extract useful information and generate a basic preliminary geological model).
- Automated, non-subjective approach.
- Decision support at the beginning of a project to accelerate planning and control costs.

2.2 Machine learning

Artificial intelligence techniques, such as machine learning, offer a robust method for integrating geosciences data to provide valuable information for effective management of subsurface risks. ML, which is generally considered as a subset of AI, allows computers to learn from data without being explicitly programmed to do so. This capability is particularly useful in geosciences, where vast amounts of complex data need to be analysed and interpreted.

ML algorithms, such as CatBoost (Sharchilev, et al., 2017) (Prokhorenkova, et al., 2018) and Random Forest (Breiman, 2001) (Pedregosa, et al., 2011) (Kuhn, et al., 2018) are employed to build predictive models based on data samples. These models can analyse input data, such as geophysical measurements, and correlate them with labelled data, such as lithological information or CPT data. This process involves a learning phase where the algorithm is trained to recognize patterns and make accurate predictions. Compared to unsupervised methods like clustering, ML requires this learning phase to effectively match input data with labelled data. This enables the creation of models that can predict subsurface and underground characteristics and potential risks, thereby aiding in the management and mitigation of these risks.

ML typically involves the following steps: First, a matrix of training data is generated, containing geophysical and geotechnical (or geological) information. This training matrix serves as the learning source for algorithms like CatBoost and Random Forests. The learning phase is performed using a series of trees where each tree gives a response associated to an input. The final model prediction corresponds to the “average” response of the forest of trees (for regression problem) (Figure 3). To evaluate model prediction accuracy, the training matrix is divided into five equal parts. Four-fifths of the dataset is used for training and model building, while the remaining fifth is used to compare predictions against actual data. This process is

repeated, using different parts as the training dataset, until all combinations have been tested. The result with the best fit score is then applied to the entire dataset.

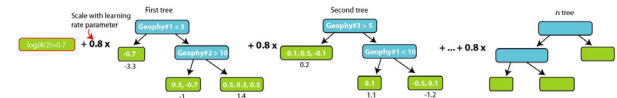


Figure 3 : Machine learning method principle

3 CASE STUDIES

3.1 Netherlands

As part of GroundIQ®, Fugro’s patented and innovative ANT (Ambient Noise Tomography) technology provides geophysical data to determine surface properties and structure. Using passive seismic data, ANT Fugro technology generates a 3D shear wave velocity (V_s) soil model. This first example (Figure 4) illustrates the final result obtained using the ANT technology at the site of Fugro’s future headquarters in Nootdorp, Netherlands.

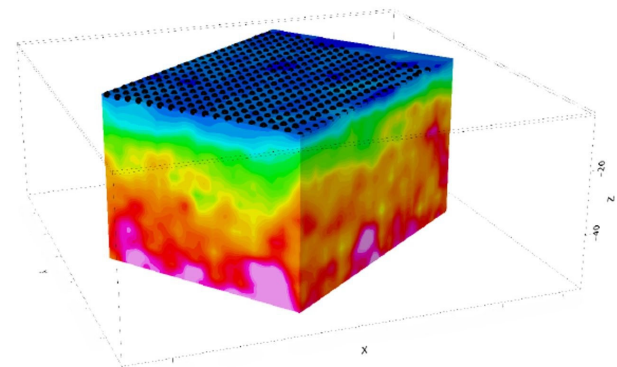


Figure 4 : Nootdorp 3D ANT V_s model

The advantage of using ANT or other geophysical methods lies in their ability to determine surface homogeneity right from the surface site investigations. This early insight allows for the optimization of the number and location in-situ tests based on the geophysical results obtained. Initially, consolidation efforts focused solely on shear wave velocity (V_s) data. The goal was to create a detailed image of the subsurface stratigraphy and correlate this information with Cone Penetration Test (CPT) data to evaluate the suitability of the foundations design recommendations.

In this case study, the clustering technique was employed to test its effectiveness in detecting geological transitions between surface clay and deeper sand layers. The model derived from the unsupervised analysis revealed the presence of two distinct layers, characterized by V_s velocity intervals ranging from 50 to 240 m/s and 240 to 660 m/s, respectively. These findings are visually represented in Figure 5, which clearly illustrates the correlation between the sand-clay interface and the increase in cone resistance measurements obtained from CPT.

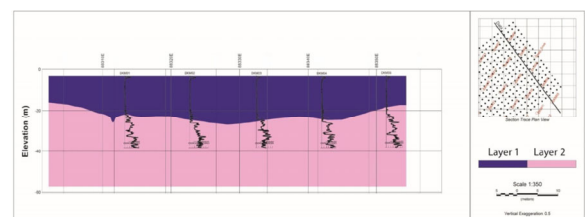


Figure 5 : Cross section obtained from clustering analysis

3.2 Belgium

The second example is set within a context of more complex geology. Like the previous case study, the ANT geophysical survey method was employed over an area of 20,000 m². Clustering was used to differentiate the stratigraphy based solely on shear wave velocity (Vs). As represented in Figure 6, the model identify four distinct layers: the first layer corresponds to the superficial soil layers, the second to weathered shales, the third to healthy shale and the fourth and deepest layer to even deeper shale.

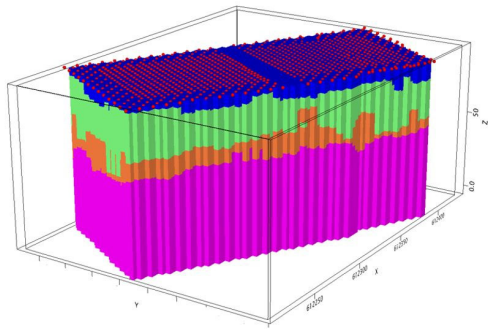


Figure 6 : 3D geological model obtained using clustering analysis

The Figure 7 shows the Vs velocity groups generated by the clustering. The last two groups are combined in a single colour (pink).

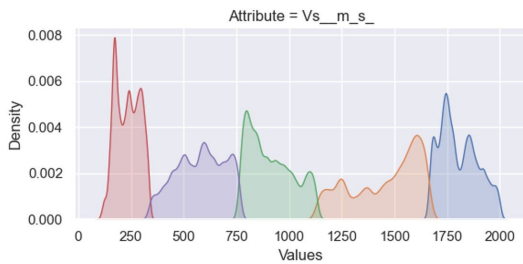


Figure 7 : Vs clustering results by group obtained from ANT data

It is evident that relying solely on geophysical classification has its limitations, especially when the layers within the top 10 meters of the ground are relatively thin. Consequently, the clustering results do not accurately reflect the actual conditions on the ground. In practice, clustering groups the first 10 meters of soil into a single cluster, whereas on-site observations reveals that these 10 meters consist of three distinct soil layers: silt, clay and gravel.

To address this discrepancy, the machine learning approach was applied to this case study. Figures 8 and 9 illustrate the results obtained through ML.

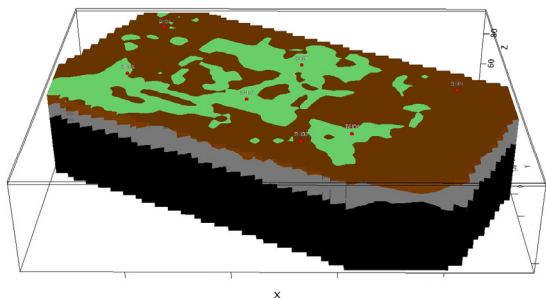


Figure 8 : 3D geological model obtained using machine learning approach

The ML algorithm uses boreholes data, which includes only four geological units. Therefore, the resulting 3D geological model represents the prediction of in-situ data (borehole) based on geophysical data (Vs).

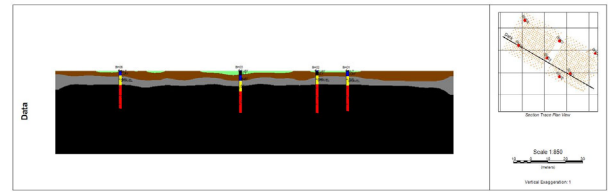


Figure 9: Cross section of the 3D geological model using ML

A notable observation is that the 3D model obtained through ML differs significantly from the clustering result. While the classification indicated five layers, machine learning identifies only four layers with varying thicknesses and locations. The prediction of borehole data based on geophysical data obtained through ML is thus much more consistent and aligned with the true nature of the soil.

3.3 Germany

Fugro conducted a survey at another site using the ANT method. Although machine learning wasn't necessary for this case study, it served as a valuable demonstration of how AI can enhance the accuracy of soil models by integrating various sources of information. During the learning phase, 15 out of the total 17 CPTs were used, while CPTs no. 7 and no. 10 were excluded to serve as test cases for evaluating the prediction quality and the overall accuracy of the model.

The prediction results, illustrated in Figure 10, show that the algorithm produced relatively consistent outcomes, which can be attributed to the simplicity of the area which is essentially characterized by planar monoclinial layers. However, the inclusion of X and Y coordinates as input data during the learning phase, significantly improved the spatial specificity of the predictions. This enhancement is evident when comparing the Qtn value predictions with and without the X and Y coordinates (represented in blue and red, respectively). The comparison clearly demonstrates that incorporating spatial information allows the predictions to more closely align with the actual conditions on the ground.

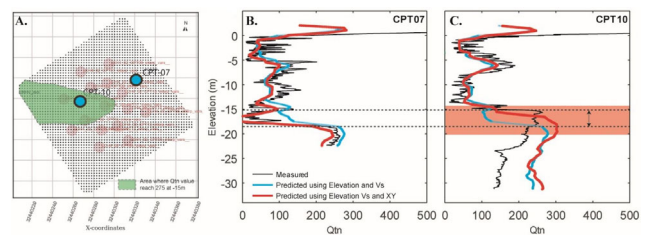


Figure 10: Comparison of predicted CPT data. A: Map of Vs data points and CPTs location. B and C: Comparison between measured and predicted Qtn values for CPT 07 and 10 respectively.

Nevertheless, the Vs/CPT correlation for CPT 07 is not optimal. This can be attributed to several factors. Firstly, as previously mentioned, CPT 07 and 10 were excluded from the database used for decision trees calculations and were instead used to verify correlation. CPT 10 exhibits a significant difference in Qtn values below 20 meters compared to other CPTs. Another possible explanation is that the boreholes nearest to CPT 10 are shallower and have different Qtn data. These factors combined suggest that greater differences in depth can lead to increased uncertainty in the correlation.

Additionally, the variability in soil structure plays an important role in the accuracy of the correlation. In areas with highly variable soil structures, the reliability of artificial intelligence can be limited. This is particularly evident in the case of CPT 10, where the soil composition and properties may differ from those of other CPTs. The presence of such variability introduces a higher degree of uncertainty in the correlation.

Furthermore, the limitations of using artificial intelligence in geotechnical investigations become apparent when dealing with complex subsurface conditions. While AI can provide valuable insights and enhance the accuracy of the predictions, it is not infallible. The model's performance is heavily dependent on the quality, quantity and consistency of the input data. In the CPT 10, differences in depth and soil properties contribute to the observed discrepancies in the Vs/CPT correlation.

Consequently, the geotechnical QTn model in Figure 11 is limited to around 20 meters depth, where the accuracy of the data correlation is more reliable. This decision ensures that the model's predictions remain within a range where the correlation between Vs and CPT data is more consistent and trustworthy. By focusing on this depth range, the study aims to provide a more accurate representation of the subsurface conditions despite the inherent limitations of the AI model.

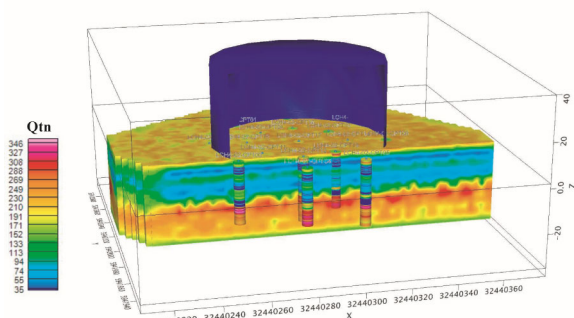


Figure 11: 3D Qtn model results based ANT Vs data

This example underscores the critical importance of thoroughly evaluating the data obtained and involving a real person, such as an engineer or geologist, in the process. Their expertise is essential for creating a 3D soil model that closely approximates reality. By integrating human insight with advanced technology, a soil model has been developed that provides predictions of Qtn values at each point within the 3D model. These predictions are derived from the learning of geophysical data coupled with CPT data.

4 CONCLUSION

GroundIQ® and Ambient Noise Tomography (ANT) provide essential information on subsurface properties and structure, forming Fugro's new comprehensive and integrated solution for subsurface risk management. ANT technology is particularly effective in urban environments, where other techniques may fall short, addressing the numerous challenges associated with large-scale infrastructure projects, such as the design of specific deep foundations.

Fugro has also developed a computational method that leverages clustering and ML to correlate geophysical data with data obtained from in-situ tests, particularly boreholes or CPT. By integrating the available survey data, it becomes possible to quickly generate 3D soil model, which can be directly used for foundation design and project risk assessment.

One example of clustering and two examples of ML applications were presented: one using clustering for layering, one using AI to

- predict geological models based on geophysical and borehole data (Belgium case),
- to generate a 3D Qtn model and evaluate prediction accuracy (Germany case).

These examples demonstrate the versatility and effectiveness of clustering and machine learning in enhancing subsurface modelling.

The efficiency of these new methods are made possible by application of AI in processing geophysical, geotechnical and lithological data. This integration of advanced technology ensures that subsurface investigations are more precise, reliable, and capable of addressing the complex demands of modern infrastructure projects.

And finally, it was demonstrated that in cases of complex soil stratigraphy, the quality and quantity of tests significantly enhance the accuracy and reliability of the models generated through AI. By conducting thorough and extensive testing, AI algorithms can better interpret the intricate variations in soil layers, leading to more precise predictions and a more faithful representation of the subsurface conditions.

This finding underscores the importance of comprehensive data collection in geotechnical investigations. High-quality and high-quantity test data provide the necessary foundation for AI models to perform effectively, ensuring that the resulting soil models are both accurate and trustworthy. This approach is particularly important in challenging geological settings, where the complexity of soil stratigraphy demands a robust and detailed analysis to inform foundation design and risk assessment.

5 REFERENCES

Breiman, L., 2001. Random forests: Machine Learning, 45(1), pp. 5-32.

Hastie, T., Tibshirani, R. & Friedman, J., 2001. *The elements of Statistical Learning*. New York, NY, USA: Springer New York Inc..

Kuhn, S., Cracknell, M. & Reading, A., 2018. Lithologic mapping using Random Forest applied to geophysical and remote-sensing data: A demonstration study from the Eastern Goldfields of Australia. *Geophysics*, 83(4), pp. B183-B193.

Pedregosa, F. et al., 2011. Scikit-learn: Machine Learning in python. *The Journal of Machine Learning Research*, 12(2).

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V & Gulin, A., 2018. *CatBoost: unbiased boosting with categorical features*. s.l., s.n.

Samuel, A., 1959. Some Studies in Machine Learning Using the Game Checkers. *IBM Journal of Research and Development*, 3(3), pp. 210-229.

Sauvin, G., Vanneste, M., Vardy, M., Klinkvort, R.T & Forsberg, C., 2019. *Machine Learning and Quantitative Ground Models for Improving Offshore Wind Site Characterization*. Houston, USA, OTC Paper 29351, pp. 6-9.

Sauvin, G., Vardy, M., Klinkvort, R.T., Vanneste, M., Forsberg, C.F & Kort, A., 2022. State-of-the-Art Ground Model Development for Offshore Renewables – TNW Case Study. pp. 1-5.

Sharchilev, B., Ustinovsky, Y., Serdyukov, P. & de Rijke, M., 2017. *Finding Influential Training Samples for Gradient Boosted Decision Trees*. s.l., s.n.

Wang, X., Wang, H., Liang, R. Y. & Liu, Y., 2018. A Semi-Supervised Clustering-Based Approach for Stratification Identification Using Borehole and Cone Penetration Test Data. *Engineering Geology*.

Xiangrong, W. & Hui, W., 2021. Unsupervised and Simultaneous Stratigraphic Interpretation of CPT Soundings at Site Scale. *International Journal of Geomechanics*, 21(8).