Machine learning based soil property prediction: A Quantitative Ground Model building approach based on 3D UHRS

R. Ruiz[1], N. Pernin[1], L. Limonta[1]

[1] TGS UK

Summary

The world is undergoing an energy transformation, with the UK aiming for 50 GW of offshore wind capacity by 2030. Achieving this goal requires technologies that can accelerate the development of accurate 3D Quantitative Ground Models (QGM) for site characterization. This study investigates the integration of 3D Ultra High-Resolution Seismic (3D UHRS) data with Quantitative Interpretation (QI) methods for this purpose.

Two machine learning (ML) models based on the Gradient Boosted Tree algorithm are tested. The first model uses post-stack 3D UHRS data to predict cone resistance from acoustic impedance and envelope attributes. The second model shifts these curves vertically to address potential time-depth conversion uncertainties. Results indicate a strong correlation between predicted and measured cone resistance, with ML methods outperforming traditional linear regression in capturing vertical and lateral variations. This research highlights the potential of ML algorithms to streamline geotechnical assessments, potentially reducing reliance on extensive CPTs, and enhancing efficiency in offshore wind site evaluations. Ongoing efforts aim to refine these methodologies and expand their applications, for example the introduction of pre-stack Relative Extended Elastic Impedance (rEEI) for predicting Gmax, although further calibration is needed due to the lack of Cone Penetration Test (CPT) data in the area.

Machine learning based soil property prediction: A Quantitative Ground Model building approach based on 3D UHRS

Introduction

The world is the midst of a major energy transformation, with the focus in cutting carbon emissions to deliver on net-zero targets, while at the same time building a reliable system that meets the worlds ever-growing demands for energy. As an example, the UK government plans to reach 50 GW of offshore wind capacity by 2030, up from 15 GW in 2024 (OEUK, 2024). To meet this ambitious target, it is crucial to implement advanced technologies that can accelerate building accurate and robust 3D Quantitative Ground Models (QGM). This study investigates the integration of 3D Ultra High-Resolution Seismic (3D UHRS) data with Quantitative Interpretation (QI) methods for soil property prediction to streamline the wind farm site characterization process.

In this study, we propose two approaches. First, where only post-stack 3D UHRS data is available, we present a workflow implemented in the Ten Noorden van de Waddeneilanden (TNW) Wind Farm Zone (Figure 1) that predicts cone resistance (qc) from acoustic impedance using Machine Learning (ML). Additionally, as the quality of 3D UHRS data acquisition and processing improves, QI techniques can be extended to pre-stack analysis to estimate geotechnical parameters such as soil types (Dalgaard *et al.*, 2024) and small strain shear modulus (Gmax) (Carpentier *et al.*, 2024). To achieve this, we explore the integration of pre-stack data from a recent near-surface survey conducted in 2024 in the North Sea for relative inversion of Gmax.

This integration of 3D UHRS data with QI methods will allow the offshore wind sector to use a more targeted approach in the design and acquisition of geotechnical Cone Penetration Test (CPT) campaigns, an optimized foundation design and enhance the efficiency of offshore wind developments.

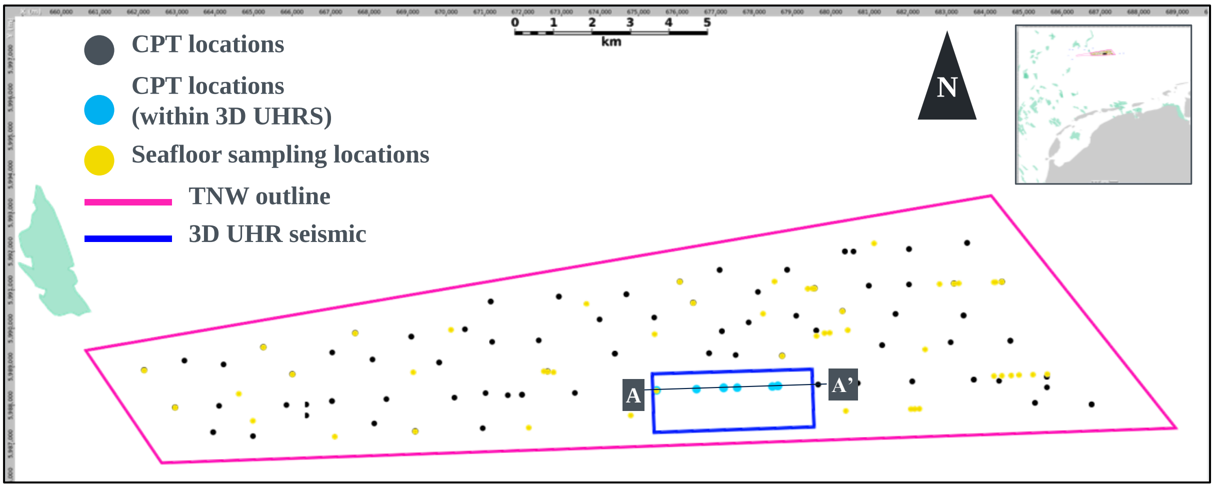

***Figure 1** Basemap showing outline of the TNW site and outline of the 3D UHRS data available for this study. Black dots represent CPT locations and yellow dots seafloor samplings. CPT locations used for training the ML model are in light blue. Prediction results are presented along profile A-A'.*

Post-stack ML methodology

Over the TNW area, only a full stack is available from the 3D UHRS data; this volume was used to derive a 3D acoustic impedance (Ip) volume (Polyaeva *et al. 2024*), and an envelope seismic attribute. Gradient Boosted Tree method, also known as XGBoost (Chen and Guestrin, 2016) was chosen to predict cone resistance, this algorithm offers two key advantages over more traditional Neural Network (NN) methodologies: it can handle missing input data (a typical problem in CPT logs, as they are not always continuously recorded) and it is less prone to over-fitting. Two different training models (Figure 2) were used for predicting cone resistance:

- **ML Model 1:** this model utilizes two inputs -envelope and Ip traces extracted at CPT locations- to predict cone resistance.
- **ML Model 2:** originally extracted envelope and Ip traces are shifted one sample up and down and used to train the model for prediction of cone resistance. Shifting the original attributes up and down, the space of training data is effectively increased from two attributes to six and now the ML model can account for potential mismatches in the time-depth conversion.

Using extracted seismic traces as input for prediction of a property such as cone resistance is not excepted from challenges and sources of uncertainties in the training when attempting to reconcile seismic and CPT datasets. First, seismic attributes are extracted along the CPT location, assuming a straight deviation, which may not always be the case. Additionally, without elastic logs available at the CPT site, it is not possible to perform an amplitude well-to-seismic tie, limiting QC of local time-depth conversion. Seismic attributes were converted to depth using root mean square (RMS) seismic velocities resampled to 1 ms. Finally, it is important to note that seismic and CPT measurements are acquired under different stress regimes, potentially affecting the reliability of the integrated data.

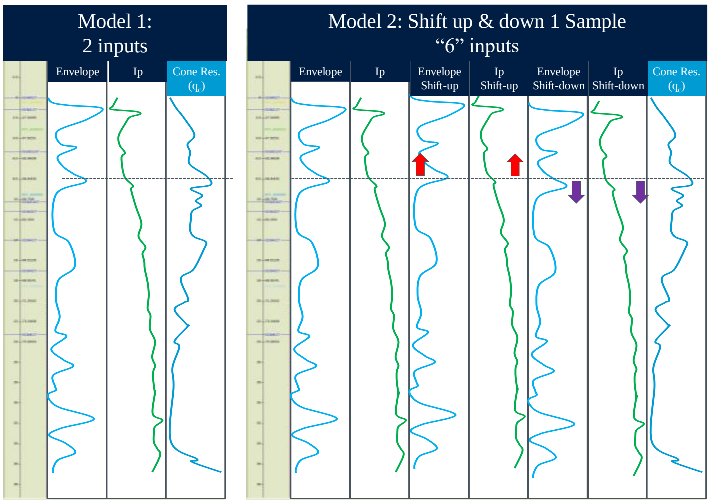

Figure 2 *Summary of two ML training datasets used in the study. By shifting upwards and downwards one sample envelope and inverted Ip it's possible to increase the number of inputs to train the model while proving the model a potential way for compensating for time-depth scaling mismatches.*

Post-stack ML results

To provide a frame to compare ML predictions, cone resistance results using a 1D correlation via linear regression between cone resistance and acoustic impedance from Polyaeva *et al. 2024* are also included and discussed in Figure 3. In general, results show good correlation between synthetic and measured cone resistance at CPT locations, still, some differences can be highlighted. First, the linear regression method required the window of interest to be split into two intervals (shallow and deep intervals) due to the varying trends in the cone resistance with increasing depth. In contrast, the ML methods effectively handled these varying trends. Furthermore, both ML methods identify zones of high cone resistance in the deep interval where high seismic velocities below CPT control are observed. Notably, Model 2 predicts higher values of cone resistance at the base of the tunnel valley compared to Model 1.

Both methodologies show more detailed lateral variations compared to existing model-driven approaches in the area, as they rely in the seismic amplitudes to provide the lateral variations that traditional geostatistical methods might overlook. However, heavily relying on seismic data requires a high degree of trusts in the fidelity of the seismic amplitudes, thus requiring the maximum care when acquiring and processing 3D UHRS data.

Pre-stack Relative Extended Elastic Impedance (rEEI)

Relative Extended Elastic Impedance (Went, Hedley and Rodgers, 2023), also referred to as rEEI, is a seismic inversion method originally devised to exploit a globally applicable rock property model. Initially devised to estimate the optimal rotation angle in the Ip and Gradient Impedance (GI) space for distinguishing lithology and fluid types using Amplitude Versus Angle (AVA) attributes, rEEI has since been adapted to predict various elastic properties, such as: shear impedance, bulk modulus, shear modulus (Gmax), and Poisson's ratio (Went 2025). Figure 4 illustrates the results of this seismic inversion technique applied to a recently acquired dataset in the North Sea. Ideally, the suitability of seismic data should be tested by comparison to borehole and CPT measurements. However, since no CPT data is currently available for calibration in this area, these results remain tentative until further calibration can be performed.

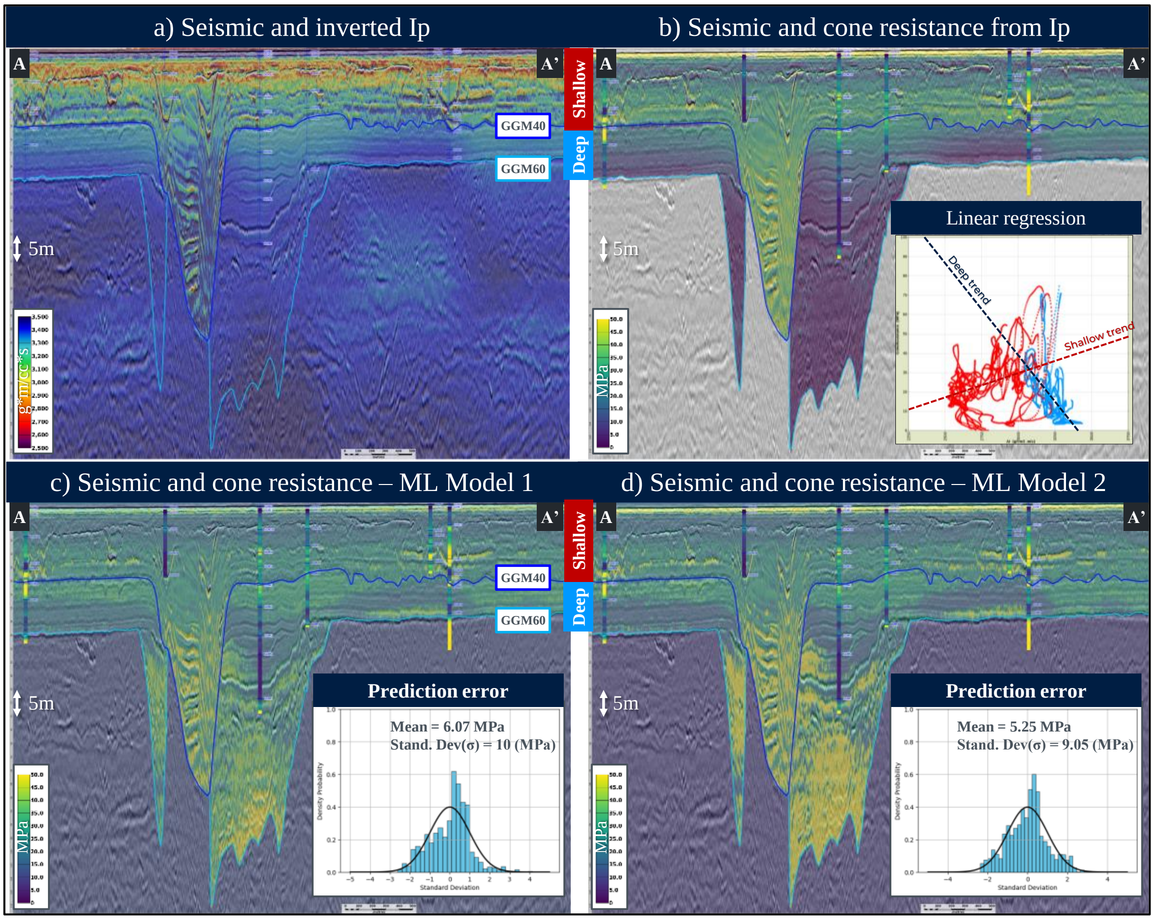

Figure 3 *(**a**) absolute acoustic impedance; (**b**) predicted cone resistance from the linear regression; (**c**) cone resistance prediction using ML Model 1; (**d**) cone resistance prediction using ML Model 2. The linear regression method does not manage to capture varying trends within the tunnel valley (deep interval), and there is not enough statistically meaningful information for splitting the deep window into another interval. Breaks in trends are well-handled by both ML models.*

Conclusions

In conclusion, this study demonstrates the efficacy of implementing machine learning algorithms for direct estimation of cone resistance (qc) from 3D UHRS data in the TNW area. The successful application of this method highlights the potential for this type of data to characterize near-subsurface strength properties, potentially reducing the reliance on extensive Cone Penetration Tests (CPTs) and borehole sampling campaigns. The method offers the potential for significant cost reductions and improved time efficiencies in both fixed and floating wind site assessments, representing a substantial advancement in geotechnical site characterization techniques for offshore wind farm development.

The implementation of QI workflows for QGM is still in its early stages, therefore ongoing research efforts are focused on further refining and expanding the capabilities of this methodology. Current developments include optimizing 3D UHRS survey design and acquisition parameters, enhancing processing and imaging techniques for improved data quality and amplitude preservation, and constructing robust velocity models for pre-stack depth migration. Additionally, efforts are underway to integrate advanced QI techniques for pre-stack inversion, such as rEEI, extend predictions to additional soil properties like Gmax, and improve the reconciliation between CPT and seismic domains. These advancements, coupled with a rigorous approach to minimizing and quantifying uncertainties throughout the workflow, aim to establish a comprehensive and reliable framework for wind site characterization.

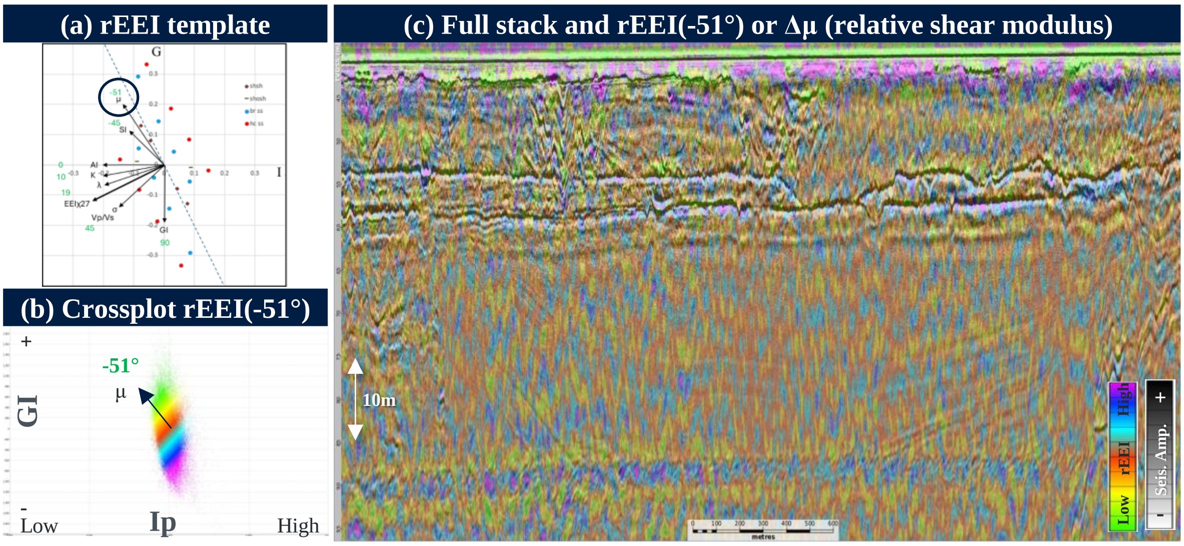

Figure 4 ***(a)*** *template showing the rotation angle to generate different elastic properties. The angle to estimate $\Delta\mu$ (relative change in shear modulus or Gmax) corresponds to rEEIχ = -51° based on a generalized rock physics model (after Went 2025);* ***(b)*** *AI-GI cross plot coloured by rEEIχ-51 (relative change in μ or Gmax);* ***(c)*** *stack reflectivity co-rendered with inverted rEEIχ-51 (relative Gmax).*

Acknowledgements

TNW input data shown in this abstract are made publicly available by Netherlands Enterprise Agency (RVO) and the authors would like to thank RVO as Creative Common 4.0 CCBY SA license holder.

References

Carpentier, S., Atkins, S., Boullenger, B., Vandeweijer, V., Peuchen, J. and Murali, M. [2024]. AI-Based Gmax Prediction from Seismic Data with Improved Confidence by Improved Inputs. *Fifth EAGE Global Energy Transition Conference & Exhibition (GET 2024)*, Extended abstract.

Chen, T. and Guestrin, C. [2016] XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM and SIGKDD international conference on knowledge discovery and data mining*, August, 785-794.

Dalgaard, E., Hansen, H.J., Horn, F., Jakobsen, A.F., Kuppens and S., Stuyts, B. [2024]. Energy Island - CPT interpretation and soil classification from UHRS pre-stack data. *Fifth EAGE Global Energy Transition Conference & Exhibition (GET 2024)*, Extended abstract.

NGI [2022]. Ten noorden van de Waddeneilanden Wind Farm Zone Integrated Ground Model. Doc No. 20190798-04-R. Client RVO

Offshore Energies UK [2024]. Offshore wind insight. May 2024.

Polyaeva, E., Pernin, N., Caselitz, B., Ruiz, R. and Lee, N. [2024]. Soil Properties Prediction Based on 3D Ultra-High Resolution Seismic: A Data Driven Inversion Workflow. *Fifth EAGE Global Energy Transition Conference & Exhibition (GET 2024)*, Extended abstract.

Went, D. [2025]. A graphical approach to determine the relationship between intercept, gradient and the common seismic rock properties: global model and application. *The Leading Edge* (in press).

Went, D., Hedley, R. and Rodgers, J. [2023]. Screening for AVA anomalies in siliciclastic basins: testing a seismic inversion method in the Mississippi Canyon, Gulf of Mexico. *First Break*. 41(9), 75-81.