Proceedings of ISFOG 2025
5TH INTERNATIONAL SYMPOSIUM ON
FRONTIERS IN OFFSHORE GEOTECHNICS
Nantes, France | June 9-13 2025

ISBN 978-2-85782-758-0

A review of benefits and pitfalls for prediction of geotechnical parameters in offshore wind ground models using synthetic CPTs

G. Michel*
Venterra Group, London, United-Kingdom

M. Coughlan
University College Dublin School of Earth Sciences, Dublin, Ireland, Country

R. Stanca, P. Narożny, L. Trindade and T. Joseph
Venterra Group, London, United-Kingdom

**guillaume-michel@venterra-group.com (corresponding author)*

ABSTRACT: Synthetic cone penetration test (CPT) profiling recently appeared as a buzzword in the geoscience community. While no unanimous definition exists, synthetic CPTs are often defined as predicted geotechnical parameters based on auxiliary variables. These synthetic datasets notably attracted the attention of the offshore wind sector, with the opportunity to characterise subseabed conditions over large areas where no prior measured data exist. These tools are foreseen as the next generation of tools for early site characterisation and de-risking, but also to support conceptual and detailed design of foundations, as well as cable routing and burial. The methods to generate synthetic parameters vary from basic spatial interpolation to complex 3D interpolation, or machine learning algorithms. The scope of methods is nearly as large as the number of published studies on synthetic CPTs and each method presents advantages and limitations, dependent on dataset availability and quality, amongst other factors, which must be understood. The application of such methods must be considered and adapted to the study objectives. This paper aims to provide an overview of the methods employed to predict geotechnical parameters, as well as benefits and limitations of synthetic CPTs and geotechnical parameter generation. This paper intends to inform potential end-users on how to assess the applicability of different synthetic CPT prediction methods and understand their limitations, from early site characterisation to foundation design and cable burial.

Keywords: Offshore Wind, cone penetration testing, ground models, synthetic parameters.

1 INTRODUCTION

The development of offshore wind sites heavily relies on robust ground condition characterisation. This is key to assess project feasibility, especially during the early phases, or to support detailed design during later phases. This characterisation is based on the integration of geophysical and geotechnical data. Geophysical data are usually acquired across the entire site with a dense coverage, but with limited vertical resolution and limited sediment property information. Conversely, geotechnical data provide stratigraphic and lithological information with a very high resolution but only for discrete locations. Along with boreholes, Cone Penetration Testing (CPT) is usually the chosen tool for geotechnical investigations which provides a continuous multi-parameter soil profiling method.

While CPT investigations represent a reliable source of high-resolution information, numerous CPTs are required to obtain a detailed understanding of ground conditions, and the acquisition of many CPTs represents a significant commitment in terms of cost and time.

Recently, methods have been developed to derive predictive geotechnical parameters, so-called ‘Synthetic CPTs’ (Figure 1) (Forsberg et al., 2017; Sauvin et al., 2019; Siemann et al., 2024). As part of several studies, geo-statistical methods and machine learning were used to reconstruct CPT profiles from limited measurements (Wang and Zhu, 2016). Interest in synthetic CPTs rapidly increased during the last decade due to the need of advancing offshore wind developments with limited initial data. Methods to derive synthetic CPT parameters vary from 1D (Wang and Zhu, 2016) and 2D (Carpentier et al., 2021; Sauvin et al., 2019; Siemann et al., 2024) to more complex 3D interpolation methods (global or local per geological layers, e.g., NGI, 2022). In addition to geo-statistical methods, different types of machine learning algorithms were used to derive synthetic parameters, e.g., Random Forest, Support Vector Machines, and different types of Neural Networks (Carpentier et al., 2021; Sauvin et al., 2019; Shoukat et al., 2023).

Overall, synthetic CPTs represent an appealing alternative to reduce the cost of site investigations. They are often seen as the ultimate tool for replacing extensive CPT data acquisition but present many pitfalls.

This paper aims to provide an overview of the methods employed in literature to predict geotechnical parameters, as well as a summary of the associated benefits and limitations of synthetic CPTs and geotechnical parameter generation.

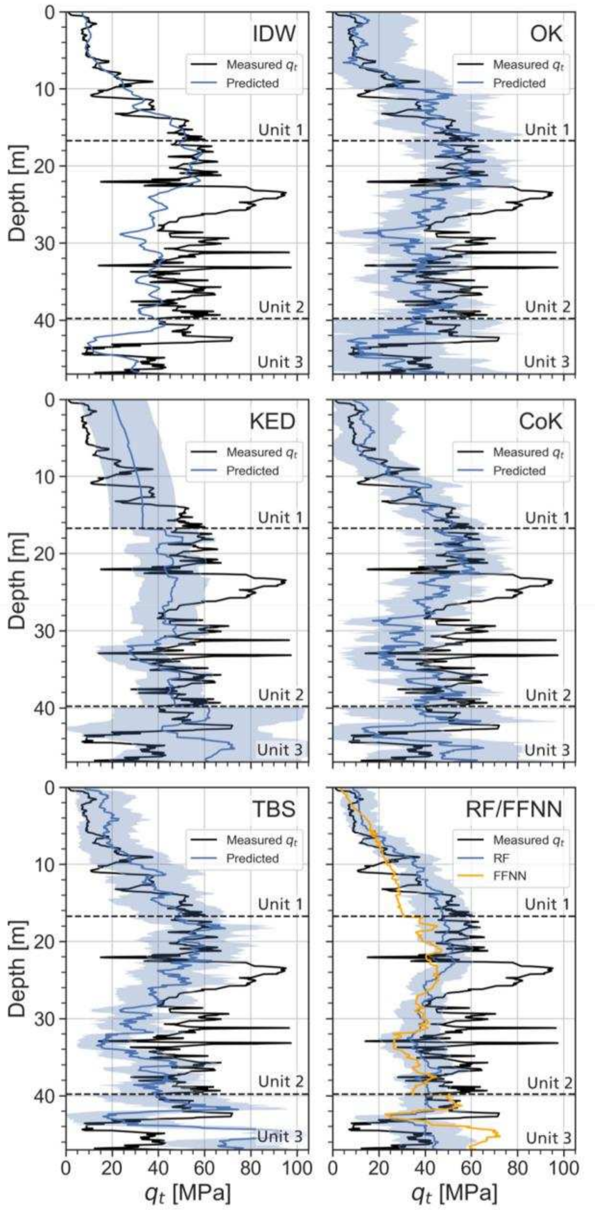

Figure 1 Example of synthetic CPT predicting corrected cone resistance (qt) using Inverse Distance Weighting (IDW), Ordinary Kriging (OK), Kriging with External Drift (KED), Collocated Kriging (CoK), Turning Band Simulation (TBS), Random Forest (RF) and Feed Forward Neural Network (FFNN) (Siemann et al., 2024)

This paper also intends to inform potential end-users on how to assess the applicability of synthetic CPTs and understand limitations, with a specific focus on the different steps of a windfarm site development, from early site description to turbine foundation design and cable routing.

2 METHODS SUMMARY

Amongst the large number of methods used to generate synthetic CPTs, two main categories can be distinguished: geo-statistical and machine learning algorithms. It is important to note that the methods from these two categories are not incompatible and can be used together. Furthermore, some methods can integrate auxiliary data, such as geophysical data (e.g., seismic stratigraphy, attributes or results from seismic inversion) (NGI, 2022; Sauvin et al., 2019; Vardy et al., 2023) (Figure 2).

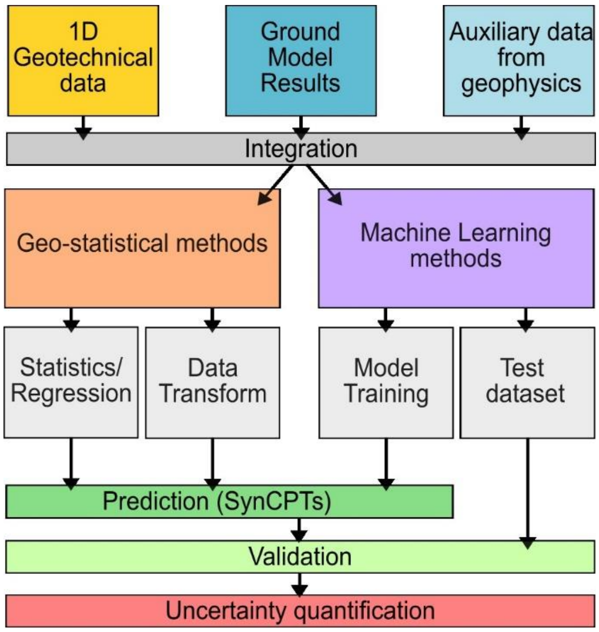

Figure 2 High-level workflow schematic for producing Synthetic CPTs from discrete 1D geotechnical data, supported by the integration of ground model results and auxiliary data. The two different paths correspond to the two main categories of methods, including geo-statistical and machine learning ones.

2.1 Geo-statistical algorithms

Geo-statistics was amongst the first methods that have been explored to generate synthetic CPTs (Forsberg et al., 2017). All geo-statistical methods rely on a similar principle: the integration of discrete data to spatio-temporally describe and model data, predict at unsampled points and evaluate uncertainties of the estimated parameters. Most of these geo-statistical methods have been developed for other purposes, notably including ore deposits mapping (e.g., Matheron, 1963) and petrophysical modelling. Furthermore, the geo-statistical methods are versatile as they can be applied to 2D or 3D problems with adaptations of their formulas.

The geo-statistical methods encountered in literature include natural neighbours, minimum curvature, spline, inverse distance weighting, and kriging interpolations (NGI, 2022). Many variations of each method exist, but the most diverse remains kriging, with multiple types of kriging available, e.g., simple, ordinary, universal, and collocated (e.g., NGI, 2022).

2.2 Machine learning algorithms

The application of machine learning algorithms rapidly increased in many fields to deal with extensive and complex datasets. With their rapid development and the democratization of their use, machine learning methods rapidly became a tool of choice to predict geotechnical parameters.

As per the geo-statistical methods, many different types of machine learning algorithms exist. Machine learning algorithms commonly encountered in publications about synthetic CPTs correspond to supervised learning methods and include, for example, Bayesian Machine Learning (BML) (Wang and Zhu, 2016), Artificial Neural Networks (ANN) (Sauvin et al., 2019; Shoukat et al., 2023), and Convolutional Neural Networks (CNN) (Carpentier et al., 2021).

3 BENEFITS AND LIMITATIONS

Based on a literature review, the results of several studies aimed at generating synthetic CPTs are presented in Table 1. The different methods used in these studies are summarised with their respective advantages, limitations, and some key reference examples. Studies which focused on the reconstruction of 1D synthetic CPTs were not included in Table 1.

According to multiple studies (Table 1), a significant limitation inherent to all methods is the vertical resolution of synthetic CPTs, which is more limited compared to in-situ data. Most of the time, this limitation is tied to the need for upscaling CPT data to simplify high vertical geotechnical variability, reduce computational time or to align with the resolution of auxiliary information such as parameters derived from seismic data (NGI, 2022; Sauvin et al., 2019).

It is difficult to assess the uncertainties and reliability of the different methods as the latter may vary depending on input datasets (quality, density, type, etc.), as well as the application (see section 4). Nevertheless, a general assessment of the two main categories of methods can be provided based on the results from the literature review.

In general, geo-statistical methods come with higher uncertainties, mostly associated with the methods themselves. Geo-statistical methods cannot deal effectively with non-stationarity and are intrinsically limited. Additionally, any uncertainties in the datasets are likely to be carried over into the analysis and results. However, these methods usually come with several metrics allowing to assess their performance, hence assuring a relatively high level of reliability. For the second category of methods, based on machine learning algorithms, most of the studies show an apparent lower level of uncertainty in their results. It is important to note that this observation might be biased due to the extensive volumes of data included in the datasets used in this type of studies. Generally, the reliability of the machine learning methods is on par with that of geo-statistical methods, although the validation methods and metrics can be different. While no comment was provided in that sense in the studies reviewed here, a consensus on machine learning methods is that it is not always straightforward to understand why a model performs well or why it underperforms, thus possibly reducing the reliability of these methods.

4 APPLICATION OF SYNTHETIC CPTS TO OFFSHORE WIND

In offshore wind farm development, site characterisation and ground models are constantly reassessed and improved with the integration of additional CPTs. Synthetic CPTs can be used to virtually increase geotechnical information density at a lower cost, thus representing an attractive alternative to in-situ measurements. However, as stated in multiple studies, synthetic CPTs can be considered as fit-for-purpose at some stages but cannot substitute for in-situ measurements (Sauvin et al., 2019; Vardy et al., 2023).

To generate synthetic CPTs, the adequate method should be decided based on available datasets and output requirements. When assessing the applicability of synthetic CPTs, other factors should be considered, including data type, density, and quality.

Table 1. Summary table of literature review for synthetic CPTs.

Methods	Benefits	Limitations	References
Natural Neighbour, Minimum curvature, Spline, Inverse Distance Weighting	- Easy to implement. - Statistically robust. - Require limited number of variables.	- Simplistic. - Low vertical resolution. - Heavily dependent on distances between in-situ locations.	Forsberg et al., 2017; NGI, 2022; Rahman et al., 2021; Sauvin et al., 2019
Kriging	- Able to predict finer variations both horizontally and vertically. - Different Kriging methods. - Error quantification.	- Low vertical resolution. - Dependent on spatial distribution and ability of sampling to capture geotechnical variations.	Coughlan et al., 2023; He et al., 2022; Liu et al., 2021; NGI, 2022; Rahman et al., 2021; Siemann et al., 2024; Xie et al., 2022
Collocated Kriging	- Same as Kriging. - Integration of auxiliary variables to 'inform' Kriging where input is limited.	- Same as Kriging. - Requires certain levels of correlation between variables.	Sauvin et al., 2019; Siemann et al., 2024; Xie et al., 2022
Bayesian Statistics	- Reconstruct signal from sparse sampling points. - Non-parametric method able to deal with non-stationarity.	- Method is probabilistic. - Struggles to capture local variations. - Outliers would be difficult to predict.	Tian and Wang, 2023; Wang et al., 2019; Zhao and Wang, 2020
Random Forest	- Can be used as a 'classifier' or a 'regression' tool. - Possibility to integrate multiple auxiliary variables.	- Requires enough 'samples' to train the model. - Variations not captured in any of the auxiliary variables might result in erroneous predictions	NGI, 2022; Rauter and Tschuchnigg, 2021; Siemann et al., 2024
Neural Networks (Artificial NN and Convolutional NN)	- Higher fidelity (major and minor variations). - Possibility to integrate multiple auxiliary variables (multi-attribute regression).	- Requires enough 'samples' to train the model. - Parametrization can be complex and time consuming (e.g., architecture and hyperparameters finetuning).	Carpentier et al., 2021; NGI, 2022; Sauvin et al., 2019; Shoukat et al., 2023; Vardy et al., 2023

In fact, during the early phases of projects, datasets are usually limited to a few geotechnical locations and seismic lines, often sourced from public databases or previous investigations. The quality of these historical datasets is usually good enough for a preliminary interpretation but not suitable for the generation of synthetic CPTs, especially in complex geological environments where significant spatial heterogeneities can occur (e.g., formerly glaciated continental shelves) (Figure 3). For any application, a proper assessment of the required precision, accuracy, and uncertainties of the prediction must be performed.

For preliminary site characterisation, the objective is to get an initial understanding of the geological units present at the site, as well as an overview of their respective geotechnical properties.

Usually, data available during preliminary site investigations are limited, so methods relying on a minimum amount of input data, such as Kriging or Neural Networks may not be adequate (Figure 3). However, simple interpolation methods, providing more simplistic and low-resolution predictions, might be fit for purpose and provide an overview of the geotechnical variability at the site (Forsberg et al., 2017; NGI, 2022) (Figure 4).

For survey planning, the context is usually like the preliminary site investigation stage, with geotechnical data sparsity being relatively high. Simpler interpolation methods, along with the geotechnical variability overview they can provide, can be a valuable tool to plan or revise additional target locations.

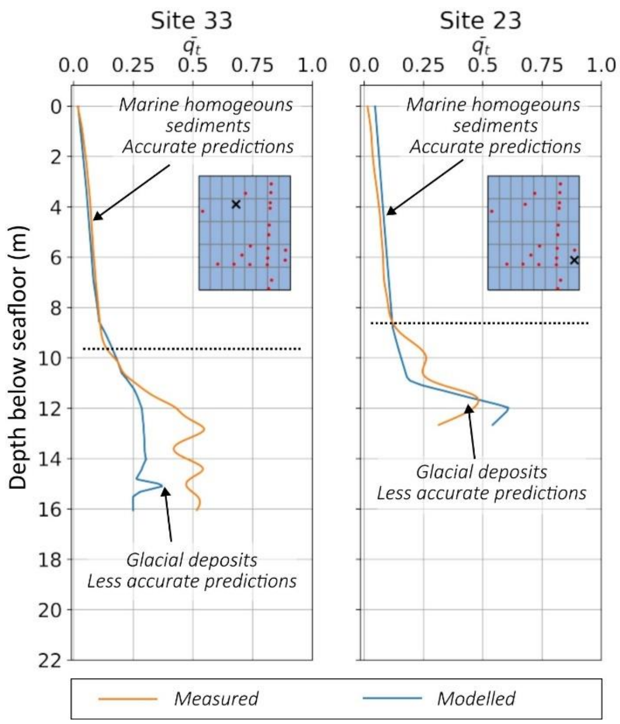

Figure 3 Example of challenges for SynCPT predictions, with predicted (blue) corrected cone resistance against measured parameter (orange). Accurate predictions are made in homogenous sediments, but the predictions are less accurate in complex glacial deposits. Note that the corrected cone resistance values were normalized to 0-1 range. Modified from Shoukat et al. (2023).

During later phases of site development, higher resolution, precision, and accuracy would be required. Usually, more geotechnical data are available so more complex methods such as Kriging and Neural Networks, relying on more inputs, can be used (Figure 4). These methods can be applicable to the planning of detailed surveys and can be considered for conceptual design. Furthermore, if these methods are supplemented by the integration of geophysical data, the results from synthetic CPT predictions will be more robust (NGI, 2022; Sauvin et al., 2019; Vardy et al., 2023).

For conceptual foundation design, synthetic CPTs can provide valuable information under the form of best case, worst case, and best estimate, which can help identify design challenges before the detailed foundation design phase.

For detailed design, the requirements in terms of precision, accuracy, resolution, and confidence in the results are very high. However, even the most complex methods to generate synthetic CPTs, with integration of auxiliary information, will always have a lower vertical resolution compared to in-situ data and higher levels of uncertainty. This should not be acceptable at this stage of project development (Figure 4).

For cable routing and design, a very fine scale spatial mapping and a very high vertical resolution are required, but for a more limited depth range below the seabed. The use of synthetic CPTs for such applications has not yet been described or discussed in literature. While a very fine scale spatial mapping can be achievable, most of the methods currently available are not able to predict CPTs at the same vertical resolution as in-situ measurements. Therefore, the applicability of synthetic CPTs for cable routing and design requires further investigations.

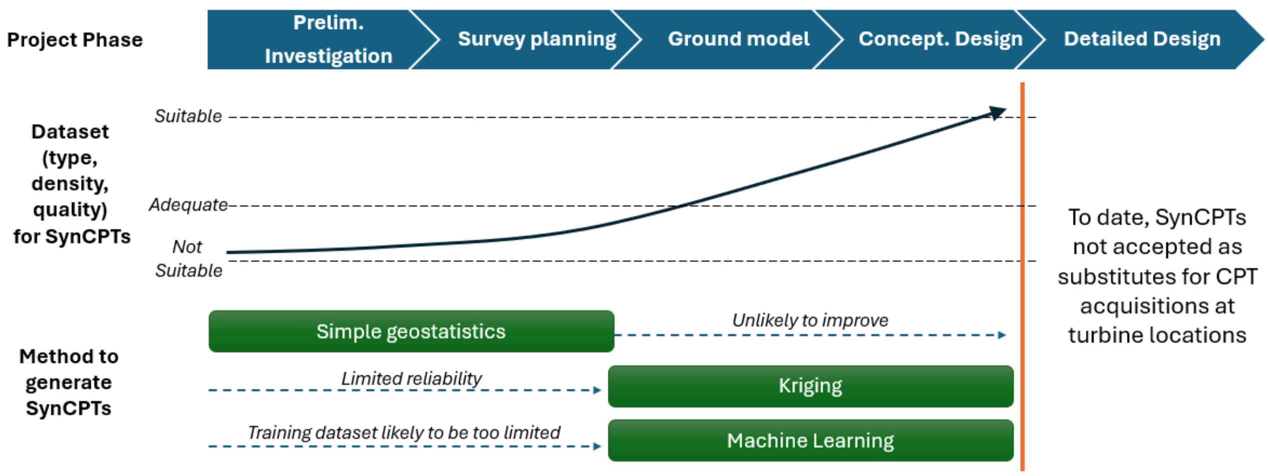

Figure 4 Schematic summary diagram of SynCPTs applicability for different OWF project development phases, with indication of qualitative dataset suitability and methods compatibility based on their respective limitations.

5 CONCLUSIONS AND RECOMMENDATIONS

With a large range of methods available, the prediction of CPT parameters is an appealing tool to virtually increase the density of geotechnical information, while minimising costs and time required for additional geotechnical data acquisition.

Depending on the requirements and available data type, quality and density, including supporting datasets such as geophysical ones, synthetic CPTs can be applied at different stages of offshore wind site development. The use of synthetic CPTs, as well as the method, must be assessed in terms of precision, accuracy, and uncertainties against project phase requirements.

A common pitfall is to consider synthetic CPTs as the ultimate tool. As stated in multiple studies, synthetic CPTs are incapable at present to replace in-situ measurements and the results from synthetic CPTs should always be considered with caution.

With the growth of interest for synthetic CPTs within academia and industry, the development of new methods but also reflexive studies on the limitations, precision, and accuracy will help strengthen the methodology and provide a better understanding of uncertainties, the latter being a requirement for any application within industry.

AUTHOR CONTRIBUTIONS STATEMENT

Guillaume Michel: Conceptualization, Data curation, Investigation, Writing- Original draft, Project administration. **Mark Coughlan**: Conceptualization, Investigation, Writing – original draft. **Roxana Stanca**: Validation, Writing – original draft. **Pawel Narozny**: Writing – original draft **Liliana Trindade**: Supervision, Writing- Reviewing and Editing. **Trevon Joseph**: Supervision, Writing-Reviewing and Editing.

ACKNOWLEDGEMENTS

The authors are grateful to the two anonymous reviewers for their valuable comments.

REFERENCES

Carpentier, S., Peuchen, J., Paap, B., Boullenger, B., Meijninger, B., Vandeweijer, V., Van Kesteren, W., Van Erp, F., 2021. Generating synthetic CPTs from marine seismic reflection data using a neural network approach. 2nd EAGE Workshop on Machine Learning 2021 8–10. https://doi.org/10.3997/2214-4609.202132008

Coughlan, M., Long, M., Michel, G., Desmond, C., 2023. Mapping Geotechnical Variability in the Shallow Subsurface: An Irish Sea Case Study, in: NSG2023 1st Conference on Sub-Surface Characterisation for Offshore Wind. European Association of Geoscientists & Engineers, pp. 1–5. https://doi.org/10.3997/2214-4609.202320037

Forsberg, C., Lunne, T., Vanneste, M., James, L., Tjelta, T., Barwise, A., Duffy, C., 2017. Stothetic CPTS from Intelligent Ground Models Based on the Integration of Geology, Geotectoics and Geophysics as a Tool for Conceptual Foundation Design and Soil Investigation Planning, in: Offshore Site Investigation Geotechnics 8th International Conference Proceedings. Society of Underwater Technology, pp. 1254–1259. https://doi.org/10.3723/OSIG17.1254

He, R., Li, J., Yang, S., He, B., 2022. Kriging analysis on CPTU data from offshore wind farm. Cone Penetration Testing 2022 961–966. https://doi.org/10.1201/9781003329091-144

Liu, Jinhao, Liu, Jinming, Li, Z., Hou, X., Dai, G., 2021. Estimating CPT Parameters at Unsampled Locations Based on Kriging Interpolation Method. Applied Sciences 2021, Vol. 11, Page 11264 11, 11264. https://doi.org/10.3390/APP112311264

Matheron, G., 1963. Principles of geostatistics. Economic Geology 58, 1246–1266. https://doi.org/10.2113/GSECONGEO.58.8.1246

NGI, 2022. Ten noorden van de Waddeneilanden Wind Farm Zone Integrated Ground Model. Doc No. 20190798-04-R. Client RVO. https://doi.org/https://offshorewind.rvo.nl/files/view/82df9a3c-502c-4f7a-b89b-28d640782801/tnw_20221014_gm_ngi_integrated-ground-model-f.pdf

Rahman, M.H., Abu-Farsakh, M.Y., Jafari, N., 2021. Generation and evaluation of synthetic cone penetration test (CPT) data using various spatial interpolation techniques. Canadian Geotechnical Journal 58, 224–237. https://doi.org/https://doi.org/10.1139/cgj-2019-0745

Rauter, S., Tschuchnigg, F., 2021. CPT data interpretation employing different machine learning techniques. Geosciences (Basel) 11, 265.

Sauvin, G., Vanneste, M., Vardy, M.E., Klinkvort, R.T., Fredrik, F.C., 2019. Machine learning and quantitative ground models for improving offshore wind site characterization.

Proceedings of the Annual Offshore Technology Conference 2019-May. https://doi.org/10.4043/29351-ms

Shoukat, G., Michel, G., Coughlan, M., Malekjafarian, A., Thusyanthan, I., Desmond, C., Pakrashi, V., 2023. Generation of Synthetic CPTs with Access to Limited Geotechnical Data for Offshore Sites. Energies 2023, Vol. 16, Page 3817 16, 3817. https://doi.org/10.3390/EN16093817

Siemann, L., Masoudi, P., Maraka, R., Opris, R., Pande, Y., Römer-Stange, N., Morales, N., Mörz, T., 2024. Comparison of Different Prediction Methods to Derive Synthetic CPT Profiles - An Offshore Wind Farm Case Study from the German North Sea, in: 7th International Conference on Geotechnical and Geophysical Site Characterization. CIMNE, Barcelona. https://doi.org/10.23967/isc.2024.233

Tian, H., Wang, Y., 2023. Data-driven and physics-informed Bayesian learning of spatiotemporally varying consolidation settlement from sparse site investigation and settlement monitoring data. Comput Geotech 157, 105328. https://doi.org/10.1016/J.COMPGEO.2023.105328

Vardy, M.E., Sauvin, G., Vanneste, M., Klinkvort, R.T., Dyer, N.J., 2023. How many CPTs does it take to make a synthetic CPT?, in: Offshore Site Investigation Geotechnics 9th International Conference Proceeding. pp. 385–392. https://doi.org/https://doi.org/10.3723/WASQ1942

Wang, C., Zhu, H., 2016. Combination of Kriging methods and multi-fractal analysis for estimating spatial distribution of geotechnical parameters. Bulletin of Engineering Geology and the Environment 75, 413–423. https://doi.org/10.1007/s10064-015-0742-9

Wang, Y., Hu, Y., Zhao, T., 2019. Cone penetration test (CPT)-based subsurface soil classification and zonation in two-dimensional vertical cross section using Bayesian compressive sampling. https://doi.org/10.1139/cgj-2019-0131 57, 947–958. https://doi.org/10.1139/CGJ-2019-0131

Xie, J., Huang, J., Lu, J., Burton, G.J., Zeng, C., Wang, Y., 2022. Development of two-dimensional ground models by combining geotechnical and geophysical data. Eng Geol 300, 106579. https://doi.org/10.1016/j.enggeo.2022.106579

Zhao, T., Wang, Y., 2020. Interpolation and stratification of multilayer soil property profile from sparse measurements using machine learning methods. Eng Geol 265, 105430.