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

ISBN 978-2-85782-758-0

Exploring the use of machine learning to support anchor pile design at unseen seabed locations

Q. Liu*
The University of Western Australia, Perth, Australia

M. Bertolacci, M.F. Bransby, T. French, P. Watson
The University of Western Australia, Perth, Australia

**qianbiao.liu@research.uwa.edu.au (corresponding author)*

ABSTRACT: This study investigates the potential role of machine learning (ML) to support design of anchor piles under axial tension, specifically for the case where location-specific data is not available. The analysis is undertaken for an existing site on Australia's North West Shelf where CPT data does exist, with results generated using 'standard' design approaches compared to those based on soil strength predicted from ML. The results demonstrate that ML approaches can generate reasonable design predictions while also providing quantitative assessment of anchor reliability, potentially offering a practical method to advance offshore foundation design in the absence of location-specific data.

Keywords: Site characterisation; pile design; geostatistics; machine learning

1 INTRODUCTION

Piles are extensively employed as foundations for bottom-supported structures, or as anchors for floating structures for oil & gas and renewable energy developments. The design of offshore piles requires knowledge of the properties of the seabed at a given location, with design approaches often supported by cone penetrometer test (CPT) data because of (a) its detailed vertical profiling capabilities; (b) the in situ nature of the testing, meaning there is no sample disturbance to consider; and (c) the relevance of the cone penetration process compared to some (pile) design states. In order to select a suitable pile size to ensure sufficient in-place capacity, deterministic design therefore requires a cone tip resistance, e.g. the net cone resistance q_{net} design line representing the lowest likely soil strength. This design line is based on the target level of conservatism and is generally expressed as a probability of non-exceedance (PN) of the collected CPT data—for example, PN =10% or P10.

The determination of these design lines presents significant challenges in current practice. Traditionally, this process relies heavily on engineering judgment, though nowadays is more usually guided by statistical methods such as quantile regression (Uzielli et al. 2019). While these approaches have been widely adopted, they introduce substantial subjectivity into the design process, potentially compromising the consistent assessment of reliability levels (Cai et al. 2021).

The limitations of traditional methods are further amplified when design requirements necessitate soil property estimation at locations without direct measurements. In such scenarios, practitioners must rely on interpolation between (or extrapolation from) tested locations, introducing additional uncertainty into the design process. This situation may occur during project development phases when foundation locations (and layouts) change, or when assessing soil properties beneath large foundations where test locations are limited. The combination of sparse data and subjective interpolation methods may lead either to overly conservative designs (increasing project cost) or unsafe conditions if the spatial variability of soil properties is underestimated.

Recent advances in machine learning (ML) and data-driven techniques have shown promise in providing a more informative and accurate estimation of CPT profiles at untested locations (Phoon et al. 2022). Studies have demonstrated that data-driven methods can reliably estimate soil properties – and their uncertainty – using raw CPT data, offering improved profiles (Phoon et al. 2022; Wang et al. 2017). While these techniques show great promise in improving CPT data estimation, their application to generating design lines for pile foundation sizing remains unverified.

Given the above, this study investigates the applicability of two ML approaches in generating design lines for vertically loaded piles at untested locations, utilising a comprehensive CPT database from an offshore site on Australia's North West Shelf. We specifically examine how ML-generated design lines (and probabilistically generated sets of realisations) influence the minimum designed pile size both with and in the absence of location-specific data. This research therefore aims to demonstrate the potential of ML techniques in optimizing pile size under sparse geotechnical data conditions, contributing to more cost-effective and reliable offshore foundation design.

2 SITE CONDITIONS AND CPT DATA

A total of 60 CPTs were performed across a 625 km^2 site on Australia's North West Shelf, as shown in Figure 1. The sediments across the site comprised normally consolidated carbonate silts and muds interspersed with occasional thin sandy layers. To investigate the capability and applicability of ML methods in site investigation (SI) data interpretation for engineering design, a CPT located approximately at the center of the site was selected as the location of interest (for the design of a hypothetical tension pile), with the nearest available CPT data points situated more than 1.9 km away. For some design conditions, the data from the site-specific CPT was revealed, while it was hidden for other cases in order to also investigate the value of acquiring location-specific data.

The CPT data were recorded with an ISO-standard resolution of 0.02 m (ISO 2023) to depths of between 2 and 76 m. Figure 2 shows the profiles of net cone resistance (q_{net}) versus depth below mudline.

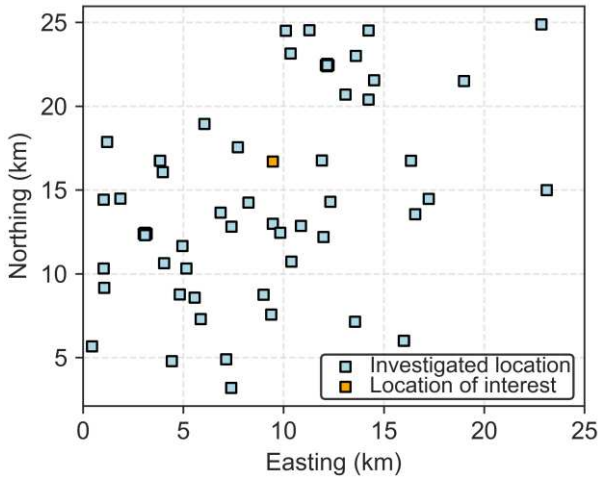

Figure 1. CPT locations (precise location not given due to confidentiality constraints)

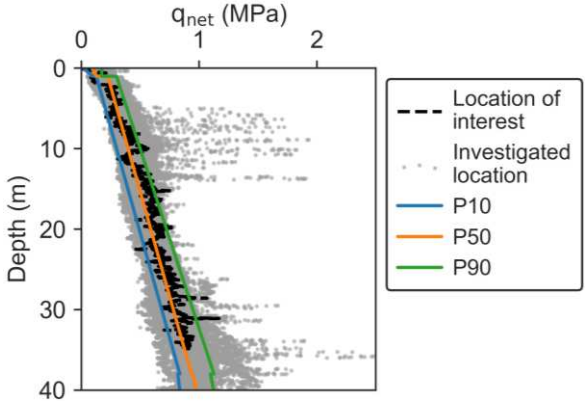

Figure 2. CPT measured net cone resistance profiles

3 DIFFERENT APPROACHES TO DETERMINE q_{net} DESIGN LINES

Two design approaches were adopted for pile sizing: load and resistance factor design (LRFD) and reliability-based design (RBD). Both use q_{net} profile with depth as the primary input to calculate foundation capacity. Consequently, the LRFD approach required definition of deterministic q_{net} design lines, whereas the reliability-based approach required generation of multiple q_{net} versus depth realisations.

For the LRFD approach, five cases were compared, each differing in the availability of field CPT data at the location of interest and the methods employed to generate the q_{net} design lines:

- Case 1: Following standard engineering practice, quantile regression (Uzielli et al., 2019) was applied to the existing dataset of 59 CPTs from the site, with the 10th percentile (P10) adopted as the design line, as shown in Figure 2. This approach does not utilise CPT data from the specific location of interest, but assumes that the soil conditions across the whole site are similar (i.e. there is one soil zone with consistent soil units across the site).
- Case 2: The design line was determined using a single CPT measurement conducted directly at the centre of the planned pile. This case assumes that an additional site investigation had been carried out to acquire CPT data at the location of interest, and that the engineer implicitly trusts the data at this location.
- Case 3 also assumed that additional CPT data had been acquired at the location of interest. However, here it was assumed that the engineer would not directly rely only on the data at the specific location, but also consider the overall variability of the site. As such a design line was selected by taking the midpoint between the P10

(derived from quantile regression of the 59 CPTs) and the measurements at the pile centre. As such, this decision involves an element of engineering judgement.

- Cases 4a and 4b: Two advanced machine learning techniques (BNN-BO and GSSLF) were employed to predict P10 profiles directly at the location of interest, assuming that location-specific data had not been acquired. These models were calibrated using the existing dataset of 59 CPTs only.

The two machine learning (ML) approaches were Bayesian neural network with Bayesian optimization (BNN-BO) (Liu et al. 2025a) and the geotechnical self-supervised learning framework (GSSLF) (Liu et al. 2025b). For these ML-based methods, the raw net cone resistance (q_{net}) data was first transformed to its natural logarithm $\log q_{net}$, assuming that the $\log q_{net}$ data at each spatial location follows a specific normal distribution. During training, the models attempt to learn the parameters of these data distributions. The trained models were then able to predict the mean and variance of the $\log q_{net}$ distribution at the untested locations (e.g., at the location of interest shown in Figure 1), and from these predicted statistics, the desired design percentiles, such as the 10th percentile (P10), can be directly obtained. Additionally, the estimated distribution also enables the generation of soil profile realisations at untested locations which was used for the RBD approach to anchor pile design.

4 PILE DESIGN PROCEDURES

4.1 Pile and problem geometry

An open-end pile subjected to uplift loading and installed at the location of interest (Figure 1) was studied. The designed pile was to have a fixed outer diameter $D = 2\,m$ and wall thickness $t = 25\,mm$ ($D/t = 80$). Consquently, the design exercise involved only selecting the required pile length L to fulfil in-place capacity requirements.

The uplift loading distribution $f(S)$ was assumed to be a Gumbel distribution, with its parameters derived from a 95th quantile (Q95) value of 1.57 MN and an assumed coefficient of variation of 0.3.

4.2 Pile capacity calculation method

The unfactored uplift capacity of the open-end pile, R was calculated using q_{net} data through the following semi-empirical formula (ISO 2016):

$$R = Q_s + Q_b + W'_{soil} + W'_{pile} \quad (1)$$

$$Q_s = \pi D \int_0^L \alpha_s \frac{q_{net}}{N_{kt}} dz \quad (2)$$

$$Q_b = \alpha_b q_{net} \frac{\pi D^2}{4} \quad (3)$$

$$W'_{soil} = \pi \frac{(D-2t)^2}{4} L\gamma'_{soil} \quad (4)$$

where Q_s is the uplift capacity due to skin friction; Q_b is the uplift capacity due to end bearing; W'_{soil} is the buoyant weight of soil plug inside the pile (assume fully plugged); W'_{pile} is the buoyant weight of pile; L is the pile length; D is the outer pile diameter; t is the pile wall thickness (defined by $D/t = 80$); α_s is an empirical adhesion factor (assumed = 0.4); q_{net} is the net cone resistance; N_{kt} is the cone factor (assumed = 13); α_b is the adopted end bearing factor (assumed = 0.2); and γ'_{soil} is the effective unit weight of soil (assumed = 6 kN/m^3).

4.3 Load and resistance factor design approach

For pile sizing with the LRFD approach, a partial material factor (γ_m= 1.25) was applied to the design cone resistance profiles in Equations (2) and (3), and a partial load factor (γ_L = 1.35) was applied to the calculated pile capacity. The design load was taken as the Q95 of the load distribution (1.57 MN) and the minimum foundation length (L) calculated to achieve the required partial factors for the given strength profile. The Q95 was selected as the characteristic load for the LRFD approach as it provides a suitably conservative design capacity (Ching and Phoon 2011).

4.4 Reliability-based design approach

With ML-based approaches being capable of estimating the statistical distribution of q_{net} at various depths for the unknown location, it is possible to support a reliability-based foundation design. This study therefore introduced two additional design cases, namely Case 5a (GSSLF) and Case 5b (BNN-BO).

For each realization of a q_{net} profile, the unfactored uplift capacity (R) for a specific pile length was determined. The results from all realizations were subsequently complied, allowing for the calculation of the capacity distribution in the form of a probability density function (PDF), $g\big(R(q_{net})\big)$, for a given pile length based on the collected data.

The estimated probability of failure p_f of a pile of a selected length then be evaluated as follows:

$$p_f = \iint_{R(q_{net}) \le S} f(S) g\big(R(q_{net})\big)\, dR(q_{net}) dS \quad (5)$$

In this study, the capacity distribution $g\big(R(q_{net})\big)$ was estimated by fitting a normal distribution to the

histogram of Monte Carlo-calculated $R(q_{net})$ values. The uplift loading distribution $f(S)$ was defined using a Gumbel distribution (as defined earlier). This then allowed the probability of failure of any given pile length (L) to be calculated using Equation (5). By repeating this exercise for a large range of pile lengths (with a constant outer diameter increment of 0.5 m), the required pile length to achieve different target probabilities of failure ($p_f = 10^{-3}, 10^{-4}, 10^{-5}$) were calculated.

Figure 3 illustrates an example probability of failure calculation for a pile with $L = 19.5$ m. The uplift loading distribution ($f(S)$) is defined by a Gumbel PDF with a location parameter of 0.926 MN and a scale parameter of 0.217. Using the GSSLF model, 10,000 realizations of q_{net} profiles were generated and the corresponding unfactored uplift capacities $R(q_{net})$ were computed using Equation (1). The capacity distribution, $g(R(q_{net}))$ was then estimated from Monte Carlo-simulated $R(q_{net})$ values. The probability of failure p_f was represented by the shaded area under the $g(R(q_{net}))$ curve where $R \leq S$, as schematically shown in Figure 3. As a result, for a pile with a length of 19.5 m, designed to resist a load of 1.57 MN, the estimated p_f was 0.974×10^{-3}.

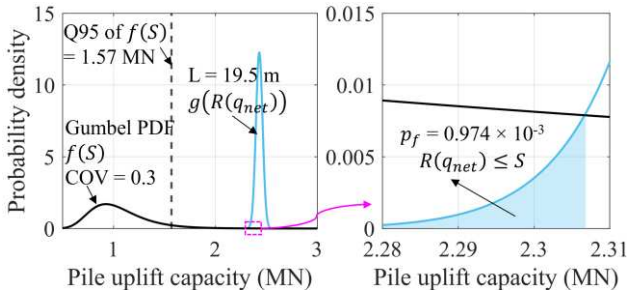

Figure 3. Example probability of failure calculation (pile with L = 19.5 using realisations from GSSLF

5 RESULTS

5.1 q_{net} design lines for LRFD cases

Figure 4 presents the q_{net} design lines derived from five different LRFD cases (Cases 1 to 4b). The baseline approach (Case 1) yields consistently lower q_{net} values across most depths, particularly above 25 m, which encompasses the typical pile design depth range in this study, compared to other design cases.

Not surprisingly, in cases with location-specific CPT data, the design line for Case 2 – based solely on a single CPT conducted at the center of the planned pile – exhibits notably higher q_{net} values at most depths compared to other design cases. In contrast, Case 3, which incorporates engineering judgment by integrating both location-specific CPT data and neighbouring site investigation data, provides a relatively conservative soil strength for the design line (i.e., has lower q_{net} values).

In the absence of location-specific data, the ML-predicted P10 design lines (Case 4a for GSSLF and Case 4b for BNN-BO) exhibit higher q_{net} values than the baseline (Case 1). Moreover, although no site-specific measurements are available for Cases 4a and 4b, their design lines align (somewhat) more closely with Cases 2 and 3 – where location-specific data are incorporated compared to the baseline. This demonstrates the potential of ML methods to effectively leverage neighbouring site investigation data for generating improved (less conservative) design lines at untested locations.

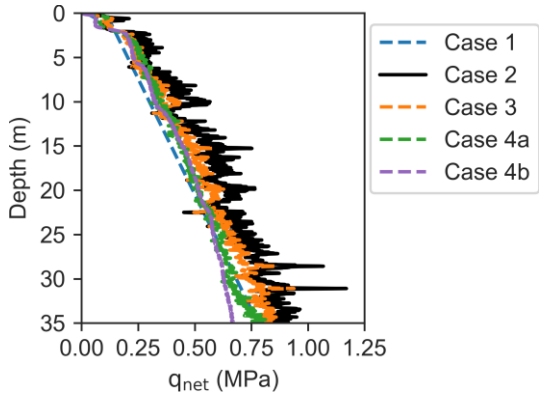

Figure 4. q_{net} design lines for the LRFD approach

5.2 q_{net} design lines for RBD cases

Figure 5 and Figure 6 show the q_{net} profile realisations from the ML models for the reliability-based design cases (Case 5a for GSSLF and Case 5b for BNN-BO). These generally exhibit trends similar to the unseen location-specific or original q_{net} data (shown as the black lines in the figures). The results shown in Figure 5b and Figure 6b indicate that almost all local variations of the original q_{net} profiles fall within the corresponding estimated 95% confidence interval, demonstrating further the reliability of the q_{net} profile realizations from ML models.

To evaluate the accuracy of the model predictions, Figure 5a and Figure 6a also show the 50% percentile profiles (median estimates, P50). To quantitatively compare the model performance, the mean absolute percentage error (MAPE) was calculated between estimated q_{net} profile and the unseen location-specific q_{net} profile. The MAPE is defined as follows:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \qquad (6)$$

where N is the number of q_{net} values at profiles; y_i is the measured data; $\hat{y}_i$ is the q_{net} value from the estimated P50 profile. Lower MAPE indicates better performance.

Results of quantitative comparison indicate that GSSLF achieves a MAPE of 11.60%, followed by BNN-BO with 12.68%, whereas quantile regression (Case 1) using the remaining 59 CPT profiles results in a higher (i.e. worse) MAPE of 14.49%. These findings indicate that the ML models more effectively leverage the neighbouring CPT data to produce more realistic and accurate q_{net} predictions compared to the quantile regression method. More importantly, the ML methods also estimate the statistical distribution of q_{net} at a location, thereby enabling reliability-based design, while quantile regression cannot provide the complete probabilistic characterization of q_{net}.

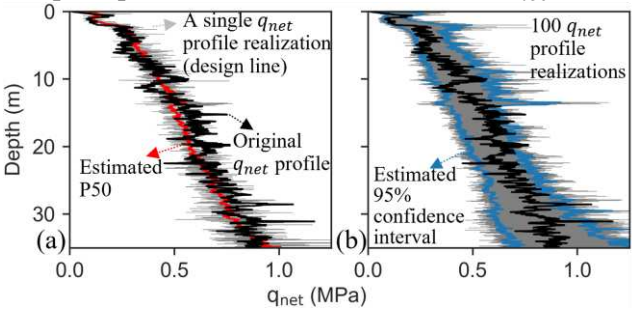

Figure 5. q_{net} design lines for the location of interest generated from the GSSLF method: (a) an example of a single q_{net} profile realization and (b) 100 q_{net} profile realizations.

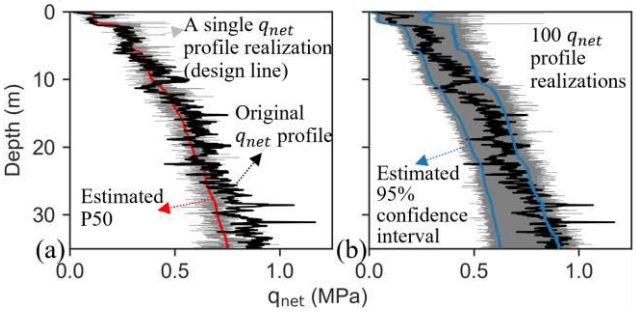

Figure 6. q_{net} design lines for the location of interest generated using the BNN-BO method: (a) an example of a single q_{net} profile realization and (b) 100 q_{net} profile realizations

5.3 Designed pile length results

The results in terms of pile length for each case are presented in Figure 7. Not surprisingly, results of the LRFD approach indicate that cases with location-specific CPT measurements (Case 2 and Case 3) yield shorter pile lengths compared to cases without such measurements (Case 1, Cases 4a and 4b), although the acquisition of additional location-specific CPT measurements would inevitably increase site investigation costs (especially if requiring a standalone campaign). Moreover, it is noteworthy that Case 2, which determined the design line using a single CPT profile at the centre of the planned pile location, yielded the shortest pile length of 19.5 m. While this method may seem straightforward, it may lead to unconservative design, especially for large piles, due to its inability to fully capture the spatial variability of soil properties (Cai et al. 2021). Therefore, the design outcomes from Case 2 should be interpreted with caution, considering the inherent risks associated with potentially insufficient site characterization.

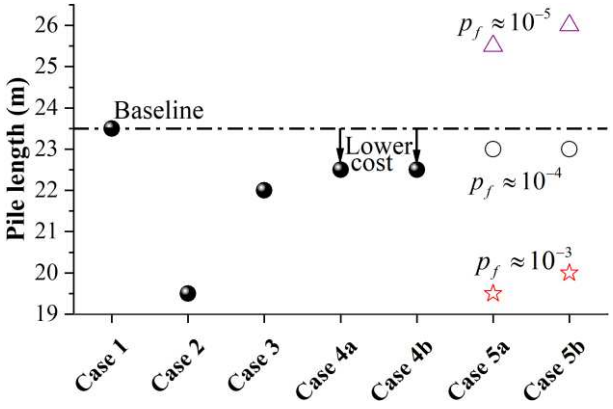

Figure 7. Designed pile length for different cases

In the absence of location-specific data, the ML-based approaches demonstrated less conservative and (somewhat) more cost-effective designed pile lengths. Specifically, both ML methods (GSSLF (Case 4a) and BNN-BO (Case 4b)) achieved a reduced length of 22.5 m, compared to the baseline quantile regression approach (Case 1) that resulted in a pile length of 23.5 m. Notably, the ML-based designs closely align with the results from Case 3, highlighting the potential of ML approaches to effectively leverage neighbouring site investigation data for offshore pile design.

The RBD, incorporating spatial variability of seabed soils allows for optimization of pile lengths based on different target probabilities of failure (p_f). As shown in Figure 7, the GSSLF and BNN-BO approaches yielded pile lengths of 19.5 m and 20.0 m respectively for a target $p_f = 10^{-3}$. When a higher reliability level is specified (target $p_f = 10^{-5}$), the required pile lengths increase to 25.5 m and 26.0 m respectively. Deterministic design methods, such as LRFD, rely on code-specified load and resistance factors but do not explicitly convey information on the probability of failure. In contrast, RBD directly incorporates uncertainties in q_{net} distribution. Consequently, the ML-based RBD approaches can play a complementary role to LRFD by providing quantitative assessment of reliability levels (e.g.,

achieved p_f), thereby enabling risk-informed decision making in foundation design, particularly in the absence of location-specific data.

6 CONCLUSIONS

This study has investigated the potential of ML models to generate q_{net} design lines for the analysis of vertically loaded piles at untested locations, utilising a CPT database from an offshore site on Australia's North West Shelf. Results indicate that ML approaches can effectively leverage neighbouring site investigation data to generate design lines for offshore pile sizing, particularly in the absence of location-specific measurements. Moreover, the ML-based RBD approaches also can provide quantitative assessment of reliability levels, thereby enabling risk-informed decision making in foundation design. These findings demonstrate the practical value of ML methods in providing robust initial design estimates for various scenarios, such as when foundation layouts change during project development phases.

However, it should be noted that ML-based model performance may vary depending on the volume and relevance of SI data availabile and the magnitude of the spatial variability of the seabed properties arising from complex geological and environmental factors. Consequently, the applicability of these methods should be investigated further for a wider range of ground conditions.

AUTHOR CONTRIBUTION STATEMENT

Qianbiao Liu: Data curation, Formal Analysis, Writing-Original draft. **Fraser Bransby & Phil Watson**: Conceptualization, Funding acquisition, Supervision, Writing-review & editing. **Michael Bertolacci & Tim French**: Supervision.

ACKNOWLEDGEMENTS

The project presented in this article is supported by the ARC ITRH for Transforming Energy Infrastructure through Digital Engineering (TIDE, http://TIDE.edu.au), which is led by The University of Western Australia (UWA), delivered with The University of Wollongong and a number of Australian and international research partners, and funded by the Australian Research Council, INPEX Operations Australia, Shell Australia, Woodside Energy, Fugro Australia Marine, Wood Group Kenny Australia, RPS Group, Bureau Veritas, and Lloyd's Register Global Technology (grant No. IH200100009). Fraser Bransby holds the Fugro Chair in Geotechnics at UWA, whose support is gratefully acknowledged. Phil Watson leads the Shell Chair in Offshore Engineering research team at UWA, which is supported by Shell Australia.

REFERENCES

Cai, Y., Bransby, F., Gaudin, C., and Uzielli, M. 2021. A framework for the design of vertically loaded piles in spatially variable soil. Computers and Geotechnics, **134**: 104140.

Ching, J., and Phoon, K.-K. 2011. A quantile-based approach for calibrating reliability-based partial factors. Structural Safety, **33**(4-5): 275-285.

ISO. 2016. Petroleum and natural gas industries – Specific requirements for offshore structures – Part 4: Geotechnical and foundation design considerations. *In* ISO 19901-4, International Standards Organization.

ISO. 2023. Oil and gas industries including lower carbon energy – Offshore structures – Part 8: Marine soil investigations. *In* ISO 19901-8, International Standards Organization.

Liu, Q., Bertolacci, M., Bransby, F., French, T., and Watson, P. 2025a. The novel integration of Bayesian hyperparameter optimization with a Bayesian neural network to inform geotechnical site characterization. Manuscript in preparation.

Liu, Q., Bertolacci, M., Bransby, F., French, T., and Watson, P. 2025b. Predicting profiles of cone resistance at untested locations using a novel large language model inspired machine learning approach. Manuscript in preparation.

Phoon, K.-K., Shuku, T., Ching, J., and Yoshida, I. 2022. Benchmark examples for data-driven site characterisation. Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards, **16**(4): 599-621.

Uzielli, M., Zei, M., and Cassidy, M.J. Probabilistic assignment of design undrained shear strength using quantile regression. *In* Proceedings, 7th International Symposium on Geotechnical Safety and Risk. 2019. Research Publishing. pp. 0-0.

Wang, Y., Akeju, O.V., and Zhao, T. 2017. Interpolation of spatially varying but sparsely measured geo-data: A comparative study. Engineering Geology, **231**: 200-217.