Proceedings of the 7th International Conference on Geotechnical and Geophysical Site Characterization
Barcelona, 18 - 21 June 2024

Can Seabed Spatial Uncertainty Be Quantified Using Advanced Statistical Approaches?

Juan Valderrama[1#], *Michael O'Neill*[1], *Fraser Bransby*[1], *Phil Watson*[1], *Michael Bertolacci*[2] and *Andrew Zammit-Mangion*[2]

[1]*The University of Western Australia, Oceans Graduate School, M053, Perth, WA, Australia*
[2] *University of Wollongong, School of Mathematics and Applied Statistics, Northfields Av., Wollongong, NSW, Australia*
[#]*Corresponding author: juan.valderramagiraldo@research.uwa.edu.au*

ABSTRACT

Knowledge of seabed properties away from investigated locations is often required, for example, when geotechnical surveys are sparse or when the field layout changes between investigation and construction phases. In such cases, design lines that appropriately incorporate the uncertainty of the seabed properties must be defined to ensure reliable (yet not overly costly) design. This paper explores how two different approaches, traditional engineering judgement and advanced statistical methods, fare at quantifying the uncertainty of a real offshore site. This is achieved by 'hiding' different data and 'scoring' the predictive performance of the methods against a range of criteria. The work reveals that complex geological sites are significantly more challenging to represent than the stationary random fields often examined in research and suggests that more advanced approaches incorporating broader data sets are required to reduce uncertainty.

Keywords: Bayesian prediction, spatial variability, quantile regression.

1. Introduction

Geotechnical site investigation, especially in offshore developments, is constrained by high costs and tight schedules, typically resulting in relatively sparse data being available for design purposes. Additionally, it is not uncommon for infrastructure to be relocated as projects progress, resulting in a degree of misalignment between the investigation and construction sites. In this context, geotechnical practitioners need to either interpolate data between investigated sites or extrapolate to unsampled locations, which is a predictive exercise.

These predictions, which aim to inform design by modelling spatial variability, can have a significant impact on foundation cost and performance. As such, a growing body of research has been published on the topic (Uzielli 2022).

Research on the topic (Wang and Zhao, 2017; Cai et al. 2019; Parra-Gomez et al. 2022) often uses data from Piezocone Penetration Tests (PCPT). These tests are repeatable, cost-effective, and relatively fast, measuring near continuous data along vertical profiles, which makes them suitable for modelling spatial variability. Current research, however, often uses relatively homogeneous data (either synthetic or from low variability sites) to investigate spatial predictive methods. In contrast, this study uses real data from a complex geological site to demonstrate the nuances involved when making real-world predictions.

This study uses GeoWarp (Bertolacci et al. 2024), a new Bayesian hierarchical model designed to predict volumes of sediment properties while taking into account the nonstationary and anisotropic nature of geotechnical data. Our analyses were conducted using a cross-validation scheme in which hidden q_{net} profiles (derived from PCPT data) were predicted using different arrays of PCPT data as input.

The performance of the predictive model was assessed in terms of accuracy (how close the prediction is to the actual value) and uncertainty (how 'sure' the model is of the predicted value). For comparison, a simpler statistical analysis using Quantile Regression (QR) is included as a point of reference.

Our study provides tentative answers to the following questions:

1. Can real and unobserved PCPT profiles be accurately predicted?
2. Is there value in implementing advanced statistical approaches?
3. What is the value of acquiring more data?

2. Dataset

The dataset used in this study comprised 13 PCPTs from a site roughly 105 m by 105 m in plan dimension (Fig. 2), located offshore north west Australia.

In the region where the tests were performed, the seabed is composed mainly of carbonate sediments (Watson et al. 2019). In the site, as shown in Fig. 1, two distinct layers compose the first 20 m of the seabed. The first has been described as turbidites (very soft carbonate sandy mud) and pelagic sediments (very soft carbonate muds). It goes from the seabed surface to a depth that varies between 8.5 m and 13 m. The second layer, which has been described as debris flow material and is composed of carbonate clasts, sands, and clays, goes down to 20 m depth. It shows highly variable conditions and overall higher cone tip resistance.

Fig. 2 shows the grid layout used to collect the test data. The depth (in m) at which the 'jump' to the second layer occurs was manually determined and is given in parentheses under each test marker, while contour lines show the approximate location (obtained through interpolation) of the transitions between layers.

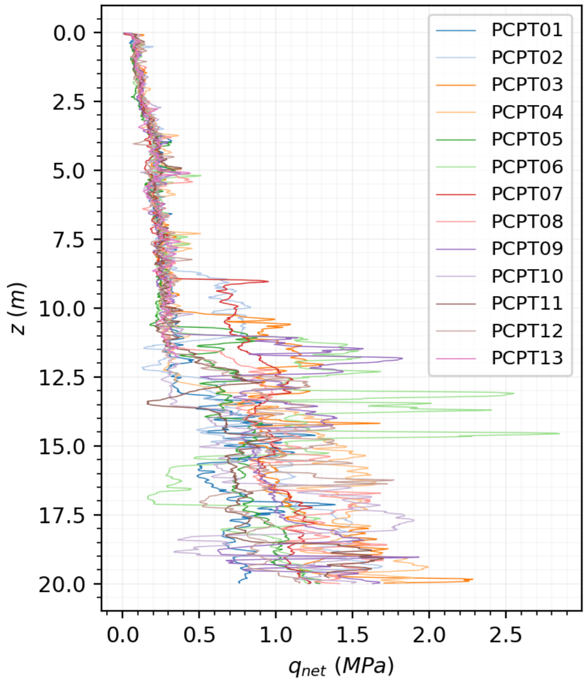

Figure 1. Profiles of net cone resistance (q_{net}) used in the analysis.

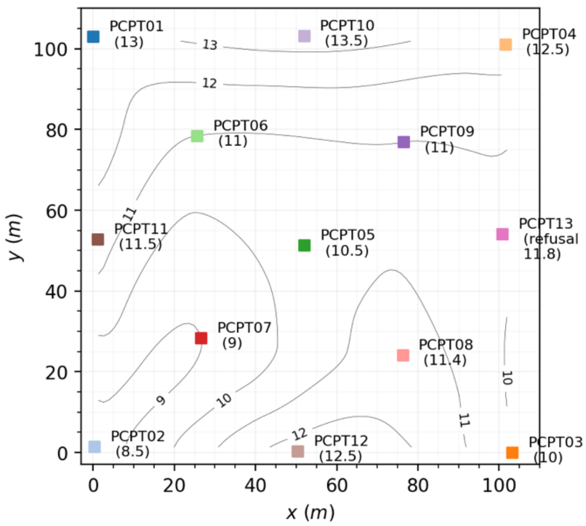

Figure 2. Plan view of the location of the PCPTs and isopachs to the second layer.

This site was selected for this study due to the challenges it poses for predictive models, which are evident in Figs. 1 and 2 and outlined below:

- **Distinct statistical properties:** The seabed profile of this site is composed of two layers that differ in their mechanical and statistical properties. The first layer, primarily pelagic, is relatively uniform and can be reasonably modelled using a linear trend and a stationary Gaussian process, where the variance and scale of fluctuation of the process do not vary with depth (Valderrama, et al. 2024) in a conventional random field framework. The second layer, a mass transport deposit, is highly nonstationary (i.e., it should be modelled using a stochastic process with spatially varying variance and scale of fluctuation).
- **Extreme values:** The second layer exhibits extreme values that manifest at specific depths. For instance, PCPT06 displays notably high resistances between 13 and 15 metres, followed by relatively low resistances between 16 and 17 metres. These variations, which are discussed later, present challenges of predictability in a cross-validation scheme.
- **Transition depth:** Given the distinct features of the units that make up the profile, the accuracy of any predictive method greatly depends on its ability to predict the transition between these two layers.
- **Data resolution:** While a relatively 'data-rich' site by the standards of many offshore projects, the grid-like layout of the data leads to the minimum distance between any given pair of tests to never be less than 33 m. This creates a gap in the information that the dataset offers: the data suggest high variability in depth, but the tests are not close enough to understand the horizontal variability.

3. Statistical models

Practical geotechnical engineering involves the translation of site investigation data into design lines. For offshore infrastructure, two design lines are typically required: a low estimate for calculating load-bearing capacity and a high estimate for determining installation requirements. This paper presents two statistically based practical approaches for defining these design lines.

The first approach, which serves as a benchmark, is QR, a relatively simple method based on descriptive statistics and engineering judgement that nevertheless can accommodate changes in the variability with depth. The second approach is GeoWarp, a method that can account for variability in a statistical model-based framework.

Both approaches can adapt to scenarios with varying amounts of data and yield measures of variability, making them usable in real-world design applications. They also produce best estimates, which can be regarded as predictions. Predictions and associated uncertainties form the basis of evaluation and comparison of the methods' performance in cross-validation schemes.

3.1. Quantile Regression (QR)

In simple terms, QR is a statistical technique that extends traditional linear regression analyses. Linear regression estimates the conditional mean of a response variable, while QR provides estimates for different quantiles of the response variable.

Following Uzielli et al. (2019) and O'Neill et al. (2022a, 2022b), Valderrama et al. (2024) demonstrate the use of QR for design line definition in geologically complex, layered profiles, using the dataset employed in this paper. In the current study, QR was applied to obtain three outputs: a low estimate (10th quantile), a best estimate (50th quantile) and a high estimate (90th quantile).

3.2. GeoWarp

GeoWarp (Bertolacci et al, 2024) is a hierarchical Bayesian modelling framework for inferring the three-dimensional geotechnical properties of subsea sediments. To achieve this, the framework decomposes the modelled property into two components:

1. **A vertical mean process ($\boldsymbol{\mu}$):**
   This component remains constant across the study region and is modelled using B-splines.
2. **A deviation process ($\boldsymbol{\delta}$):**
   This component is used to capture spatially dependent deviations of the property from the mean and is modelled as a spatial Gaussian process. This stochastic Gaussian process has expectation zero and a nonstationary covariance function.

The core mathematical expression of GeoWarp is given by the following equation, where $Y(\cdot)$ is the modelled property.

$$Y(x, y, z) = \mu(z) + \delta(x, y, z) \tag{1}$$

The most important feature of GeoWarp is the covariance function of the deviation process. For two points (x_1, y_1, z_1) and (x_2, y_2, z_2), the covariance is given by

$$\mathrm{Cov}\big(\delta(x_1, y_1, z_1), \delta(x_2, y_2, z_2)\big) = \sqrt{\sigma_\delta^2(h_1)\sigma_\delta^2(h_2)}\mathcal{M}_v(d_{12}), \tag{2}$$

where $\sigma_\delta^2(h)$ is the depth-dependent variance of the property and $\mathcal{M}_v$ is the Matérn correlation function, which is a function of the warped Euclidean distance d_{12} between the two points. The warping of 3-D space allows the model to accommodate nonstationarity in the correlation scales, both within and between layers (Zammit-Mangion et al. 2022).

While GeoWarp offers advantages over other state-of-the-art models (Bertolacci et al. 2024), its predictive ability naturally depends on the features of the observed data. In instances where insufficient information is available or spatial correlation in the data is lacking, the predictions exhibit "regression to the mean" in that the prediction defaults to the mean process.

GeoWarp is implemented as an R (2023) package that uses spatially referenced data, such as q_{net} profiles, to generate a user-defined number of simulations at user-defined locations.

In this paper the simulations produced by GeoWarp were post-processed to produce three q_{net} profiles representing the low, best, and high estimates at a specific location. Fig. 3 illustrates this, presenting a statistical analysis of 1000 simulations to derive the three profiles. The mean serves as the best estimate, while the 10th and 90th percentiles offer low and high estimates, respectively. These quantities can be used for different design purposes.

3.3. Performance evaluation

As noted above, the evaluation of predictive performance involves assessing both accuracy and uncertainty. This assessment can be done using cross-validation frameworks, where subsets of the observed data are kept hidden from the predictive model to test its performance at predicting them (Stone 1974).

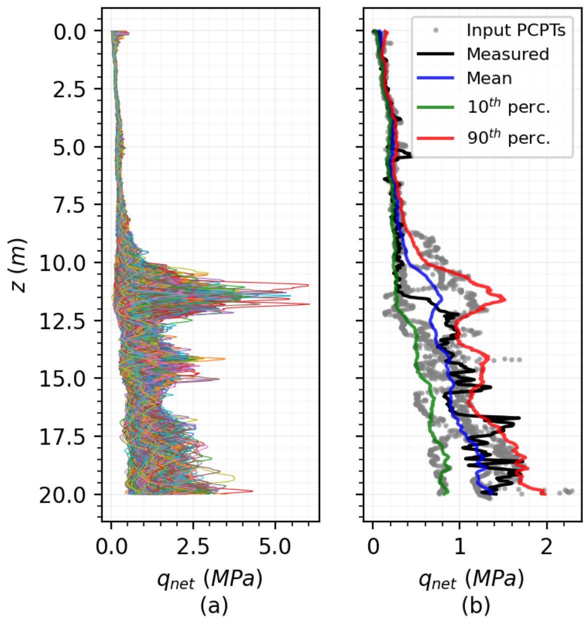

Figure 3. Typical GeoWarp output at a single horizontal location (prediction node). (a) 1000 simulations. (b) outputs for performance analyses.

3.3.1. Accuracy

Measuring accuracy entails comparing the model's best predictions with hidden (real) values. In this study, accuracy is quantified the root-mean-square error (RMSE) metric. This is calculated as

$$RMSE = \sqrt{\frac{1}{N_P}\sum_1^{N_P}\left(F_i - \widehat{F}_i\right)^2} \tag{3}$$

where N_p is the number of measured data points, $\boldsymbol{F} = (F_1, \ldots, F_{N_p})'$ is the vector of measured values, and $\widehat{\boldsymbol{F}} = (\widehat{F}_1, \ldots, \widehat{F}_{N_p})'$ is the vector of predicted values (typically chosen to be the prediction mean, labelled "best" in Fig. 3 (b)).

3.3.2. Uncertainty

For each predicted value $i = 1, \ldots, N_p$, and for a given prediction width $\alpha \in (0,1)$, both QR and GeoWarp can produce prediction intervals $[F_i^{\text{lower}}, F_i^{\text{upper}}]$ that are expected to contain the true value with probability α. For example, the 10th and 90th percentiles in Fig. 3 (b) can be used to define an $\alpha = 0.8$ (80%) prediction interval. The coverage proportion (CP_α) metric used by Lyu et al. (2023) assesses whether these intervals empirically achieve the nominal width α. This is done by counting the proportion of observed values that fall within the prediction interval. More precisely,

$$\mathrm{CP}_\alpha = \frac{1}{N_P}\sum_{\mathrm{i}=1}^{N_P} I_i, I_i = \begin{cases} 1, & F_i \in [F_i^{\text{lower}}, F_i^{\text{upper}}] \\ 0, & \text{Otherwise} \end{cases} \tag{4}$$

Whether the intervals are appropriate can be assessed by calculating $\mathrm{DCP}_\alpha = \mathrm{CP}_\alpha - \alpha$, the difference between the empirical and the nominal coverage. Methods that produce intervals that are too wide on average will have $\mathrm{DCP}_\alpha > 0$, those that are too narrow or that otherwise do not contain the true value often enough will have $\mathrm{DCP}_\alpha < 0$, and those with intervals that are an appropriate width will have $\mathrm{DCP}_\alpha \approx 0$.

4. Prediction exercises

Three modelling scenarios were investigated to answer the three questions motivating this paper:

- Whether profiles can be accurately predicted was assessed by exploring the performance metrics of GeoWarp for scenarios in which the input data were varied.
- Whether it's worthwhile using advanced statistical methods was assessed by comparing the relative predictive performance of QR and GeoWarp.
- Whether there is value in collecting more data was assessed by exploring the impact of increasing the amount of data used.

The following sections present the results of these scenarios. In all scenarios, an important concept is the representative distance between a predicted PCPT and the PCPTs used as input for its prediction. This distance was employed in various plots in the following sections.

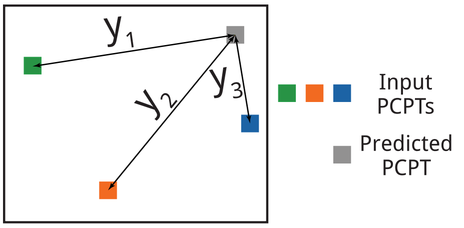

Figure 4. Representative distance scheme.

Fig. 4 illustrates a typical prediction scenario, where q_{net} profiles from three PCPTs are used to predict a q_{net} profile at the location of a test concealed from the predictive model. The distance from each input point is denoted as $y_i, i = 1,2,3$, and the 'representative distance' $\bar{y}$ is defined as the average of these three values.

4.1. Question 1: Can real profiles be accurately predicted?

To assess prediction accuracy, we adopted a "pick-three-predict-one" cross-validation approach, in which we predict one profile using three profiles as input. The targets of the prediction were the inner PCPTs (05 to 09 in Fig 2), which were predicted using every other combination of three PCPTs. This resulted in 1100 separate predictions.

The RMSE results (Fig. 5) indicate that for the 'pick-three-predict-one' scenarios, representative distance does not appear to correlate with prediction accuracy. Counterintuitively, predictions are not consistently more accurate when the input PCPTs are closer to the predicted PCPT. This may be because GeoWarp defaults to predicting the site-wide mean when a dataset lacks sufficient data quantity or density to characterise spatial variability. In such cases, the model is expected to achieve better accuracy metrics for tests that are similar to the mean profile – and this expectation aligns with the observed results: Fig. 6 provides a summary of the RMSE results in the form of boxplots, while Fig. 7 displays a comparison of the five predicted tests with a site-wide mean. As is clear, accuracy is notably higher for PCPTs 05, 07, and 08 – each of which is closer to the site-wide mean than PCPTs 06 and 09.

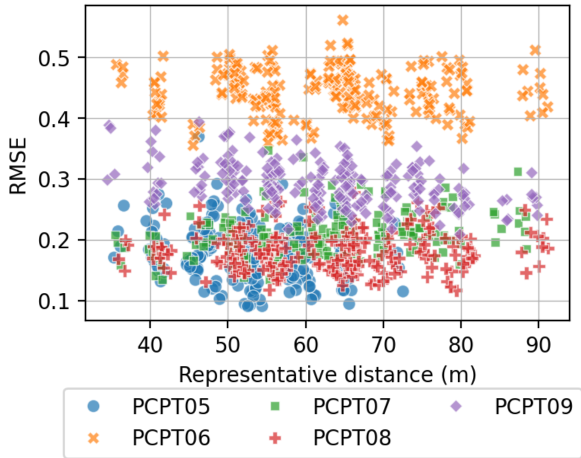

Figure 5. RMSE results.

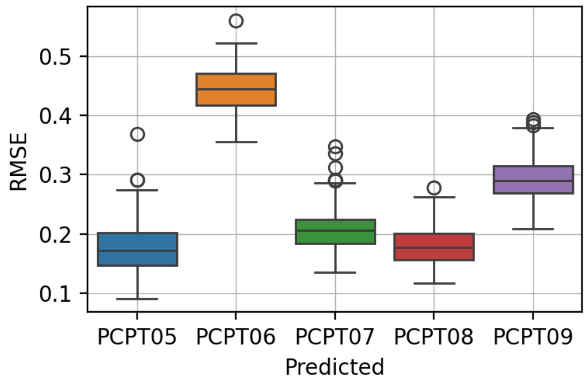

Figure 6. RMSE boxplots.

The low accuracy (high RMSE) of the predictions for PCPT06 may be attributed to the test's unique nature (Fig. 1): with features unseen elsewhere in the site, it is effectively 'unpredictable'. Fig. 8 (a) illustrates this – it is evident that if the only inputs the model receives are PCPTs 01, 04 and 05, accurately predicting the variations in PCPT06 between 11 and 18 m depth is nearly impossible.

The outlier PCPT06 is an extreme case, however. Consider Fig. 8 (b), where PCPT05 is predicted. While the profile is predicted reasonably well, as reflected in the relatively low RMSE, there is noticeable smoothing, particularly evident in the transition between units. In this case, the smoothing may be attributed to limited data input and uncertainty about the location of the layer change. Given the potential design implications of a sharp change in cone resistance, such as for installation, this example raises a new question: what defines a good prediction?

If the goal is to predict natural profiles and accurately represent geological features at times unpredictable (such as layer transitions), then it is crucial to recognise that relying on a single source of information (e.g., PCPT) challenges predictive models with a near-impossible task. This limitation could be mitigated by considering additional sources of information – for instance, geophysical surveying is typically conducted at various stages of project development, and such data could inform the prior distributions that form the foundation of GeoWarp. Other data sources, like borehole logs, should also be duly considered – and this is an area of future work.

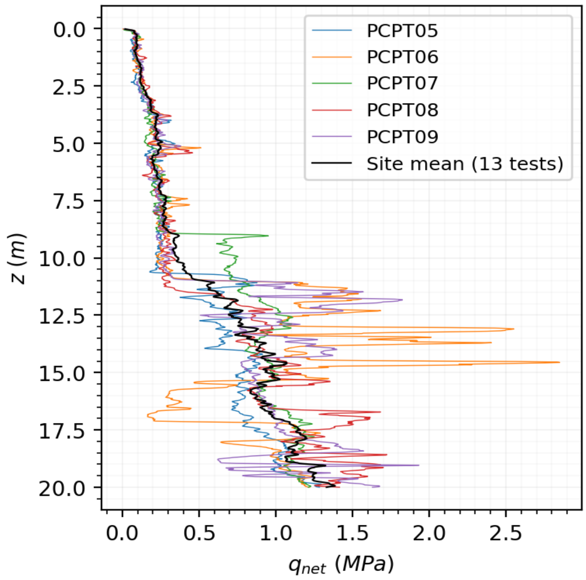

Figure 7. Comparison of the 5 predicted PCPTs and a site-wide mean profile

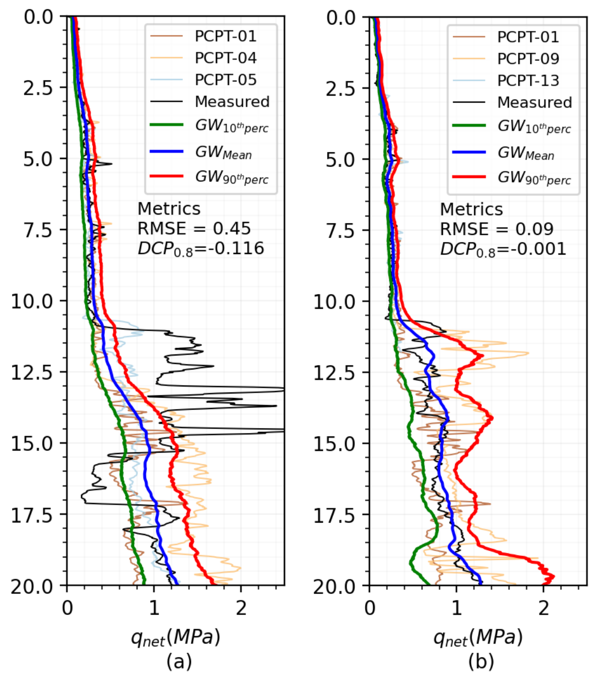

Figure 8. Prediction examples. (a) prediction of PCPT06 (b) prediction of PCPT05.

Aside from the potential modelling improvements, it is also worth assessing the appropriateness in geotechnical applications of metrics like RMSE. RMSE provides an idea of the overall quality of a prediction but may fail to capture aspects of high impact for geotechnical practice. For example, it assigns equal penalty to both under- and over-estimation of the quantity, which may not be appropriate. Consequently, the choice of metrics tailored to geotechnical design represents a promising avenue for future research.

Another key observation from the analysis relates to the importance of capturing site variability in the input data. To explore this, changes in prediction accuracy when PCPT06 is included as one of the three inputs were examined. Fig. 9 shows boxplots representing the differences – for PCPTs 07, 08, and 09, the accuracy slightly improves, and while the finding may not be generalisable, it draws attention to the fact that if outliers are not captured as part of an investigation, then the overall knowledge of the site is poorer, and the predictive ability of models (such as GeoWarp) potentially decreases. This is also an area of further investigation.

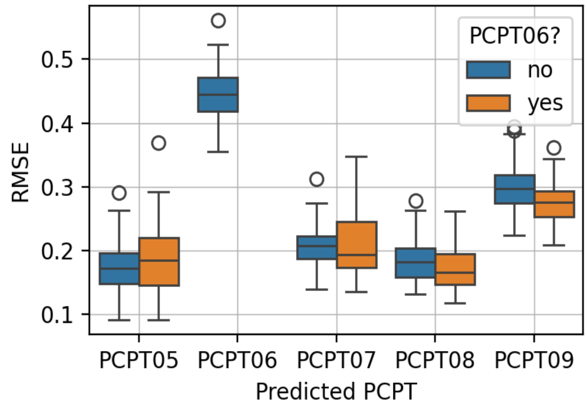

Figure 9. RMSE boxplots for comparing the effect of including PCPT06.

Up to this point, we have focused only on prediction accuracy. Fig. 10 shows the relationship between $DCP_{0.8}$ and RMSE for all 'pick-three-predict-one' analyses. The results tend toward negative values that, in some cases, reach values as low as -0.5. This implies that the predictions contain the true value too infrequently, which can arise either because the intervals are too narrow, because they are poorly placed, or both. The worst coverage is, not surprisingly, achieved on the outlier PCPT06. In general, however, the coverage is too low at most locations. This may be an artefact of using only three PCPT profiles as input (i.e., low variability in the inputs yields narrow prediction intervals). A clearer understanding would require the inclusion of cases in which additional PCPTs are included.

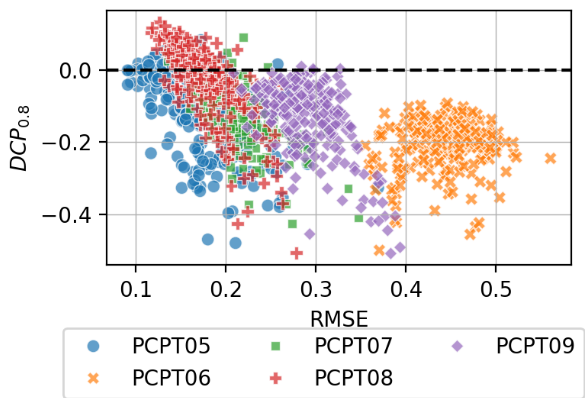

Figure 10. Relation between $DCP_{0.8}$ and RMSE.

4.2. Question 2: Is there value in implementing advanced statistical approaches?

To address this question, we examined 21 representative 'pick-one-predict-three' cases of the 220 cases where PCPT08 had been predicted. The choice of PCPT08 was motivated by its strategic location within the inner investigated 'box' (Fig. 2), allowing for both interpolation and extrapolation scenarios. Furthermore, its off-centre location in the site allows for a wider range of representative distances than PCPT05. Finally, while still presenting a challenging scenario, PCPT08 exhibits a lesser degree of unpredictability in comparison to PCPTs 06, 07 and 09.

The predictions were benchmarked against the performance of QR, which, as outlined by Valderrama Giraldo et al. (2024) involves the definition of layering systems. We adopted a pragmatic layering approach that is thought to mirror the methodology typically employed by engineers in a design setting. While introducing subjectivity, this aligns with the overall objective: to evaluate whether advanced frameworks like GeoWarp exhibit superior performance compared to other (simpler) tools.

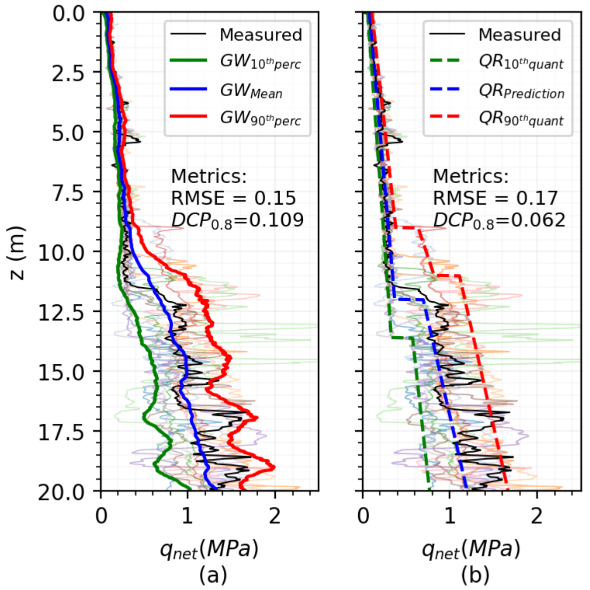

Figure 11. Prediction examples with tests 02, 04, and 06 used as inputs. (a) GeoWarp. (b) QR.

Fig. 11 illustrates the outputs for both methods using PCPTs 02, 04, and 06 as inputs to predict PCPT08. This case highlights GeoWarp's capability to represent depth-dependent variance (heteroscedasticity) in the data, evident when comparing the lower and upper bound from both methods: while both approaches predict the layer transition 'jump' at around 11 m depth, QR provides an overly simplified representation of the variability below that depth. Despite the impact of data-quantity restrictions inherent in the 'pick-three-predict-one' approach (which possibly causes GeoWarp's smooth best estimate as well), this example suggests that GeoWarp provides a more comprehensive representation of site variability.

For the remainder of the cases, the comparison is based on their performance metrics. Fig. 12 compares the two methods in terms of accuracy, with the vertical axis representing RMSE from GeoWarp and the horizontal axis representing RMSE from QR. A 1:1 line in the plot serves as a reference: points below the line suggest GeoWarp performs better, while points above indicate QR performs better. Albeit the trend is not particularly strong, GeoWarp appears to outperform QR accuracy-wise. More cases (including those with more input PCPTs) would be needed to form a clearer picture.

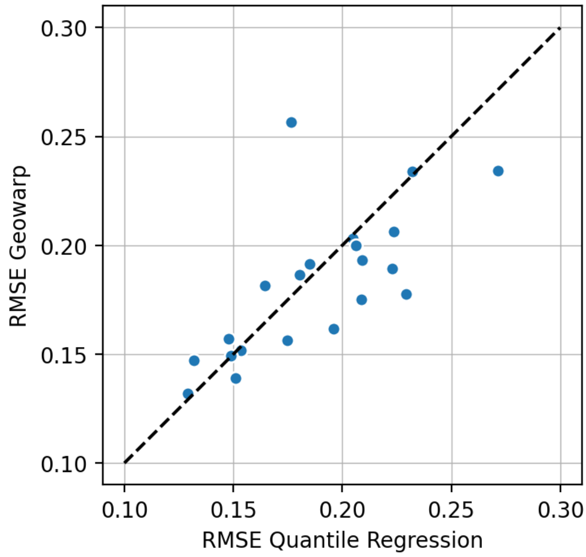

Figure 12. RMSE comparison.

Fig. 13 compares $DCP_{0.8}$, incorporating histograms to display the data distribution, which is helpful when interpreting data (given that the ideal DCP_{80} is zero). While the QR outputs show a relatively even spread between -0.1 and 0.1, the results from GeoWarp are more heavily concentrated around values between 0 and 0.1 – which suggests better (appropriately conservative) performance from GeoWarp (as negative values imply overconfident predictions). However, extreme values at -0.2 and -0.3 in the GeoWarp results indicate instances of overconfident (and inaccurate) prediction that warrant more research.

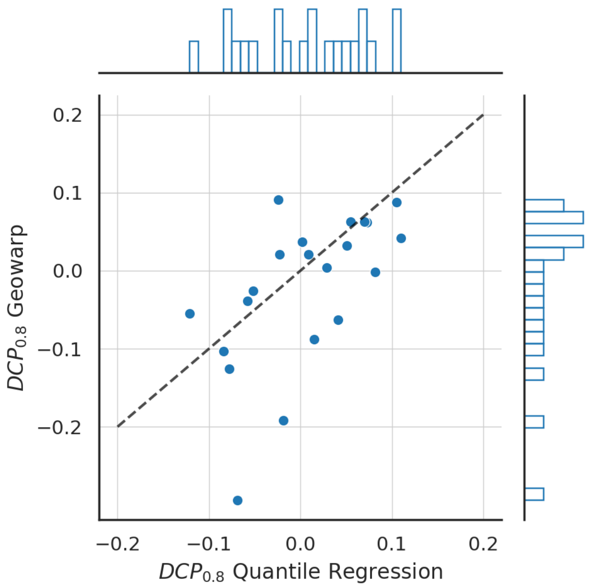

Figure 13. $DCP_{0.8}$ comparison.

While the limitations of the 'pick-three-predict-one' approach (which are to be addressed through further research) influenced the results, this section suggests that while metric-wise, straightforward statistical methods may yield comparable (but slightly inferior) outputs, implementing advanced models like GeoWarp provides a more thorough representation of a site's variability.

Aside from considerations of prediction accuracy and uncertainty calibration, GeoWarp also has the capability to produce predictive simulations, not just quantiles or best-guess predictions. These simulations, which can have more realistic statistical properties than prediction means, can be used as stand-ins for data at missing in follow-up design analyses.

4.3. Question 3: What is the value of acquiring more data?

Moving beyond the 'pick-three-predict-one' framework, we now consider the value of acquiring more data. To do this, we conducted an incremental inclusion of data in a predictive exercise focused on PCPT08. We systematically expanded the input data, commencing with three initial PCPTs (01, 10, and 11), and progressively incorporated closer tests until PCPT08 was entirely encompassed by the input PCPTs. The progression unfolded as follows: the first additional test to be introduced was PCPT06, followed by 02, 04, 05, 07, 09, 12, 13 and 03, giving a total of 10 cases.

For comparison purposes, QR analyses were also conducted for each case. The results for both approaches are shown in Fig. 14. The upper plot indicates a trend toward improved accuracy with larger input datasets, as expected. Interestingly, the incremental improvement seems to stall when 6 PCPTs or more are used.

The results for $DCP_{0.80}$ are displayed in the lower plot in Fig. 14. These show only positive values for cases with 4 or more input PCPTs, which implies wide prediction intervals that slightly overestimate the site's variability. This is in contrast to Fig. 10 for the 'pick-three-predict-one' cases where $DCP_{0.80}$ is consistently negative. Also, the behaviour exhibited by the DCP values closely mirrors that of RMSE, meaning, firstly, that the values also tend to stabilise after the inclusion of the sixth test, and secondly, that while the accuracy (RMSE) of GeoWarp is better than QR for all cases, the opposite holds for $DCP_{0.80}$, meaning that QR consistently produces DCP values closer to zero.

While these results may be influenced by the sequence in which the data was added (making the findings somewhat anecdotal) a comparison between the first and last cases suggests that given the large variability of this site, increasing data allows GeoWarp to improve its characterisation of the variability, which in turn leads to wider prediction intervals (i.e. an acknowledgment of higher uncertainty), and hence larger DCP values. However, the observed trend appears to plateau after a certain quantity of data is utilized. This intriguing finding, which may be site-specific and requires further research, suggests a potential diminishing return for additional data.

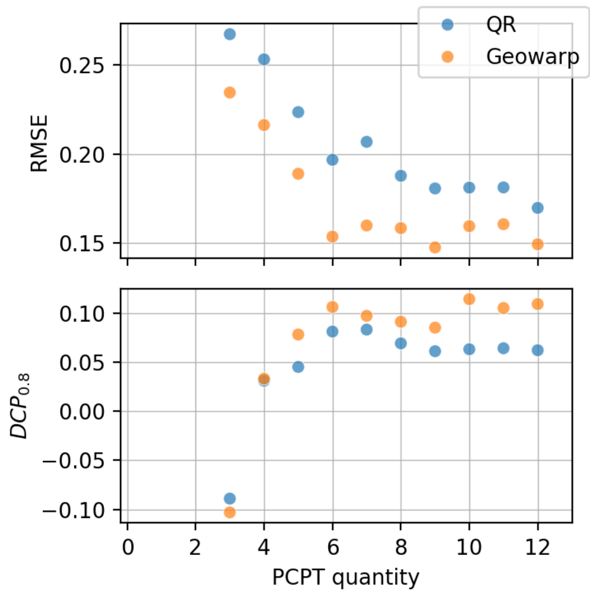

Figure 14. Relation between data quantity and predictive performance metrics.

Further comparison between QR and GeoWarp is shown in Fig. 15 by exploring actual prediction intervals for the case with all data included. The prediction interval from GeoWarp successfully describes the site's variability (and depth-dependent variance), without the spread observed at around 17 m in cases with less data (Fig. 11). However, GeoWarp still produces a smooth prediction (and prediction interval) – while it accurately represents the overall behaviour of PCPT08, it misses the sharp variations that may be important, for instance, for installation assessment. In contrast, the subjective outputs from QR arguably address this feature better, by showcasing varying depth to Unit 2 and sharp layer transitions.

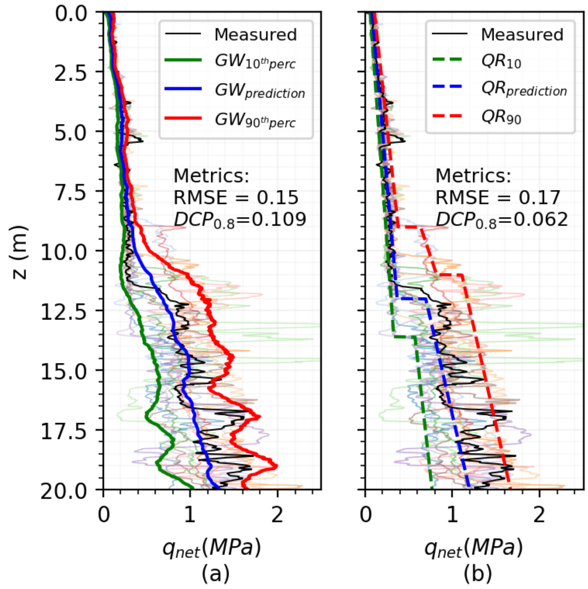

Figure 15. Prediction using all available tests but PCPT08. (a) GeoWarp. (b) QR.

It should be noted, however, that while QR provides a single output offering a 'big picture' idea of a site's features, GeoWarp is a framework for predicting 3D volumes of data. Although not explored in this work, in GeoWarp, the inclusion of additional data not only enhances predictive performance but also enables location-specific analyses of uncertainty, which, for instance, could inform planning and optimisation of a site investigation such that each PCPT location is selected to progressively reduce uncertainty (to a certain plateau). Ongoing research is dedicated to exploring these possibilities.

5. Conclusions

This study showcased the predictive performance of GeoWarp, a novel Bayesian framework, in a geologically complex site, comparing it with the simpler QR method. Metrics-wise GeoWarp demonstrated a slight advantage over QR in its predictive ability. There are, however, some aspects in which GeoWarp offers substantial advantages, the main one being that GeoWarp analyses the sediments' spatial structure, which allows an accurate representation of the depth-varying variance of geotechnical properties.

The findings also underscored that statistical approaches can output predictions in which the underlying uncertainty is quantified by prediction intervals. Yet, it was shown that predicting unsampled locations within natural deposits (particularly those influenced by mass transport processes) becomes impractical when relying solely on one sparse data source. Future research endeavours should focus on methodologies enabling the fusion of datasets, such as combining PCPT with geophysical surveying and borehole logs. In the Bayesian modelling context in which frameworks such as GeoWarp are developed, leveraging additional datasets to inform prior distributions holds the potential to enhance predictive accuracy.

Acknowledgements

This research 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

Bertolacci, M., A. Zammit-Mangion, J. Valderrama Giraldo, M. O'Neill, F. Bransby, and P. Watson. 2024. "GeoWarp: Warped spatial processes for inferring subsea sediment properties." *Journal of the American Statistical Association.* In revision.

Cai, Y., J. Li, X. Li, D. Li, and L. Zhang. 2019. "Estimating soil resistance at unsampled locations based on limited CPT data." *Bulletin of Engineering Geology and the Environment* 1435-9537. https://doi.org/10.1007/s10064-018-1318-2.

Lyu, B., Y. Hu, and Y. Wang. 2023. "Data-driven development of three-dimensional subsurface models from sparse measurements using Bayesian compressive sampling: A benchmarking study." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 9 (2): 04023010. https://doi.org/10.1061/AJRUA6.RUENG-935.

O'Neill, M., F. Bransby, and P. Watson. 2022b. "The effectiveness of Spatial Interpolation of Sparse PCPT data to optimise Offshore Design." *8th International Symposium for Geotechnical Safety & Risk (ISGSR 2022).* Newcastle, NSW, Australia.

O'Neill, M., F. Bransby, J. Doherty, and P. Watson. 2022a. "Spatial interpolation of sparse PCPT data to optimise infrastructure design." *Cone Penetration Testing 2022.* 1023-1028. 1023-1028. https://doi.org/10.1201/9781003308829-154.

Parra-Gomez, L. J., Colmenares J. E., and Bohorquez M. P. 2022. "A functional-geostatistics approach to deal with soil variability in project management and deep foundation design." *20th International Conference on Soil Mechanics and Geotechnical Engineering.* Sydney: ISSMGE. 4601-4606. https://www.issmge.org/uploads/publications/1/120/ICSMGE_2022-786.pdf.

R Core Team. 2023. "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Stone, M. 1974. "Cross-Validatory Choice and Assessment of Statistical Predictions." *Journal of the Royal Statistical Society. Series B (Methodological) 36, no. 2* 111–47. http://www.jstor.org/stable/2984809.

Uzielli, M. 2022. "Non-deterministic interpretation and applications of CPT testing data." Edited by G Gottardi and L Tonni. *Cone Penetration Testing 2022.* Bologna, Italy: CRC Press. 81-93. https://www.taylorfrancis.com/chapters/oa-edit/10.1201/9781003308829-6/non-deterministic-interpretation-applications-cpt-testing-data-uzielli.

Uzielli, M., M. Zei, and Mark J. Cassidy. 2019. "Probabilistic assignment of design undrained shear strength using quantile regression." Edited by Jianye Ching, Dian-Qing Li and Jie Zhang. *Proceedings of the 7th International Symposium on Geotechnical Safety and Risk (ISGSR).* Taipei: Research Publishing. 188-193. ISBN: 978-981-11-2725-0; doi:10.3850/978-981-11-2725-0 IS15-8-cd.

Valderrama, J., M. O'Neill, F. Bransby, and P. Watson. 2024. "A case study on quantile and percentile regression to select design lines for complex, real work profiles." *Proceedings of the 7th International Conference on Geotechnical and Geophysical Site Characterization.* Barcelona.

Wang, Y., and T. Zhao. 2017. "Statistical interpretation of soil property profiles from sparse data using Bayesian compressive sampling." *Géotechnique* 67 (6): 523-536. https://doi.org/10.1680/jgeot.16.P.143.

Watson, P. G., M. F. Bransby, Z. L. Delimi, C. T. Erbrich, I. Finnie, H. Krisdani, C. Meecham, et al. 2019. "Foundation design in offshore carbonate sediments-building on knowledge to address future challenges." Edited by N.P. et al. López-Acosta. *XVI Pan-American Conference on Soil Mechanics and Geotechnical Engineering (XVI PCSMGE).* Cancun, Mexico: ISSMGE. 240-274.

Zammit-Mangion, A., T. L. J. Ng, Q. Vu, and M Filippone. 2022. "Deep Compositional Spatial Models." *Journal of the American Statistical Association* (Taylor & Francis) 117 (540): 1787-1808. doi.org/10.1080/01621459.2021.1887741.