Applications of data science in offshore geotechnical engineering: State of practice and future perspectives

B.Stuyts[1,2], S.Suryasentana[3]

[1]*UGent Geotechnical Laboratory, OWI-lab, Technologiepark 68 Zwijnaarde, 9052, Belgium*

[2]*Vrije Universiteit Brussel, OWI-lab, Pleinlaan 2, 1050, Brussels, Belgium*

[3]*Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow G11XJ, U.K.*

ABSTRACT: Data-driven predictive models are becoming ubiquitous in society with a wide range of applications, including engineering design. For offshore geotechnical engineering, semi-empirical models make up most of the predictive models to date. These models implicitly include knowledge on soil mechanics and foundation behaviour gathered from laboratory testing, scale model testing and field observations. Modern data science techniques enable researchers and practicing engineers to leverage state-of-the-art artificial intelligence tools to re-evaluate the creation of these predictive models. This keynote explores basic and advanced applications of artificial intelligence in offshore geotechnical engineering and aims to offer a perspective on the future use of these techniques in solving complex geotechnical problems.

1 INTRODUCTION

In recent years, digitisation has caused disruption across industries. The integration of cloud-based data platforms, sensors with direct connection to the internet (IoT), autonomous predictive models and robotics have paved the way for a so-called *fourth industrial revolution* (McKinsey & Company, 2022).

While the acceleration of these technologies is most apparent in computer science and associated domains, digital innovation is also permeating legacy industries such as the (offshore) construction sector. The deployment of digital solutions for survey data collection, foundation engineering calculations and in-place monitoring can lead to increased efficiency and improved insights in the geotechnical parameters affecting foundation behaviour.

Figure 1 illustrates the use of digital data in foundation engineering calculations. The basic workflow does not differ from a conventional (non-digital) workflow. Geotechnical data is gathered in the field and is stored in a centralised database after rigorous quality control. This data, combined with a thorough understanding of soil mechanics and possibly monitoring data, is used to determine the appropriate foundation type and dimensions necessary to ensure satisfactory performance over the operational lifetime. Although this approach is standard practice in offshore geotechnical engineering, digital data transfer and automated decision making can have a significant impact oneach stage of the workflow.

Efforts have been made to facilitate the digitisation of offshore ground investigation data with the development of the AGS4.0 and DIGGS standards (AGS, 2017; Bachus, 2017) and cloud-based geotechnical data platforms have been developed to securely share geotechnical data with various project stakeholders (Doherty et al., 2018; Stuyts et al., 2022). These data platforms ensure that geotechnical data is immediately available for engineering calculations and the geotechnical engineer does not need to spend time and effort to digitise geotechnical test results.

The foundation engineering calculations which are based on the site-specific geotechnical measurements are most often calibrated semi-empirical models based on (scale) model testing, numerical modelling and in-situ monitoring data. Machine learning methods are increasingly accessible through implementations in open-source packages (Scikit-learn developers, 2015) and allow complex mathematical models to be fitted to the available data.

This is a peer-reviewed, accepted author manuscript of the following paper: Stuyts, B., & Suryasentana, S. (2023). Applications of data science in offshore geotechnical engineering: state of practice and future perspectives. Session on 'AI applications' in 9th OSIG conference, london, United Kingdom.

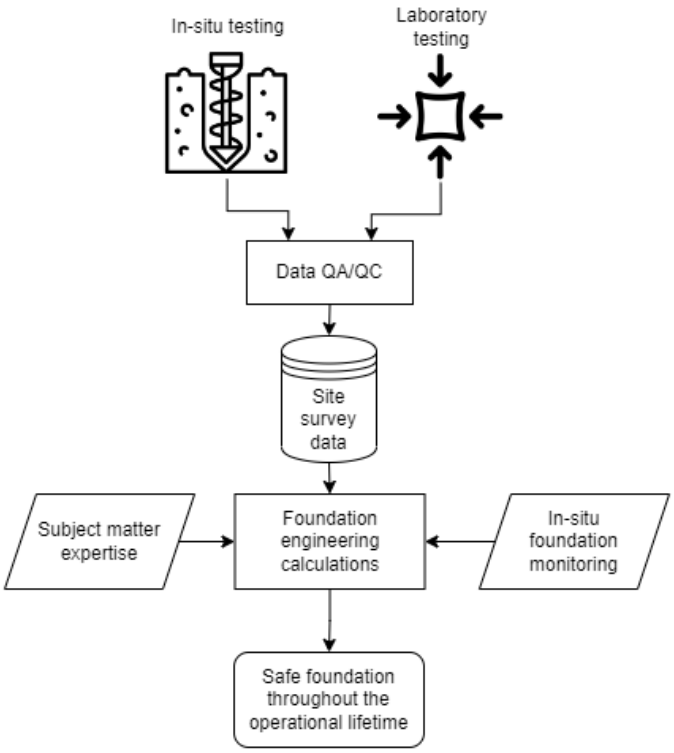


Figure 1. Flowchart illustrating the use of digital data in foundation engineering calculations.

This paper outlines the different types of machine learning models which can be applied to offshore geotechnical problems and aims to provide insight into the potential of these data-driven techniques. Importantly, evaluation criteria for machine learning model quality are also discussed. Applying these techniques without due consideration of their drawbacks should be avoided.

The different types of machine learning are illustrated in Section 2. The data used in a machine learning workflow needs to be evaluated carefully as it will affect the models which are built from it. Section 3 describes data requirements and provides basic guidance for evaluating machine learning model quality.

While some machine learning techniques (e.g. linear regression) are already familiar to most geotechnical engineers, a number of advanced model types have been developed in recent years, such as models which provide uncertainty estimates, recurrent neural networks for time series predictions, physics-informed neural networks, multi-fidelity data fusion techniques and large language models. These advanced models are discussed in Section 4 and they are applied to example offshore geotechnical applications in Sections 5 to 8.

2 TYPES OF MACHINE LEARNING

Machine learning is a subset of artificial intelligence in which a computer (*machine)* extracts information from a given dataset without the intervention of a human being. In many cases, the computer will mimic the way in which humans learn, gradually improving the accuracy of predictions as more data becomes available. The accuracy of machine learning models depends on the dataset size, the type of mathematical model being trained and the quality of the data presented to the algorithm.

Machine learning can be subdivided into four categories which are discussed hereunder.

2.1 *Supervised learning*

In a supervised learning problem, a labelled dataset is provided to the machine learning algorithm during the training phase. The label refers to the *target* outcome that the predictive model seeks to predict. This outcome can either be discrete or continuous.The input variables on which the predictions are based are called *features*.

Figure 2 shows an illustration of training a machine learning model on a dataset with m labelled samples. The dataset has n features and a target variable y which is known for the labelled samples. The objective of the machine learning is to learn the relationship between the features and the target, which would lead to a good prediction of the target on new, unseen features.

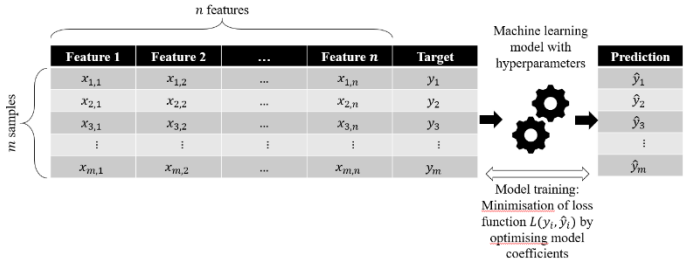


Figure 2. Illustration of supervised learning.

The machine learning model is a mathematical model which formulates the target as a function $\hat{f}$ of the features (Equation 1). This estimate may differ from the actual value y of the target. This error ε is minimised during the training process by optimising the model coefficients.

$$\hat{y} = \hat{f}(x_1, x_2, \ldots, x_n) = y + \varepsilon \quad (1)$$

A *loss function* is minimised to find the optimal set of model coefficients. The optimal fit depends on the scatter in the training data and the ability of the model to capture the underlying trends in the data.

The model also has a set of *hyperparameters*; parameters which control the behaviour of the model, but are fixed before the training phase. The hyperparameters can be *tuned* to improve the machine learning model performance.

Once trained, the model is able to make predictions for any set of feature values. The accuracy of the predictions can be evaluated against the known values of the target variable for the labelled dataset. When additional data is collected, the training can be repeated to further optimise the model coefficients and to tune the hyperparameters. As such, building a machine

learning is an iterative process, with models evolving over time.

2.1.1 *Supervised learning for continuous target variables: Regression*

In a regression problem, the value of a continuous variable is predicted based on a number of features. This type of modelling is routinely performed for calibrating semi-empirical methods in geotechnical engineering. Example applications of regression models include:

- Geotechnical parameter correlations: The dataset consists of samples where the value of the target is measured together with the values of the features. The shear wave velocity (target) has been expressed as a function of CPT measurements and in-situ stress conditions (features) (e.g. Rix & Stokoe, 1991; Hegazy & Mayne, 2006; Robertson, 2009).
- Prediction of foundation installation behaviour: When the installation behaviour of a foundation can be measured (e.g. blow count during pile driving), a model can be trained on the available installation data. Stuyts (2020) reported results from a community prediction exercise aimed at the development of a blow count model for driven tubular jacket piles for a site in the Southern North Sea. The results highlighted the importance of capturing geotechnical knowledge in the training process and of model interpretability. Figure 3 shows an example of blow count predictions with a linear regression model.

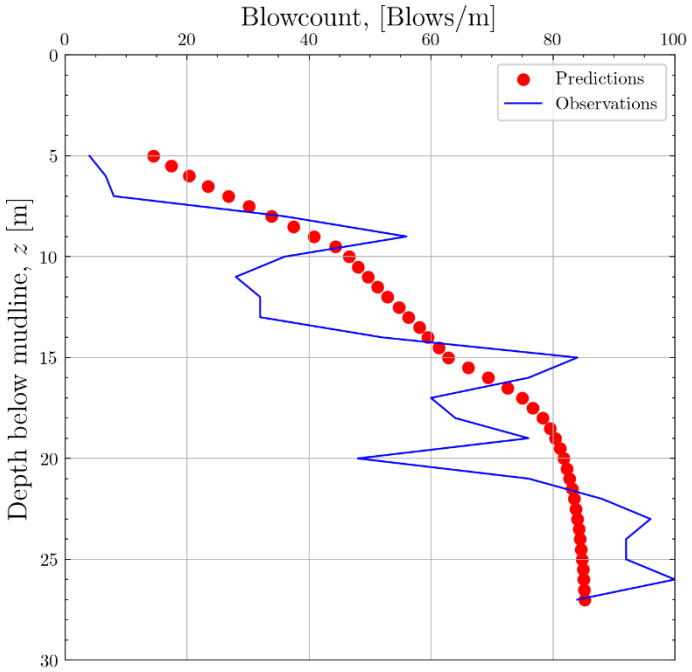


Figure 3. Example of predicted blow count from a trained regression model against blow count measurements (Stuyts, 2020)

While linear regression is the most intuitive model to understand the principles of machine learning, several other model types exist. Advanced models such as extreme gradient boosting (Chen et al., 2015) or artificial neural networks (Jain et al., 1996) do not have a straightforward mathematical formulation and can lead to *black box* behaviour. In such cases, verification of the model behaviour for variations of the features is required to ensure that the model provides physically meaningful predictions.

2.1.2 *Supervised learning for discrete target variables: Classification*

Machine learning models can be trained to predict categorical targets. The model predicts the *class* of the output from several possible classes. Example problems in geotechnical engineering include:

- Prediction of soil type class from CPT measurements. The classification chart by Robertson (2009) based on normalised cone tip resistance Q_t and normalised sleeve friction F_r is routinely used to determine whether a layer consists of clean sand, silty sand, clay, … The Robertson chart is a graphical representation of a classification model, based on training data where both soil type descriptions and CPT measurements were available. Stuyts (2020) presented a decision tree model to evaluate soil type from CPT. The importance of the tree depth (a hyperparameter of the decision tree model) was highlighted. Trees with greater depth led to inconsistent predictions.
- Installation refusal: When the presence of a hard layer or insufficient capacity of the installation equipment halts the installation of an offshore foundation (e.g. driven pile or suction caisson), the conditions leading to this refusal can be recorded. Accumulating a dataset of installations with and without refusal allows a binary classification model to be fitted. The combination of ground conditions, foundation dimensions and installation equipment can be used to predict whether the installation will be successful or not. Offshore wind farm foundation installation, with many structures being installed in the same geological settings, provides possibilities for deriving such models.

Classification models have traditionally not been used as much as regression models in offshore geotechnical engineering, but when the target is appropriately formulated, these models can add valuable insights for decision making in offshore foundation design and installation.

Several classification methods such as design trees and boosting methods, assign a probability to each

class, which makes them suitable for probabilistic predictions.

2.2 *Semi-supervised learning*

In semi-supervised learning, the model is trained on a dataset that contains both labelled and unlabelled data. The labelled data is used to train an initial version of the model which is then used to make predictions on the unlabelled data. Unlabelled samples where the probability of the prediction is high are then added to the labelled data and the training is repeated. This type of modelling can be useful when the data volume is large and the labelling requires significant effort.

The principle of semi-supervised learning can be illustrated on downhole CPT data. When starting a new CPT stroke from the bottom of the borehole, a certain amount of displacement is required before reaching the ultimate resistance of the soil. This gives rise to a near-horizontal line at the start of each push which needs to be filtered out to allow meaningful parameter selection. The displacement needed to reach the ultimate resistance depends on the soil type, the stress level at the depth considered and the cone resistance in the layer.

A binary classification problem can thus be formulated which determines whether the CPT measurement at a certain depth corresponds to the initial part of the push or not.

A dataset consisting of 80 downhole CPTs was assembled and the initial part of the push was identified on 15 out of the 80 CPTs. Figure 4 shows an example downhole CPT from the labelled dataset with the initial part of the push marked in red.

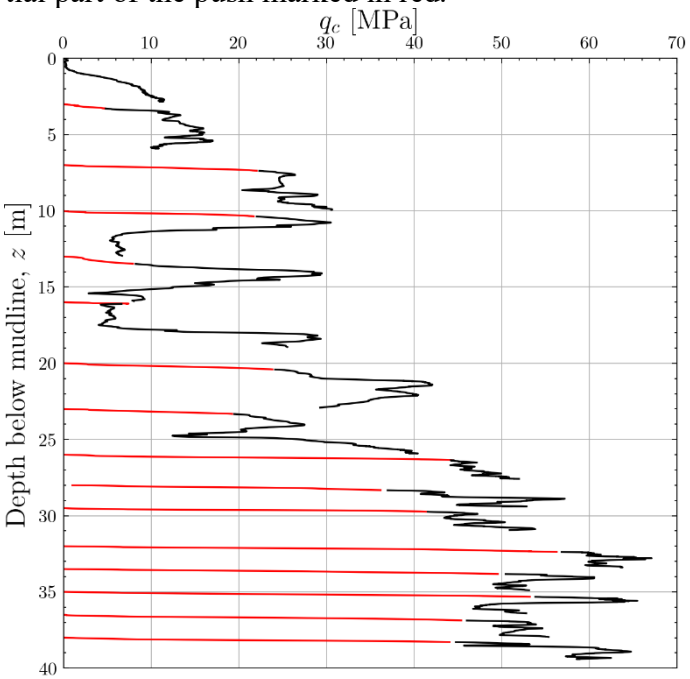


Figure 4. Cone tip resistance for a downhole CPT in the labelled dataset with initial part of the push marked in red.

The model features were selected as follows:

- Distance from top of push: For each push, the CPT penetration relative to the bottom of the hole was recorded;
- Soil behaviour type index I_c: Robertson (2009) shows that I_c can be used to differentiate between cohesionless and cohesive behaviour. Figure 5, in which the initial part of the push is color-coded in blue, shows that the ultimate resistance takes more displacement to mobilise as I_c becomes smaller (more cohesionless soil);
- Cone tip resistance q_c: In soils with high cone resistance, more displacement will be required to reach the ultimate resistance. The cone tip resistance allows differentiation between sands of different relative density.

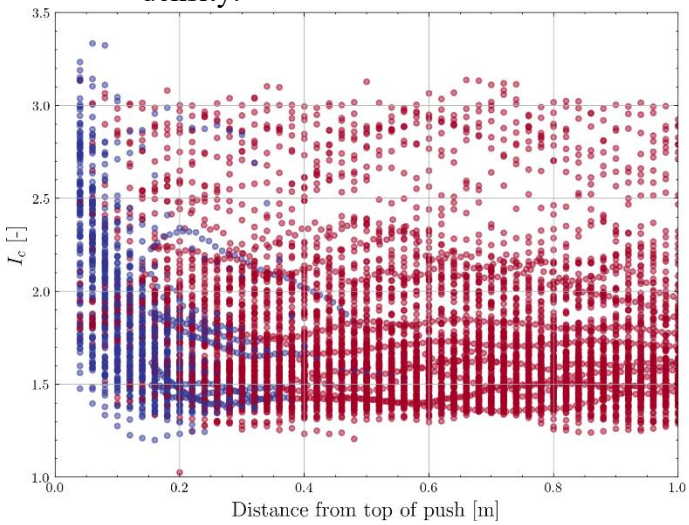


Figure 5. CPT measurements belonging to the initial part of the push color-coded in blue as a function of soil behaviour type index and distance from the top of the push

A decision tree classifier with a maximum tree depth of 3 was chosen for the semi-supervised learning. The decision tree model was first trained on the 15 CPTs. Subsequently, the trained model was applied to the unlabelled CPT and the prediction for which the class was predicted with greater than 75% probability were added to the training dataset. The decision tree was then retrained with this additional data.

The model was evaluated by calculating the predictions for I_c=2 and q_c=10MPa (more cohesionless behaviour) and I_c=3 and q_c=0.5MPa (more cohesive behaviour) for the supervised learning on the 15 labelled CPTs and the semi-supervised learning on the entire dataset. Figure 6 shows the probability of belonging to the initial part of the CPT push as a function of the distance from the top of the push. Points located less than 0.2m from to the top of the push have a high probability of belonging to the part that needs to be filtered out. The results show that adding the non-labelled points (see dashed lines) leads to changes in the depths where the transition to the ultimate resistance is expected. For clay, a sharper transition is noticed, whereas for sand the probability of

belonging to the initial part of the push increases for points less than 15cm away from the top of the push. Without explicitly having to label the remaining 65 CPTs, the estimates of the transition depths are refined.

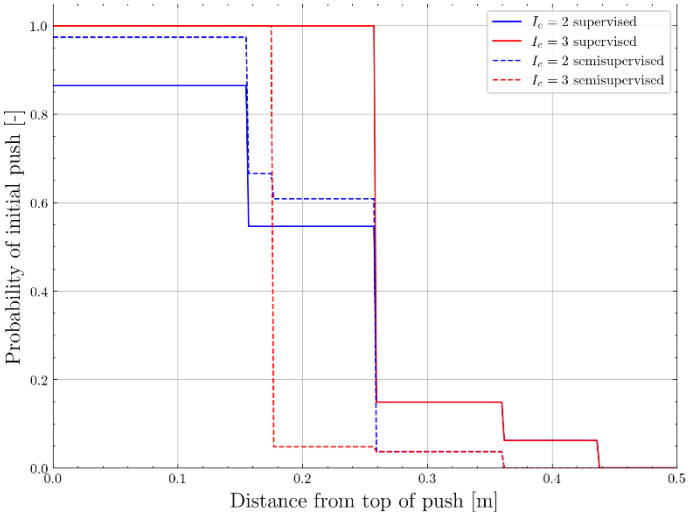


Figure 6. Comparison of performance of supervised and semi-supervised learning models for predicting the initial part of the CPT push.

The trained semi-supervised model can be applied to an unseen location as shown in Figure 7. The model is capable of distinguishing between various soil types and can also differentiate between lower and higher values of cone resistance. The black line shows the cone resistance trace with the initial part of each push filtered out. This trace appears to be representative for the ultimate resistance of the soil.

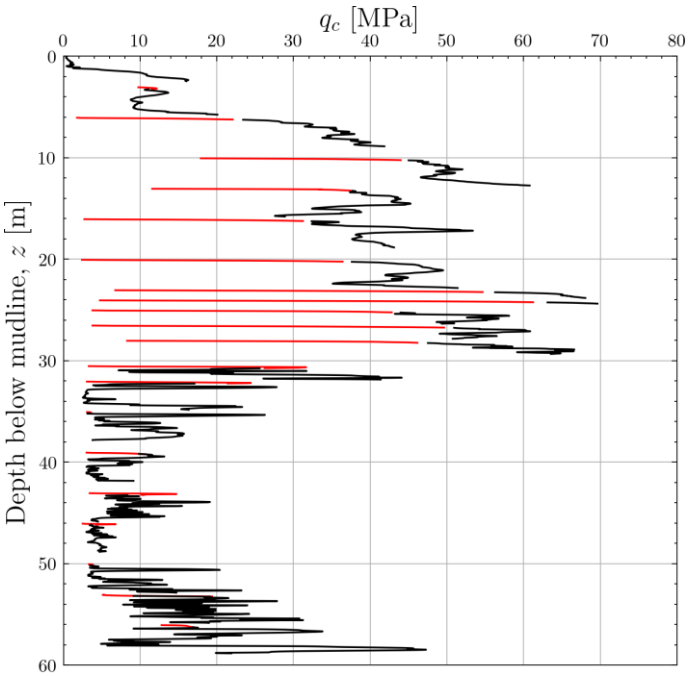


Figure 7. Application of the trained semi-supervised model to an unseen CPT. The initial part of the push identified by the model is marked in red.

2.3 *Unsupervised learning*

In unsupervised learning, the data is not labelled and the aim of the machine learning is to differentiate clusters with significantly different behaviour.

For geotechnical engineers, this is best illustrated with the clustering of offshore wind turbine locations across a project site. Locations with similar geotechnical conditions are grouped together based on several features (e.g. depth to a load bearing stratum, presence of soft soil). Depending on the geological setting at the site, the number of clusters required to capture the differences between the individual location groups will vary. At geologically homogeneous sites, a limited number of clusters may suffice whereas strong site heterogeneity may lead to more clusters being required.

In an unsupervised clustering analysis, the parameter space of the features is subdivided into a number of clusters based on similarity between the features of the individual data points. The most intuitive clustering algorithm is the K-means clustering algorithm. This algorithm calculates a generalised distance between the cluster centres and each datapoints (Equation 2). The optimal cluster centres are found when the distance between the points inside the cluster is small and the distance between cluster centres is large. Equation 2 shows the equation for the within-cluster sum-of-squares. For the n datapoints in the cluster C, the distance between the cluster centre μ_j and the individual datapoints is minimised. For datasets with significant scatter, identifying meaningful clusters can be challenging.

$$\Sigma_{i=0}^{n} \min_{\mu_j \in C} \left(\|x_i - \mu_j\|^2 \right) \quad (2)$$

An example of K-mean clustering for determining soil layers from CPT data is shown in Figure 8. The depth below mudline z, cone tip resistance q_c and soil behaviour type index I_c are used as the features for the unsupervised learning. The profiles of q_c and I_c were smoothed (see red dashed lines) to ensure homogeneous layers. If smoothing is not applied, the assigned cluster varies rapidly with depth and the layer homogeneity is not respected. The figure shows that the clustering algorithm is able to identify the individual layers. However, the layer boundaries, especially for the orange layer in the top 10m are debatable and might be assigned differently by an engineer.

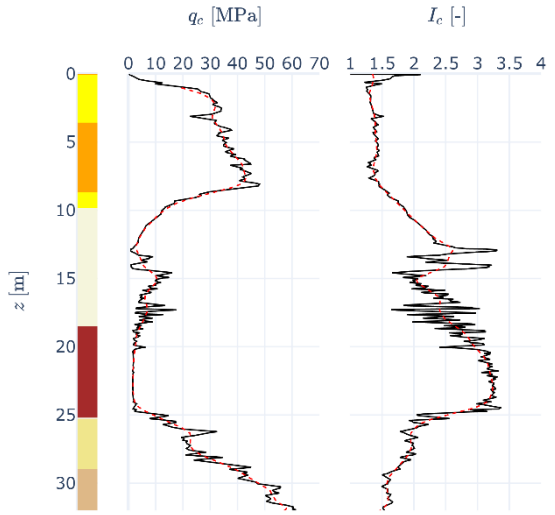


Figure 8. Determination of soil layers based on K-means clustering.

2.4 *Reinforcement learning*

Reinforcement learning is a type of machine learning that differs from both supervised and unsupervised learning approaches. In reinforcement learning, algorithms interact with an environment, which may be a real-world setting or a computer program. The actions taken by the algorithm impact the environment, and based on these actions, a reward may be given.

The training process does not rely on a predefined dataset. Instead, the model learns through repeated interactions with the environment, where it receives feedback in the form of rewards or penalties. This encourages the model to learn the optimal behaviour.

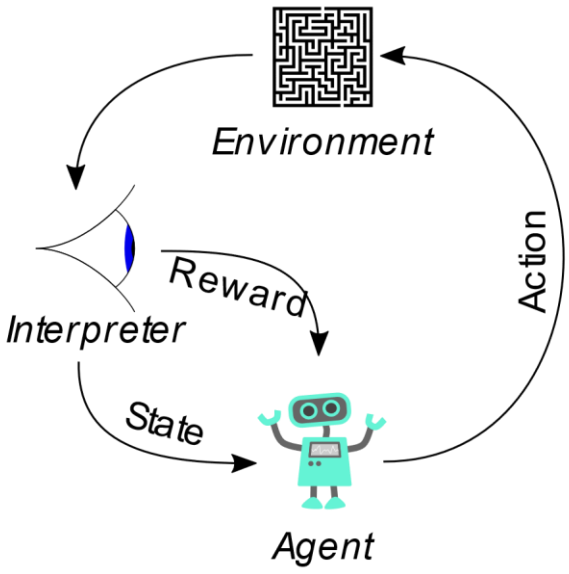


Figure 9. Schematic representation of reinforcement learning. An algorithm (agent) interacts with the environment and based on the interpreted outcome, a reward can be given (Image source: Wikimedia Commons)

Reinforcement learning is applied in areas where there is an opportunity to repeatedly interact with the environment. In situations where there is a clear reward such as games, algorithms have been very successful in learning optimal behaviour.

In the development of the control algorithms for autonomous vehicles, learning interactions with the environment in a controlled setting is necessary as the agent may select actions which lead to harmful interactions with the environment.

At the time of writing, there are no known applications of reinforcement learning in offshore geotechnical design. However, it is striking that many of the actions taken by operators of offshore equipment are inspired by the same principles as reinforcement learning. Experienced operators have learned what leads to a successful outcome by being exposed to many different situations (e.g. equipment settings, ground conditions, metocean climate) and have adapted their actions to these settings. It is plausible that control algorithms for offshore equipment (e.g. pile driving hammers, trenching assets) will eventually make use of reinforcement learning to make the equipment more autonomous and to reduce operator-dependence.

3 DATA REQUIREMENTS FOR MACHINE LEARNING METHODS

3.1 *Data quality*

Supervised and unsupervised learning methods heavily rely on the data that is presented to them during the training phase. Therefore, uncertainties in this data arising from measurement errors or statistical variations can have a significant impact on the resulting models. One example from offshore site investigations is the measurement of undrained shear strength S_u using relatively crude, manually operated tools (e.g. torvane, pocket penetrometer), intermediate quality tests such as the laboratory vane or the UU triaxial test and more accurate DSS and CU triaxial tests (Figure 10). Although these tests target the same geotechnical parameter, the test method has a significant impact on the accuracy of the S_u estimate.

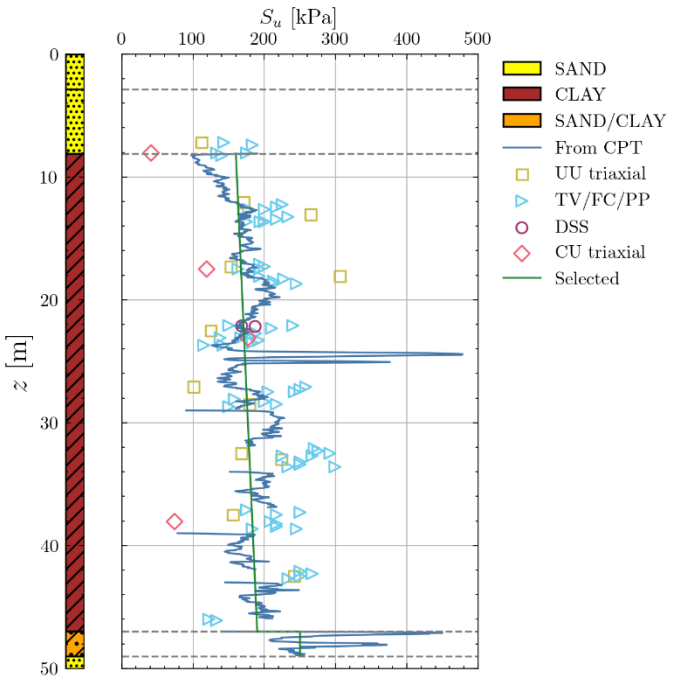


Figure 10. Undrained shear strength determined from various techniques.

Moreover, datasets can exhibit high levels of imbalance and bias if they contain more data for certain regions in the parameter space than others. For example, site investigation for offshore wind farms in the Dutch and German North Sea shows that cohesionless soils dominate in areas. Any attempt to train models which also need to predict the behaviour for cohesive soils will be affected by a lack of data.

In general, obtaining highly accurate data (e.g., from detailed computer simulations or costly site investigation tests) may not always be possible due to cost and time constraints. In such cases, lower-fidelity data are often used as an alternative (e.g., simplified physics-based model calculations instead of detailed computer simulations). Information from different sources of data with varying accuracy and acquisition costs can be combined using multi-fidelity data fusion techniques to provide more accurate predictions than using a single source of data. This approach offers several benefits, including cost savings by performing detailed computer simulations or carrying out costly site investigation tests only in critical areas where there is the greatest design uncertainty.

A series of standards for machine learning is currently under development by ISO/IEC with the ISO/IEC DIS 5259-1 standard covering data quality.

3.2 *Data quantity*

Machine learning models in areas such as image analysis, natural language processing or product recommendation are often trained on millions or even billions of samples. The cost of obtaining such data is relatively low and data can be readily found in the public domain. Computation challenges arising from the data volume are often more challenging than data acquisition.

Offshore geotechnical data is generally more costly to acquire as specialised vessels and equipment are required. Moreover, data is often confidential, especially when the behaviour of high-value foundation assets is being monitored. Knowledge on the site conditions and foundation behaviour offers a competitive advantage to the operator.

A typical offshore wind farm consists of 40 to 100 geotechnical borehole locations leading to a limited data volume for a single project. This limited amount of data imposes limitations on the modelling accuracy which can be achieved with machine learning methods. From probability theory, it is known than when a limited number of samples is available, the uncertainty on the estimated statistical parameters from the data will be high. This is further complicated by the geological origins of the site. Building models for geologically heterogeneous sites is more challenging as the number of samples in each geological unit will be smaller than for a geologically homogeneous site.

In supervised learning, the predicted target variable often depends on several features. This leads to the so-called *curse of dimensionality*; as the number of features considered in the model grows, the data becomes sparse and investigating the impact of each feature on the predicted target becomes more challenging.

Combining data across projects offers possibilities to expand datasets for machine learning models. Storing data in centralised databases which keep the data from legacy projects *alive* after project commissioning can allow insights to be developed on a much larger dataset. The recent initiatives by RVO, Crown Estate and BSH to bring offshore site investigation data into the public domain, allow an extensive publicly available dataset to be developed which can supplement any project-specific data in similar ground conditions. The available data is restricted to selected projects in the North Sea and Baltic Sea but this model could be adopted globally to develop a basic understanding of regional ground conditions.

ISSMGE TC209 is currently investigating the impact of data quality and quantity on foundation design models through a dedicated task group.

3.3 *Model quality*

A model will only be fit-for-purpose if it provides accurate predictions which also respect the overall trend in the data. If the model does not capture the trends identified in the data, it is said to *underpredict* the data. If the model perfectly predicts the training data but performs poorly on unseen data, it is said to *overpredict*. Figure 11 shows an example of three models which predict the same dataset. The linear model clearly does not capture the underlying trend while a decision tree model with a high tree depth provides

perfect predictions on the training data, leading to high-frequency oscillations which are not physically meaningful. The quadratic model strikes a balance between under and overfitting and is therefore the preferred model.

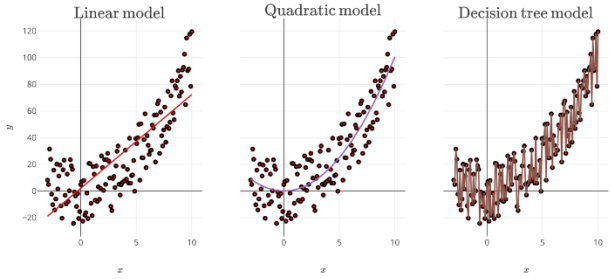


Figure 11. Under- and overprediction in machine learning models (Stuyts, 2020).

3.3.1 *Accuracy*

Calculating the model accuracy is the first component of a model quality assessment. Several metrics exist which quantify the error between the predicted value of the target and its measured value in supervised learning. The root mean squared error (RMSE) from Equation 3 is a much-used metric in regression modelling and it is often used as the loss function.

$$RMSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

For classification, the cross entropy loss is commonly used as the loss function. Alternatively, the accuracy of the model can be assessed by checking the ratio of the number of correct label predictions to the total number of training samples. The number of correct predictions for each class can be captured graphically through a *confusion matrix*. An example is shown in Figure 12. The predicted labels are shown on the X-axis and the true labels on the Y-axis. If the model is accurate (predicted label corresponding to the true label), the cells of the matrix diagonal will have the highest colour intensity.

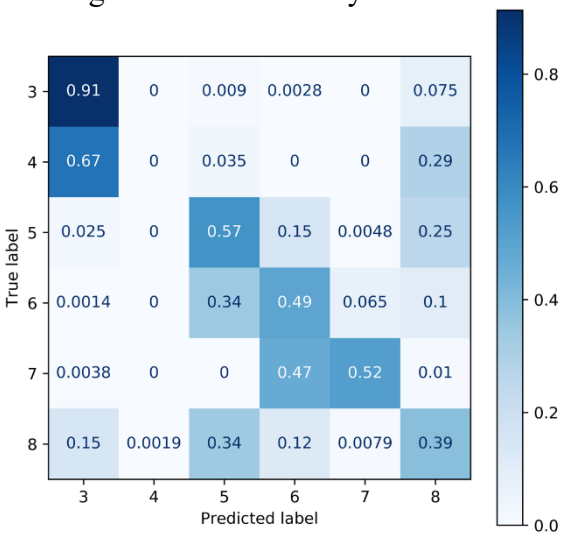


Figure 12. Confusion matrix for soil type class prediction (Stuyts, 2020).

3.3.2 *Generalisation*

Accurate predictions on the training data is only one of the components of assessing model quality. The model should also perform well on unseen data. For this purpose, the available dataset is partitioned in a training dataset and a test dataset. The test dataset is not used during the model training and after completing the training phase, the prediction accuracy is evaluated on this test set.

If the model *generalises* well, it will provide comparable accuracy on the test data as what was achieved for the training data. If the accuracy on the training data is significantly higher than the accuracy on the test data, the model is likely to overfit the training data. The decision tree model from Figure 11 has a very high accuracy on the training data but is not likely to be able to achieve the same accuracy on data it has not seen before.

3.4 *Model interpretability*

The advent of mathematically complex model such as deep neural networks or Extreme Gradient Boosting (XGBoost) algorithms has allowed engineers to train more accurate models. However, as the mathematical formulations of these models are too complex to inspect, it is not straightforward to understand the internal workings of a trained model.

Models with a low complexity and a closed-form mathematical expression allow the influence of the features on the target to be inspected quickly. When the mathematical equations capture physically meaningful trends, such models can be used for extrapolation beyond the training data.

For more complex models, extrapolations beyond the range of the features in the training dataset carries much more risk. Predictions can diverge or lead to physically inconsistent behaviour. Insight in the behaviour of the model needs to be gathered using alternative techniques. One possibility is to vary one of the input features and checking if the model output is physically meaningful. Figure 13 shows the predicted shear wave velocity for a model with a simple stress-dependence as given in Equation 3 and an XGBoost regression model. Both models are trained on a dataset of 2000 S-PCPT measurements from the Belgian and Dutch North Sea and show comparable accuracy. The model coefficients α and β in Equation 4 are expressed as linear functions of the soil behaviour type index I_c and the model coefficients are optimised using the available training data.

$$V_s = \alpha \cdot \left(\sigma'_{vo}/\sigma'_{ref}\right)^{\beta} \quad (4)$$

A profile of shear wave velocity is generated for two combinations of total cone tip resistance q_t and soil behaviour type index I_c (Figure 13). As expected, the higher cone resistance and lower I_c (lower fines content) leads to higher values of shear wave velocity.

However, the difference for the XGBoost model is not consistent with depth and the increase of shear wave velocity with depth does not follow the expected exponential law. In the shallow region above 5m depth, the shear wave velocity also shows significant oscillations. The training data did not have much data in this shallow region.

This shows that even when models can have comparable accuracy, the model which better captures the underlying physics will always lead to more consistent results. Any trained model should therefore be subjected to a thorough review to ensure that predictions are physically meaningful.

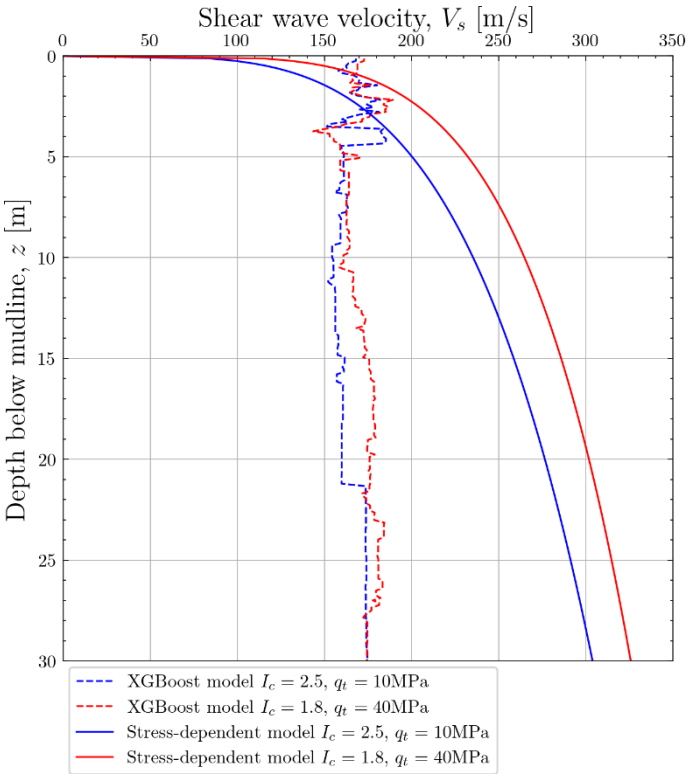


Figure 13. Variation of shear wave velocity predictions with depth for two combinations of cone tip resistance and soil behaviour type index for a machine learning with a simple stress dependence and an XGBoost model.

3.5 *Machine learning model development tracking*

The development of a machine learning model is an iterative task in which several model types, feature combinations, combinations of hyperparameters, train/test splits etc. are used. Keeping track of the different combinations can be challenging especially if model development is performed as a team effort.

In Machine learning operations (MLOps) (Kreuzberger et al., 2023), each version of the model is stored on a central server and each new iteration in model development is tracked. This keeps a history of the model development and improves the efficiency of the development process.

4 ADVANCED MACHINE LEARNING MODELS

While classification and regression models are being increasingly used in (offshore) geotechnical practice due to their readily available implementations in packages such as scikit-learn (Scikit-learn developers, 2015), a number of advanced model types are currently being researched.

The following sections describe selected model types and their possible applications in offshore geotechnical engineering.

4.1 *Machine learning with uncertainty prediction*

Most machine learning models provide deterministic predictions, which can be inadequate as they do not provide any information about the model's confidence in its predictions. However, there are machine learning models that can provide probabilistic predictions and uncertainty quantification for classification and regression problems. There are several key advantages of such models, namely (i) better decision making, (ii) improved trust, and (iii) improved robustness.

Uncertainty quantification can aid users in making more informed decisions. For instance, if a model predicts significant uncertainty around a certain output, users can give less weight to that prediction when making the final decision. Uncertainty quantification also increases user trust in the machine learning model by allowing them to understand when the model is unsure, thereby instilling more confidence in knowing when to trust or not trust a model's prediction. Additionally, probabilistic machine learning models are typically more robust to noisy data since the noise in the training data is explicitly modelled in the analysis. This makes model predictions more resilient to outliers and noisy training data.

There are two main types of uncertainty that contribute to overall uncertainty in a model's predictions: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty, also called ‘data uncertainty’ or ‘irreducible uncertainty’, refers to the inherent randomness or variability in the data, such as measurement errors or natural randomness in the underlying processes being studied. This type of uncertainty cannot be reduced, even with a perfect model. On the other hand, epistemic uncertainty, also known as ‘model uncertainty’ or ‘reducible uncertainty’, arises from incomplete knowledge of the system being modelled or the use of an imperfect model. This type of uncertainty can be reduced by improving the model or collecting more data.

Epistemic uncertainties can be modelled in neural networks using simple techniques such as ensemble or Dropout methods. The ensemble method (Lakshminarayanan et al., 2017) involves training multiple neural networks with different architectures. The

variance among the predictions of the ensemble can then be used as a measure of epistemic uncertainty. Dropout (Gal and Ghahramani, 2016) is another technique used to model epistemic uncertainty. It is a regularisation technique where randomly selected neurons are 'dropped out' during training. It has been shown that dropout can be interpreted as a Bayesian approximation in deep Gaussian Processes, allowing for the quantification of epistemic uncertainty.

To explicit model both aleatoric and epistemic uncertainty in neural networks, Bayesian Neural Network (BNN) can be used (Lampinen and Vehtari, 2001). BNN applies probabilistic modelling of the network's weights and biases, resulting in a posterior distribution over the model parameters. This allows for the quantification of epistemic uncertainty. Aleatoric uncertainty can be considered by the network by explicitly modelling noise in the training data. Markov Chain Monte Carlo (MCMC) is commonly used to estimate the posterior distribution of network's parameters and make probabilistic predictions. MCMC is a computational method used in Bayesian inference to approximate complex probability distributions by iteratively sampling from them. It works by creating a Markov chain that converges to the posterior distribution of the parameters after a large number of iterative sampling steps.

There are different types of MCMC algorithms, which are used to explore the parameter space efficiently (Andrieu et al., 2003). Three common sampling algorithms are Metropolis-Hastings, Gibbs, and Hamiltonian Monte Carlo (HMC). The Metropolis-Hastings algorithm proposes a new step in the chain based on a proposal distribution, and the proposal is accepted with a probability that depends on the ratio of the posterior probabilities of the current state and the proposed state. Gibbs sampling updates one parameter at a time while keeping all other parameters fixed. The algorithm iteratively samples from the conditional probability distributions of each parameter given the other parameters. HMC is an advanced MCMC algorithm that makes use of gradient information to propose new steps in the chain.

However, MCMC can be computationally intensive and require careful tuning to ensure convergence to the correct distribution. Thus, alternative methods have been proposed to approximate the posterior distribution. One such method is variational inference, which approximates the posterior distribution using a simpler distribution that can be optimised using gradient-based methods (Graves, 2011). Nevertheless, MCMC can be readily applied to simpler models with few parameters (e.g. simple linear regression models) without much computational complexity, which can provide valuable insights into the posterior distributions of the model parameters (Stuyts et al., 2012).

Another popular probabilistic machine learning model for regression problems is Gaussian Process (GP) regression (Williams and Rasmussen, 2006). Unlike deterministic machine learning models that try to identify the best estimate of a single function that models the relationship between input and output, a GP regression model provides a probability distribution over the possible functions that can fit the data. Therefore, GP regression provides not only a point estimate of the output (through the mean of the predictions), but also a measure of epistemic uncertainty (through the confidence interval of the predictions). Besides their ability to quantify the uncertainty of the predictions, GP regression models have other advantages. For instance, they can model a wide range of complex and non-linear relationships between inputs and outputs. GP regression models do not make any assumptions about the functional form of the relationship and can capture patterns that are difficult to model with traditional regression techniques.

4.2 *Physics-informed machine learning*

In scientific applications, machine learning poses a significant challenge as it may not always follow the physical principles of the systems it is applied to, resulting in physically inconsistent predictions, particularly when extrapolating from the training dataset. To enable machine learning models to learn the physical principles from data, a considerable amount of training data is required, which can be costly to acquire. Physics-informed machine learning (PIML) (Karniadakis et al., 2021) seeks to overcome these limitations by integrating prior knowledge and physical principles into machine learning models.

PIML offers two significant advantages: physically consistent and scientifically sound predictions for unseen scenarios, and effective training with less data than a purely data-driven machine learning model. There are three main approaches to integrate physical principles and prior knowledge into the machine learning models: (i) hard constraints; (ii) soft constraints, and (iii) data augmentation.

Hard constraints are conditions that must be satisfied exactly by the model, while soft constraints are conditions that the model attempts to satisfy to the best of its ability. Soft constraints are implemented by including the conditions as additional regularisation terms in the loss function used for training the machine learning model, as shown in Figure 14. This will penalise models that violate the conditions and encourage the model to satisfy the conditions during training. Common types of soft constraints include initial conditions, boundary conditions. symmetry principles and partial differential equations that should be satisfied within the input domain. This method is made popular by the Physics-Informed Neural Networks (PINN) proposed by Raissi et al. (2019), which allows for easy computation of the partial derivatives of the output using automatic differentiation, and has been demonstrated to provide more robust

predictions, even in the presence of noisy data. A schematic diagram of the PINN is shown in Figure 14.

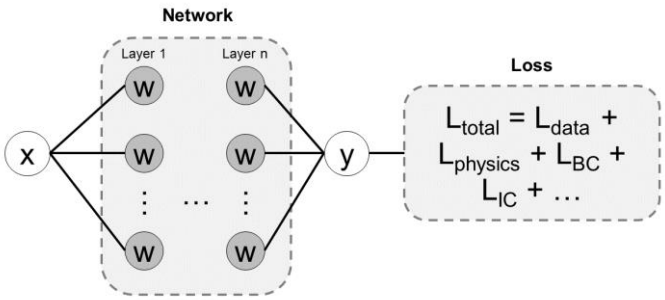


Figure 14. Schematic diagram of the PINN, where x and y are the input and output, respectively. L_{data}, $L_{physics}$, L_{BC} and L_{IC} refer to the losses due to data-predictions mismatch, physical conditions, boundary conditions and initial conditions, respectively.

Implementing hard constraints can be more challenging, and proposed methods for implementation include the augmented Lagrangian method (Lu et al. 2021) or designing specific components of the machine learning model to meet the conditions exactly (Zhu et al. 2021). Data augmentation is simple to implement, as it involves the generation of synthetic data based on physical principles to augment the training dataset. This approach is particularly useful when available data is limited or noisy, such as in medical image analysis (Omigbodun et al. 2019; Dahiya et al. 2021).

The most commonly used approach among these three is the soft constraints method, primarily due to its ease of implementation. Nevertheless, this method still presents some challenges, such as the possibility of unsatisfied conditions, resulting in physically inconsistent predictions.

4.3 *Time series predictions with LSTM*

The machine learning models that have been discussed so far are designed to predict a target variable based on input features, where each new prediction is independent of previous ones. However, there are some problems in which the value of the target depends on the feature values of the previous samples. Time series prediction is an example where the value of the target at a certain point in time is influenced by the feature values of a number of previous samples. Figure 15 shows the time series of pore pressure around an offshore monopile. The low-frequency tidal variations can clearly be seen in the data and high-frequency oscillations due to wave actions are observed in the signal for the shallowest depth. A predictive model for pore pressures at time t requires knowledge of the measured pore pressure at preceding times.

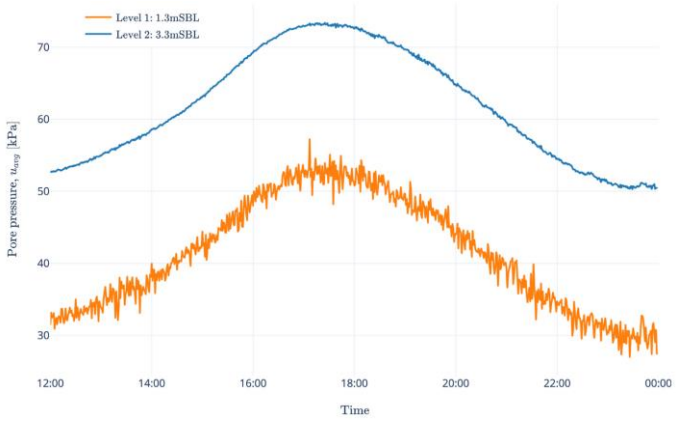


Figure 15. Pore pressures around an operational offshore monopile measured at two levels (1.3m and 3.3m) below mudline (Stuyts et al., 2023).

Recurrent Neural Networks (RNN) achieve this by storing information from preceding time steps in a hidden state. Figure 16 shows a schematic representation of an RNN. The hidden state $a^{<t-1>}$, which is the result of the time preceding time t, is combined with the feature values $x^{<t>}$ to predict the target value $y^{<t>}$. It is clear that the hidden state incorporates the information from the time steps preceding the time step under consideration.

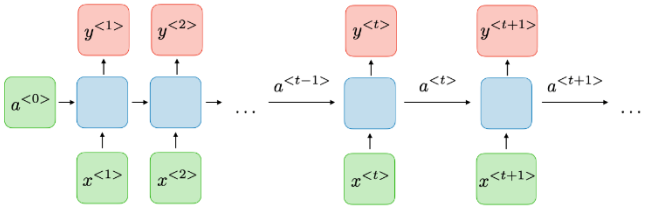


Figure 16. Schematic representation of a recurrent neural network (Amidi & Amidi, 2022).

The weights W_a in the neural network (shown as the blue block in Figure 16) are shared between all time steps, leading to computational efficiency. When training the model, the weights in the neural network are optimised. This requires the calculation of the derivative of the loss function with respect to the model weights for each time step. However, low-frequency can be difficult to capture with conventional RNNs as the gradients can combine to lead to exponential increases or decreases of the model weights. This is known as the *vanishing/exploding gradient* problem.

To mitigate this problem, additional functionality was added to the blue cells from Figure 16. Long Short-Term Memory (LSTM) units (Hochreiter and Schmidhuber, 1997) add a cell state $c^{<t>}$ to the RNN cells and use four gates to update the hidden and cell states based on the feature values in the timestep:

- Forget gate Γ_f: Determines whether the information from previous time steps should be forgotten. When sudden changes in response are observed, the forget gate can erase the effects of preceding time steps;
- Update gate Γ_u: Determines how much the feature values $x^{<t>}$ affect the updating of the hidden and cell states;

- Relevance gate Γ_r: Decides which information from the preceding time steps is relevant;
- Output gate Γ_o: Updates the hidden state based on the previous and current cell states.

At each gate, *activation functions* are used. These functions transform the output of the gate to a well-defined range and thus stabilise the model. The sigmoid function transforms the output to a value between 0 and 1, while a tanh activation function transforms the output to a value between -1 and 1.

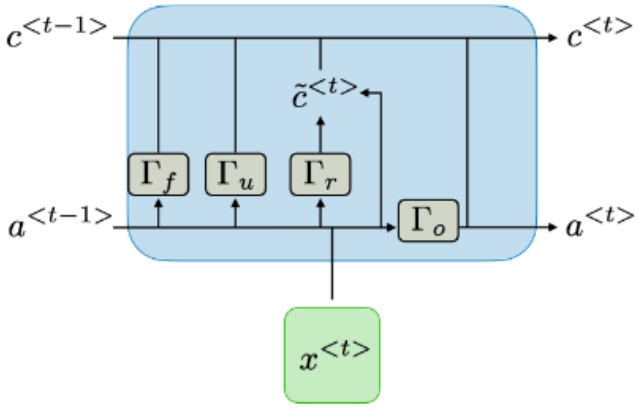


Figure 17. Schematic representation of an LSTM cell (Amidi & Amidi, 2022).

With an LSTM model, both low- and high-frequency oscillations can be learned by the model, as long as sufficient data is provided during the training phase. While time series prediction may be relevant for certain time-dependent geotechnical problems such as in-place monitoring of foundations, the LSTM techniques can also be determined to depth profiles of geotechnical parameters. By taking replacing the time axis by the depth axis, trends of parameter variation with depth can be learned by the model.

4.4 *Multi-fidelity data fusion*

Multi-fidelity data fusion (MFDF), also known as multi-fidelity modelling or information fusion, is a technique for combining information from multiple data sources with varying levels of accuracy and sparsity to produce more accurate and reliable predictions. This approach is valuable when high-fidelity data are difficult or expensive to obtain, while lower-fidelity data sources are more readily available.

MFDF allows the use of simplified models to make low-cost but potentially inaccurate predictions (known as "low-fidelity predictions") of some geotechnical design performance. These predictions can then be 'corrected' by fusing a small set of accurate but high-cost predictions (known as "high-fidelity predictions") obtained from experiments or more detailed computational simulations such as three-dimensional (3D) finite element |(FE) modelling. This is achieved by using machine learning to learn and exploit the cross-correlations between the low- and high-fidelity predictions. Figure 18 illustrates the training and operational phases of the MFDF approach. MFDF has been shown to provide greater predictive accuracy than machine learning models trained on high-fidelity predictions alone, and requires much less training data (Meng and Karniadakis, 2020). MFDF is particularly advantageous in situations where acquiring high-fidelity predictions is expensive, such as in the design of large-scale foundation projects like offshore wind farms. It can provide comparable accuracy to 3D FE modelling for the entire wind farm site, while only requiring 3D FE calculations for a small fraction of the total number of foundations.

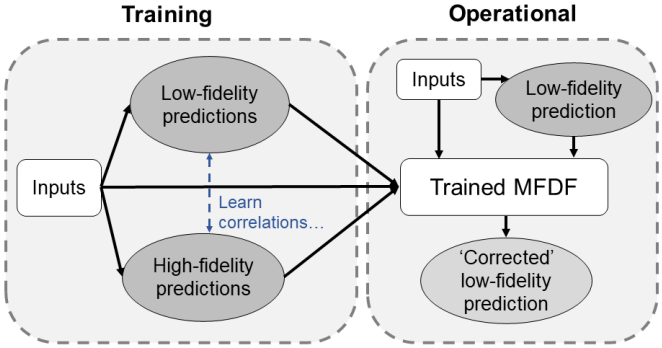


Figure 18. Schematic diagram of the training and operational phases of the MFDF model

MFDF has been successfully applied in aerospace engineering, such as in the design of space vehicles (Minisci et al., 2011). It is only recently that there is a nascent interest in MFDF for geotechnical applications (Dey et al., 2021; Zhang et al., 2022). Various MFDF methods have been developed (Peherstorfer et al., 2018), with those based on Gaussian Process (GP) being the most popular because they can quantify the uncertainty of fused predictions. The most widely used GP-based MFDF method is that proposed by Kennedy and O'Hagan (2001), which proposes the following relationship between the low- and high-fidelity predictions:

$$y_{HF}(x) = \rho\, y_{LF}(x) + \delta(x) \tag{5}$$

where $y_{LF}(x)$ and $y_{HF}(x)$ are GP regression models trained on the low- and high- fidelity predictions. ρ is a constant multiplicative correction factor and $\delta(x)$ is a GP regression model that represents an additive correction factor. Figure 19 provides an illustrative example of MFDF, where Equation 5 is used to combine information from the seemingly uninformative low-fidelity data points with the sparse high-fidelity data points to predict the true function that generated the high-fidelity data points. The prediction obtained through MFDF is in excellent agreement with the true function and compares favourably to the inaccurate prediction obtained from a standard GP regression model trained only on the high-fidelity data.

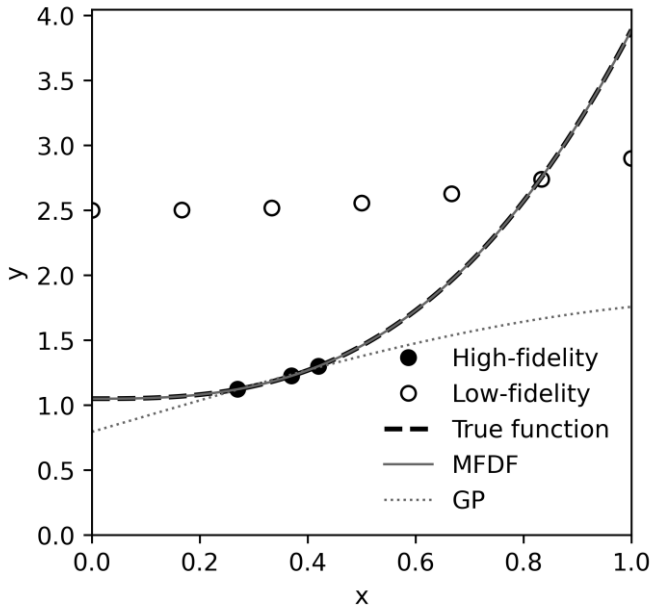


Figure 19. Illustrative example of a MFDF prediction, where the low- and high-fidelity data points are used to predict the true function that generates the high-fidelity data. The prediction by a GP regression model trained on only the high-fidelity data is also shown for comparison.

Equation 5 can only capture linear correlations between the low- and high-fidelity predictions, but more advanced GP-based MFDF methods have been developed to model non-linear correlations. For instance, Perdikaris et al. (2017) proposed the NARGP model, while Cutajar et al. (2019) introduced the Deep GP model. However, a key limitation of GP-based MFDF methods is that they may not perform well for high-dimensional problems. To address this, MFDF methods based on deep neural networks have been proposed as alternatives (Meng et al., 2020; Pawar et al., 2022). With the growing amount of data from multiple sources, there is a need to combine the information from these sources effectively to make more accurate and reliable predictions. Therefore, MFDF is expected to play an increasingly important role in various scientific and engineering applications.

4.5 *Large language models*

Since its release at the end of 2022, ChatGPT has sparked a remarkable increase in the use of Large Language Models (LLMs) in practice. LLMs belong to a class of machine learning models called *generative AI* and differentiate themselves from other machine learning models by being able to generate large sections of text (*response*) when being presented with an input question (*prompt*).

LLMs are trained on very large amounts of text which was collected from the internet. GPT-3 (Brown et al., 2020) was trained on 45TB of text data and the model has 175 billion parameters. The computational costs associated with training the model is estimated to be around 4 million USD.

The initial Generative AI models are based on Generative Adversarial Networks (GANs). As shown schematically in Figure 20, these models have a generator component G which generates a response based on a prompt Z. The generated response as then compared to the real response using the discriminator D. The model parameters are optimised until the generated responses are sufficiently accurate. Most LLMs are based on a more advanced type of neural network called a transformer.

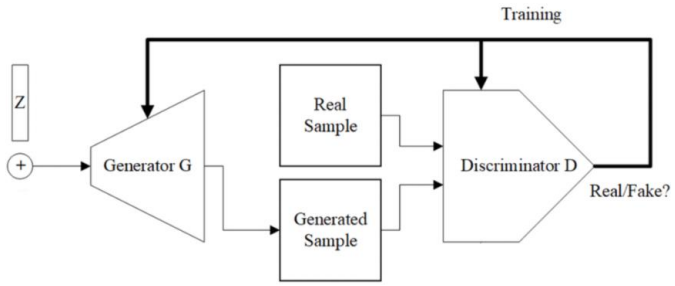


Figure 20. Representation of a Generative Adversarial Network (GAN) (Wikimedia Commons CC BY-SA 4.0)

LLMs tokenize and encode text as part of the natural language processing. Tokenisation breaks down a piece of text into smaller chunks called tokens, which can be words or subwords. Encoding is the process of representing these tokens numerically, typically using vectors or matrices, that can be processed by the neural network. Text is generated by predicting the most likely sequence of tokens.

When making use of LLMs, it is important to understand that they were trained on data which is publicly available on the internet. Any wrong information included in the public domain may persist in the responses generated by the model (Si et al., 2022). Moreover, *hallucinations* can occur in which the model presents false responses as truth. Hallucinations can occur for a variety of reasons, such as when the model is given ambiguous input, or when the model is overfitting to specific biases in the training data. Hallucinations is a significant challenge in developing LLMs, especially when the goal is to produce trustworthy text that is relevant to the input.

LLMs can be improved with specific subject matter expertise. For example, OpenAI provides the possibility to update the GPT-3 (OpenAI, 2022) with new subject material knowledge. Verified combinations of prompts and responses can be presented to the model to improve its accuracy in specific knowledge domains. Such a curated set of prompts and responses could be gathered under the auspices of a professional organisation such as ISSMGE.

5 MACHINE LEARNING WITH UNCERTAINTY QUANTIFICATION: SMALL-STRAIN STIFFNESS PREDICTIONS

Small-strain stiffness (G_{max}) is an increasingly important parameter in offshore foundation design, especially if the stiffness of the soil is driving the design. For monopile foundations, the parameter is included in the recent PISA guidance (Burd et al., 2020).

The small-strain stiffness can be measured in-situ with the seismic CPT test (S-PCPT) or in the laboratory with the bender element or resonant column test. However, due to a limited site investigation budget, site-specific G_{max} measurements are not available at every foundation location. CPT-based correlations were developed to allow location-specific stiffness profiles to be estimated. These correlations are calibrated semi-empirical models (e.g. Rix & Stokoe, 1991) which are trained on locations where both G_{max} and conventional CPT measurements are available.

Given the uncertainty on the measurement and the imperfect knowledge of the underlying physical principles, the data inevitably shows scatter around the predicted trend.

Rix & Stokoe (1991) proposed the formula in Equation 6 which allows estimate of the average small-strain shear modulus G_{max}, cone tip resistance q_c and vertical effective stress level σ'_{vo}. The two calibration parameters α and β were derived from calibration chamber testing on washed mortar sand with a median grain size of 0.35mm and less than 1% fines. Values of $\alpha = 1634$ and $\beta = -0.75$ are proposed.

The correlation was developed for near-surface soils with a depth less than 13m. The authors propose an uncertainty band on the predicted average of ±50%.

$$\left(\frac{G_{max}}{q_c}\right)_{avg} = \alpha \cdot \left(\frac{q_c}{\sqrt{\sigma'_{vo}}}\right)^{\beta} \tag{6}$$

The predictions from the correlation were checked against a dataset of 1749 G_{max} measurements on clean silica sand obtained with the S-PCPT in the Belgian and Dutch North Sea (Peuchen et al., 2019; Stuyts et al., 2022). Figure 21 shows the available data relative to the trend proposed by Rix & Stokoe (1991). The prediction clearly underestimates the actual small-strain shear modulus and a significant proportion of the data plots above the upper bound of the uncertainty band of the correlation (upper black dashed line).

Bayesian modelling can be used to describe the uncertainty on the model parameters α and β. Instead of predicting a single value for the calibration parameters, their uncertainty is characterised through a probability distribution. Based on the available data, the correlation can be recalibrated probabilistically. Posterior distributions can be generated using the MCMC algorithm (Salvatier et al., 2016).

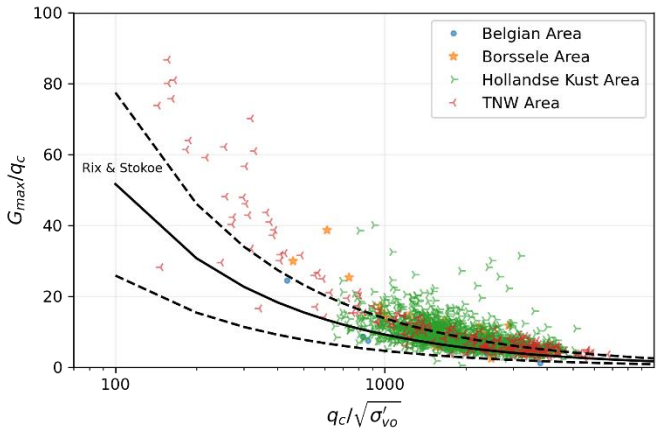


Figure 21. Overview of the available data small-strain shear modulus data on clean silica sands from the Belgian and Dutch North Sea plotted together with the proposed trend by Rix & Stokoe (1991).

If the correlation from Equation 6 is assumed to apply, the Bayesian modelling can be described in relatively simple terms. By taking the logarithm of each side of the equation, a linear equation is obtained (Equation 7). $\log_{10}(\alpha)$ is the intercept which can be written as α' and β is the slope. The dependent variable $\log_{10}(G_{max}/q_c)$ can be written as y and the independent variable $\log_{10}\left(q_c/\sqrt{\sigma'_{vo}}\right)$ as x

$$\log_{10}\left(\frac{G_{max}}{q_c}\right) = \log_{10}(\alpha) + \beta \cdot \log_{10}\left(\frac{q_c}{\sqrt{\sigma'_{vo}}}\right) \tag{7}$$

This linear regression model can be written in probabilistic terms as shown in Equation 8. This equation describes that the dependent variable is normally distributed around a mean value described by the deterministic Equation 7. The mismatch between the predicted average and the measurements is characterised with an error term ε.

$$y = \mathcal{N}(\mu = \alpha' + \beta \cdot x, \varepsilon) \tag{8}$$

Bayes' theorem was applied to infer the posterior distributions of calibration coefficients α and β given the available G_{max} data from the Belgian and Dutch North Sea. Figure 22 shows the posterior distributions with the proposed average coefficients by Rix & Stokoe (1991) as grey vertical lines. The mean value of the distribution of α (α_{mean}=2242) has shifted to higher values compared to the initially proposed correlation. The value of β is only slightly modified by the Bayesian updating.

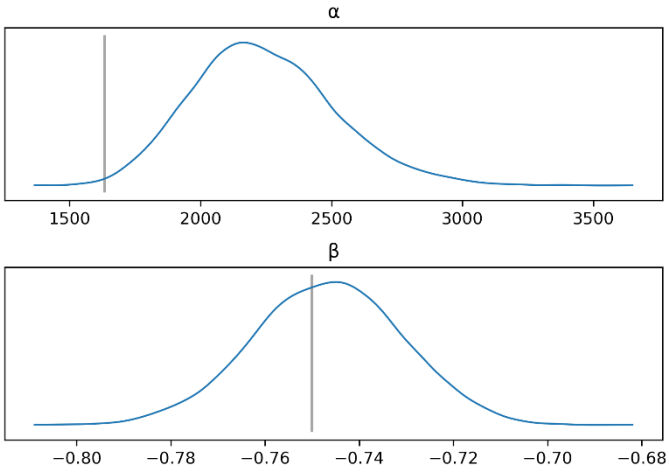


Figure 22. Posterior distribution of calibration coefficients α and β obtained using MCMC. The proposed averages by Rix and Stokoe (1991) are shows as grey vertical lines.

The predictions with the updated model are shown in Figure 23. The prediction of the mean is improved by the Bayesian updating and the uncertainty on the mean can be assessed (shown as the grey shaded area). This uncertainty increases with lower values of $q_c/\sqrt{\sigma'_{vo}}$. The model uncertainty can also be quantified by computing the difference between the model predictions and the data. The standard deviation of the model uncertainty is shown with black dashed lines in Figure 23

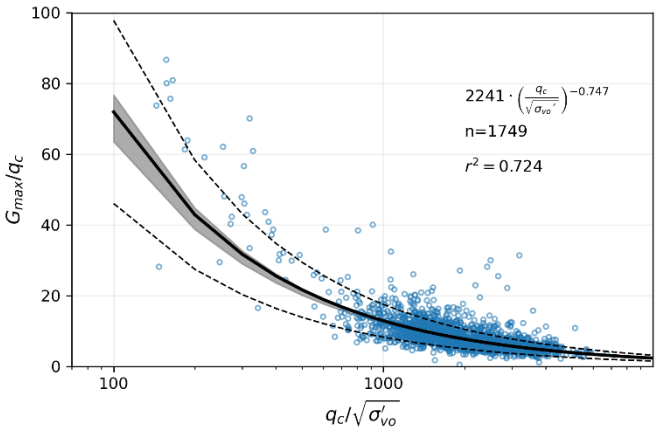


Figure 23. Recalibrated correlation for clean silica sand. The uncertainty on the mean is shown as the grey band and the standard deviation of the correlation is shown with dashed black lines.

By using Bayesian updating the model predictions are improved and information on the distribution of the calibration parameters is obtained. This makes the model suitable for further probabilistic studies.

6 PHYSICS-INFORMED MACHINE LEARNING EXAMPLE

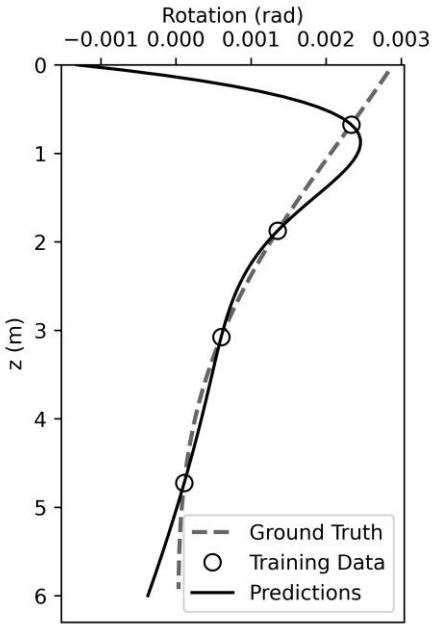
To demonstrate the benefits of PIML over purely data-driven machine learning models, a PINN will be compared with a purely data-driven neural network (PDNN) for the prediction of the rotation and bending moment profiles of a monopile. This approach may be useful for situations where there are limited inclinometer readings available for interpreting the rotation profile. However, in this example, a synthetic dataset was used instead of real-world data, as it provides the ground truth data that can be used to assess the accuracy of the PINN and PDNN. The synthetic dataset for the rotation and bending moment profiles is obtained using a Winkler modelling framework for a monopile of diameter 1m, embedded length of 6m and wall thickness of 0.01m.

The monopile is assumed to be embedded in a Dunkirk sand profile (Zdravković et al. 2020) and the monopile structure is modelled using Timoshenko beam elements, while the soil response is modelled using the PISA soil reactions for sand (Burd et al. 2020). The monopile is subjected to a horizontal load of 80kN and a moment load of 640kNm at ground level, resulting in the rotational and bending moment profiles labelled as 'Ground Truth' in Figure 24. The training dataset for both PINN and PDNN consists of the monopile cross-section rotation and bending moment values at four depths, as shown in Figure 24. This may seem insufficient for effective learning of neural networks. This example will demonstrate how incorporating physical principles into the neural network can improve the training data efficiency.

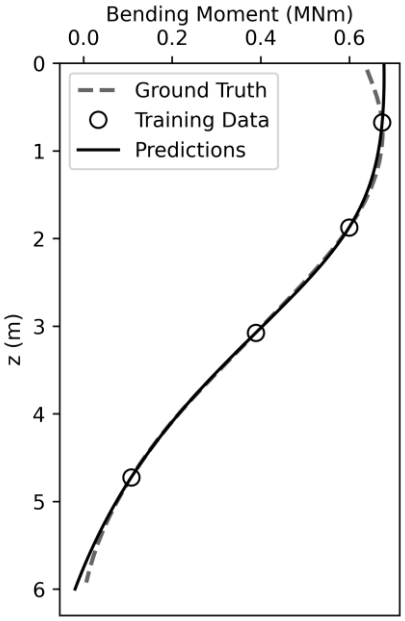
The PDNN consists of three hidden layers, each with 32 neurons and a hyperbolic tangent activation function. The input for PDNN is the depth below ground level (z), and the output is either the rotation of the monopile cross-section (θ) or the bending moment of the monopile (M) at a particular depth. The training process Involves optimising the weights and biases of PDNN using the ADAM optimizer to minimise the loss function of PDNN, as shown below:

$$L_{\text{total}} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\text{pred}} - y_i^{\text{true}}\right)^2 \quad (9)$$

where y represents the output (either θ or M). 30,000 epochs were completed during the training process. Figure 24 shows the rotation and bending moment profiles of the monopile that were predicted by the trained PDNN. Although the bending moment profile predictions of PDNN are reasonably accurate, the rotation profile predictions are inaccurate when compared to the ground truth. In order to enhance the accuracy of the predictions, the neural network will be improved by incorporating physical principles and prior knowledge about the bending moment, resulting in the formation of a PINN.



(a)



(b)

Figure 24. Comparison of PDNN predictions of the (a) monopile cross-section rotation profile and (b) bending moment profile, with the ground truth profiles. The training data are also shown in the figure.

The PINN has the same architecture and training dataset as the PDNN. The only difference is the loss function, which now incorporates the physical principle that connects θ and M, as follows:

$$M = -EI\frac{\partial \theta}{\partial z} \tag{10}$$

Furthermore, the loss function also incorporates prior information about the boundary condition of the bending moment at ground level, which should be equivalent to the applied moment:

$$M_{z=0} = 640 \text{ kNm} \tag{11}$$

Therefore, the complete loss function of PINN that is minimised using the ADAM optimiser is as follows:

$$L_{\text{total}} = L_{\text{data}} + L_{\text{physics}} + L_{\text{BC}} \tag{12}$$

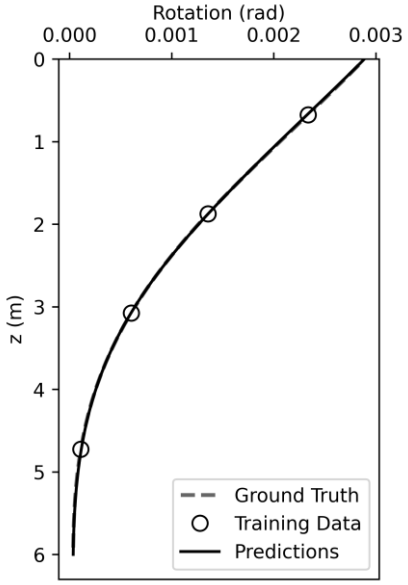
where

$$L_{\text{data}} = \frac{1}{N}\sum_{i=1}^{N}\left(\theta_i^{\text{pred}} - \theta_i^{\text{true}}\right)^2 \tag{13}$$

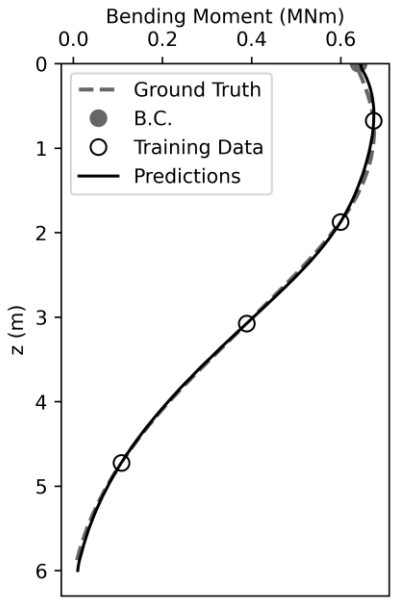
$$L_{\text{physics}} = \frac{1}{N}\sum_{i=1}^{N}\left(M_i^{\text{true}} + EI\frac{\partial \theta_i^{\text{pred}}}{\partial z}\right)^2 \tag{14}$$

$$L_{\text{BC}} = \left(EI\frac{\partial \theta^{\text{pred}}}{\partial z}\Big|_{z=0} + 640\right)^2 \tag{15}$$

During the training process, 30,000 epochs were completed. The output of PINN is the rotation profile of the monopile, which is then used to calculate the bending moment profile using Equation 10. Figure 25 shows the rotation and bending moment profiles of the monopile that were predicted by the trained PINN. The predictions are in excellent agreement with the ground truth, despite the availability of only four training data points. This highlights the much greater training data efficiency provided by PIML.



(a)



(b)

Figure 25. Comparison of PINN predictions of the (a) monopile cross-section rotation profile and (b) bending moment profile, with the ground truth profiles. The training data and boundary condition (B.C.) are also shown in the figure.

7 PILE DRIVING PREDICTIONS WITH LSTM

7.1 *Available data*

LSTM models can be used to capture time-dependent processes in which the response at future timesteps depends on the response in a number of preceding timesteps. This paradigm is also applicable for depth-dependent geotechnical problems in which predictions of the response at deeper depths can be made on the basis of observations at shallower depths.

In a pile driving problem, the soil resistance to driving (SRD) depends on the skin friction and end bearing on a pile. As the shaft area in contact with the soil increases with depth, the shaft resistance component of the SRD will also increase. At a given depth, the shaft resistance will be influenced by the unit skin friction in the layers above. SRD cannot be measured directly but the energy requirement for an increment of penetration can be used instead.

A dataset with the pile driving data for 114 tubular piles in North Sea was previously used in a community prediction exercise (Stuyts, 2020). This dataset included location-specific pile dimensions, CPT data and driving records. The data is available on a regular grid with 0.5m penetration increments. This dataset is used here to demonstrate the LSTM model capabilities. LSTM requires the features and target to be formulated as time-series. To achieve this, the location-specific data was concatenated as shown in Figure 26. The features used in the modelling are the depth below mudline z, the cone tip resistance q_c, sleeve friction f_s, pore pressure at the cone shoulder u_2 and the normalised energy E/E_0. As blowcount is an operator-dependent variable, the normalised energy required for a penetration increment $(E/E_0)/\Delta z$ was used as the target.

The models are trained on 70 locations and their performance is evaluated on the remaining 44 unseen locations.

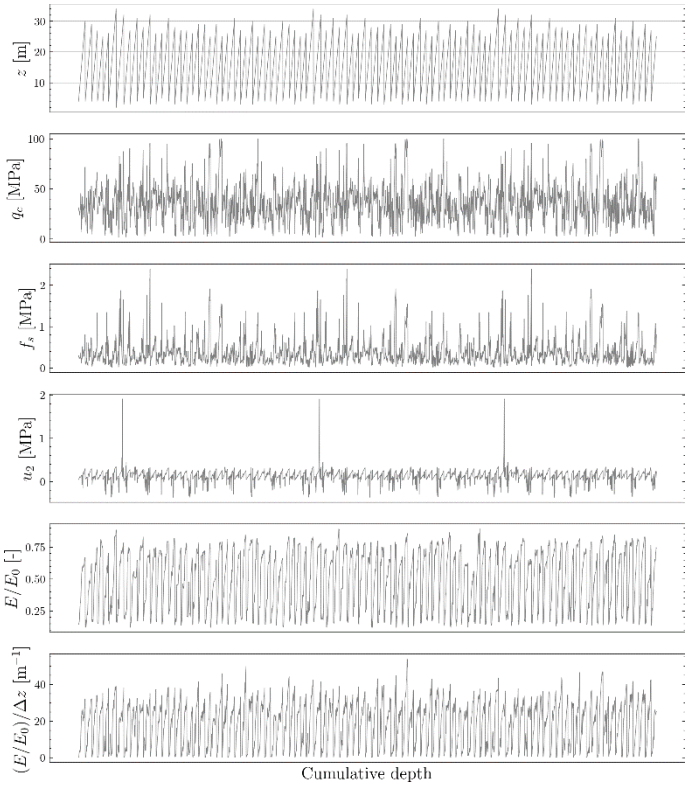


Figure 26. Series of pile driving data used in LSTM modelling.

7.2 *LSTM model types*

LSTM models can be used for various types of predictions. Depending on the model specification, the algorithm will be trained to predict the target for m future steps depending on the feature and target values at n previous steps. Figure 27 shows three modelling strategies applied to the pile driving dataset, with inputs in blue and predictions in orange:

- Single step model: The target is predicted for step t based on the values of the features and target in the previous step $t-1$;
- Multi-step model: The target is predicted for the next step (t) based on the values of the features and target in 7 preceding steps ($t-7$ to $t-1$) (1.5xOD above the current pile penetration);
- Forward prediction model: The target is predicted for the next four steps (t to $t+3$) based on the value of the features and target in 7 preceding steps ($t-7$ to $t-1$).

This type of modelling is different from a conventional prediction in which the energy required per penetration increment would be predicted for the entire profile. Instead, this modelling results in models which can be used during pile driving operations to anticipate the pile driving behaviour at deeper depths depending on the driving performance already observed. The LSTM models learn the trends from the training data and apply these in the prediction step to make forward-looking predictions.

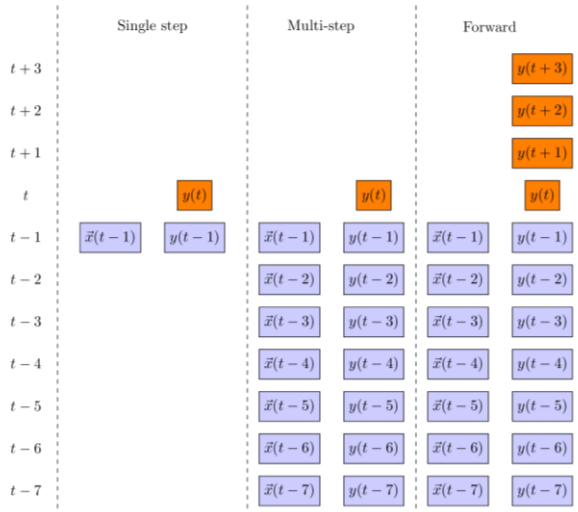


Figure 27. Model training strategy for three types of LSTM model. The predicted targets are shown in orange and the features and target used as input are shown in blue.

The three LSTM models were developed with Keras (Chollet et al., 2015) as a sequential model with a hidden layer with 50 LSTM units and an output layer with a number of nodes equal to the number of predicted steps. All models show convergence towards a minimum RMSE which was comparable for both the training and test set. This indicates that the models did not overfit the data.

7.3 *Model results*

7.3.1 *Single step vs multi-step model*

The predictions for the single step and multi-step LSTM models were evaluated for an unseen location in the test set. The predictions of each model are shown in Figure 28 together with the measured target (blue dashed line). The results for the single step and multi-step model are shown as red squares and green circles respectively. It should be noted that each prediction is based on one or more observations at shallower depths, so only the target value at the depth considered are unknown to the model.

The result show that both models are able to follow the trend observed in the measurements. As the models are only looking forward at the next 0.5m penetration increment, this is not surprising. It is also clear from the figure that the multi-step model provides slightly better predictions. Considering information from multiple preceding penetration increments leads to an improved ability of the model to replicate the behaviour observed during driving. By considering information in the 1.5xOD above the current penetration, the multi-step model is able to better learn the expected behaviour. This was also observed in the accuracy on the entire test set where the multi-step model has a lower RMSE than the single step model.

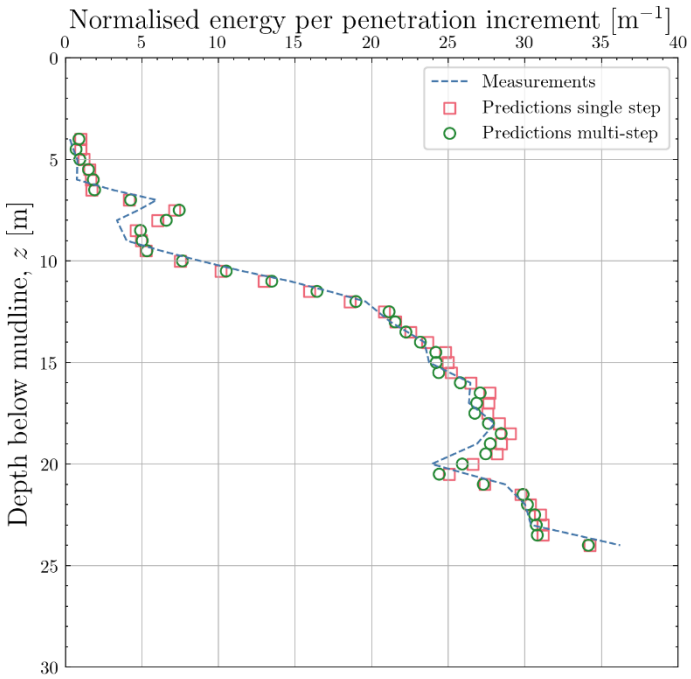


Figure 28. Predictions for the single step LSTM model (red squares) and multi-step LSTM model (green circles) at an unseen location. The observations at this unseen location are shown as a blue dashed line.

7.4 *Forward predictions for multiple steps*

When predicting the expected behaviour for multiple forward steps, the LSTM model needs to extrapolate from the known behaviour based on the trends it has learned from the training data.

Predictions with the forward LSTM model were made at four different penetrations. At each penetration, the target was predicted for the next four 0.5m increments. The predictions are shown in Figure 29. When compared to the observed target, it is observed that the prediction for the first 0.5m increment is still relatively close to the observations. However, the predictions for deeper increments show greater deviations from the expected target. This is especially true when the observed normalised energy per penetration increment shows sudden changes (e.g. between 7 and 8m depth). The model is not able to replicate those based on the trends it has learned from the training data. It is able to predict an evolution of the target which corresponds to the general trend but is unable to make detailed predictions. This is commonly observed in LSTM models; the errors accumulate the further the model is extrapolating from the known data.

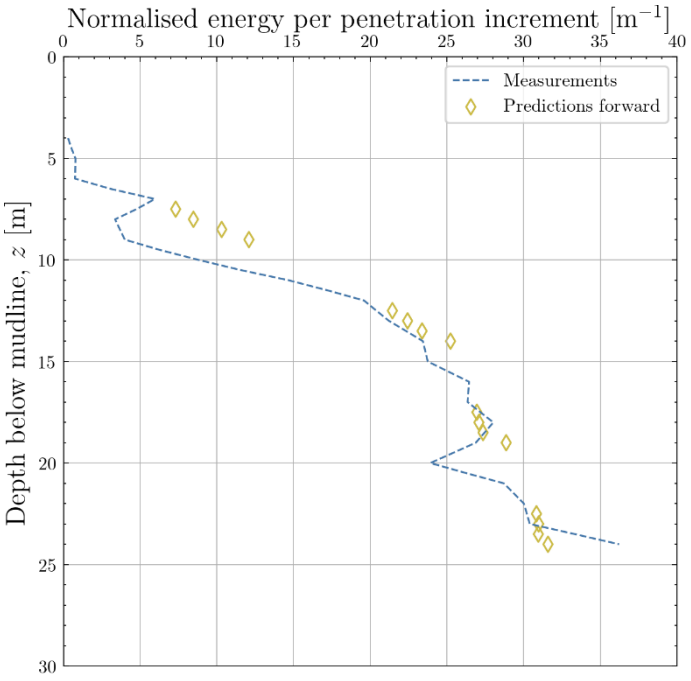


Figure 29. Predictions for the forward LSTM model at an unseen location. The predictions at 4 different penetrations are shown as yellow diamonds. The observations at this unseen location are shown as a blue dashed line.

Although LSTM models can be useful to track pile driving operations in real-time at project sites with relatively uniform ground conditions, predicting the behaviour for highly heterogeneous conditions will be challenging for any machine learning model. Use cases where contractors deploy these models in real time to justify encountering unforeseen ground conditions are more likely than an adoption during the installation planning phase where the pile driving behaviour needs to be predicted along the entire penetrated pile length.

8 MULTI-FIDELITY DATA FUSION FOR SUCTION CAISSON STIFFNESS

When designing offshore wind foundations, a combination of simplified design models and detailed 3D FE modelling is commonly used. Simplified models are employed for computationally intensive tasks such as wind farm-scale foundation design optimisation, while 3D FE modelling is used for design verification or more precise, site-specific design calculations for locations with more complex ground conditions. To demonstrate the benefits of MFDF, the following example will demonstrate how the accuracy of simplified design model calculations can be improved by combining them with a small set of 3D FE calculations. This will be illustrated using the example of estimating the vertical stiffness of a suction caisson foundation.

In this example, the vertical static stiffness for a suction caisson of length-to-diameter ratio $L/D = 1$ in various three-layered soil configurations (see Figure 30) is investigated, where the shear modulus in each layer may vary according to:

$$G_i = G_i^{\text{ref}} \left(\frac{2z}{D}\right)^{\alpha} \tag{10}$$

where $i = 1, 2$ or 3, $0 \leq \alpha \leq 1$ and G_i^{ref} is the reference shear modulus in each layer. In total, 108 different soil profiles are analysed, comprising of various combinations of $\alpha = \{0, 0.5, 1\}$, soil Poisson's ratio $\nu = \{0, 0.1, 0.2, 0.3, 0.4, 0.49\}$, and $G_i^{\text{ref}}/G_R = \{1, 2, 4\}$ where $G_R = 1000 p_{\text{atm}}$. Figure 31 shows an example of one of the soil profiles analysed.

The low-fidelity predictions for the vertical stiffness of the caisson in each soil profile were determined using OxCaisson (Suryasentana et al. 2022). OxCaisson is a Winkler model for suction caissons that models the soil response using local Winkler-type 'soil reactions', which mechanically function like independent springs acting on the caisson. There are two types of soil reactions: distributed soil reactions that act along the caisson skirt length, and concentrated soil reactions that act at the caisson base. This study employed OxCaisson as it can be used directly for any shear modulus profile, since the soil reactions are defined using local soil properties. However, it should be noted that these local soil reactions are dependent on the global shear modulus profile (Suryasentana et al. 2022). Therefore, the static stiffness estimations by OxCaisson is not likely to be accurate for the layered soil profiles examined in this study, as these profiles are not found in the calibration dataset.

The corresponding high-fidelity predictions of the vertical stiffness were obtained by 3D FE modelling with Abaqus v6.13 software (Dassault Systèmes 2014). The 3D FE model used in this study is the same as the one described in Suryasentana et al. (2022). Figure 32 shows a comparison between the high- and low-fidelity predictions for all the layered soil profiles analysed in this study. The results indicate an approximately linear correlation between the low- and high-fidelity predictions. Thus, the GP-based MFDF method (i.e., Equation 5) proposed by Kennedy and O'Hagan (2001) is used for data fusion.

To demonstrate the high-fidelity training data efficiency of the MFDF approach, only 13% of the high-fidelity predictions and their corresponding low-fidelity predictions are used as training data for the MFDF model. During the training process, the MFDF model learns the correlations between the low- and high-fidelity predictions. Once trained, the MFDF models can use the learnt correlations to 'correct' the inaccuracy of the low-fidelity predictions. The inputs into the MFDF model are $G_1^{\text{ref}}, G_2^{\text{ref}}, G_3^{\text{ref}}, \nu, \alpha$, while the output is the vertical stiffness of the caisson.

Figure 33 presents a comparison between the 'corrected' low-fidelity predictions and the corresponding high-fidelity predictions for all of the layered soil profiles examined in this study. The predictions made by a standard GP regression model, which was trained

only on the high-fidelity data and referred to as 'GP-HF,' are also shown in Figure 33 to demonstrate the advantages of fusing low- and high-fidelity data. The 'corrected' low-fidelity predictions made by the trained MFDF model are in good agreement with the high-fidelity predictions and are much better than the GP-HF predictions. Therefore, the MFDF approach allows for accurate predictions to be made at a fraction of the computational cost by combining low- and high-fidelity predictions.

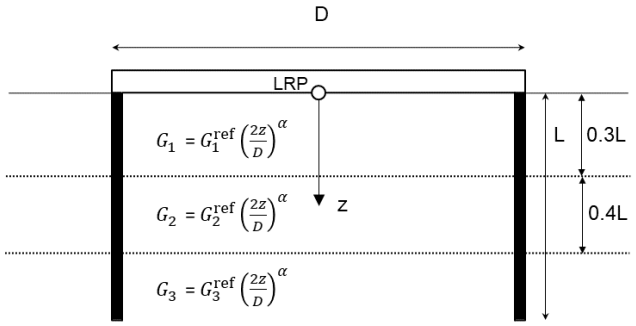


Figure 30. Schematic diagram of three-layered soil configurations investigated in this example. LRP is the loading reference point, L is the embedded length and D is the diameter.

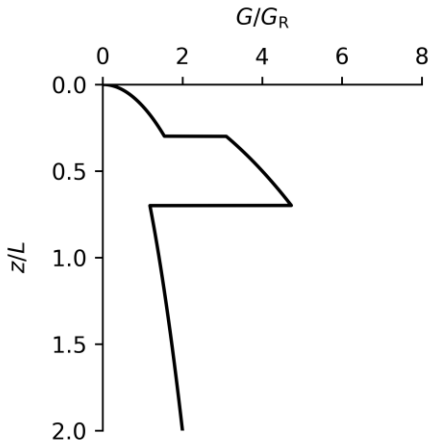


Figure 31. An example of a three-layered soil profile, where $\frac{G_1}{G_R} = 2\left(\frac{2z}{D}\right)^{0.5}, \frac{G_2}{G_R} = 4\left(\frac{2z}{D}\right)^{0.5}, \frac{G_3}{G_R} = \left(\frac{2z}{D}\right)^{0.5}$

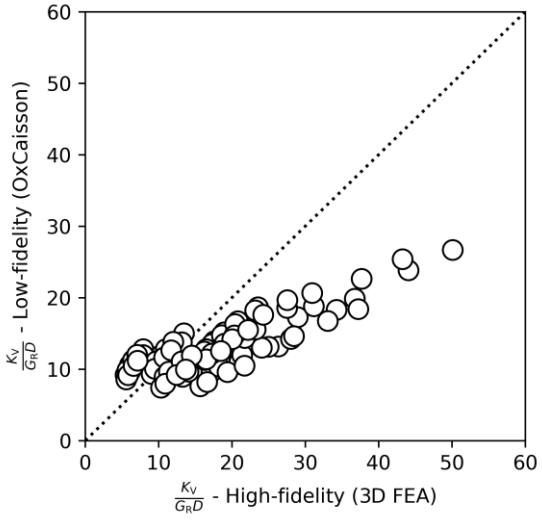


Figure 32. Comparison of low- and high-fidelity predictions of the caisson vertical stiffness.

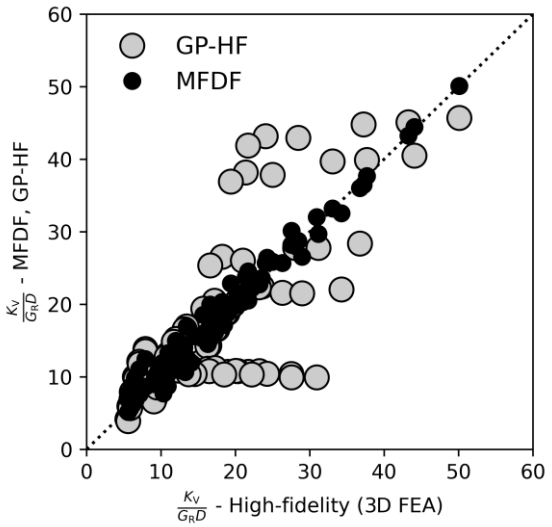


Figure 33. Comparison of 'corrected' low-fidelity predictions of the caisson stiffness made by the MFDF model, with the actual high-fidelity predictions and the predictions made by a GP regression model trained on the high-fidelity data alone (referred to as 'GP-HF' in the figure).

9 CONCLUSIONS AND FUTURE PERSPECTIVES

While the use of machine learning methods is gaining traction in research and offshore practice, there are still a number of challenges to overcome before they can be routinely applied. This paper presents the basics of unsupervised (clustering) and (semi-)supervised (classification and regression) machine learning methods. Various offshore geotechnical engineering problems can be formulated as machine learning problem. An appropriate choice of model type depends on the available input data and the nature of the question that needs to be answered. A good knowledge of the available model types and their background is required.

Machine learning models learn behaviour from the trends revealed by the data fed to them during the training phase. The quality and quantity of the data determines the quality of the machine learning models trained on those data. The offshore wind industry offers possibilities for assembling large datasets of geotechnical conditions and foundation behaviour but without the use of proper database technology, there is a potential for loss of valuable data.

When conducting measurement campaigns offshore, there are a lot of factors (e.g. environmental, operator dependence, choice of equipment) which can affect the data quality. If the data shows significant scatter, the machine learning algorithms may not have sufficient accuracy. Care should also be taken to not overfit the training data as this may lead to physically non-meaningful behaviour. Techniques exist to evaluate how well a model captures the expected physical behaviour and in engineering problems, these techniques should always be applied before using a model in production.

Conventional regression and classification techniques have a strong similarity to the techniques that were already used in geotechnical research before the advent of modern data science libraries. However, advances in machine learning research has led to a number of more complex model types which can also be applied to geotechnical engineering problems. Quantifying the uncertainty on parameter estimates made with a machine learning model, is possible with the use of MCMC techniques or by using probabilistic machine learning models such as Bayesian Neural Networks and GP regression models. An application to correlations of small-strain shear modulus data with conventional CPT data, shows that the coefficients of these correlations can be updated probabilistically when site-specific G_{max} data is available to provide an updated estimate of confidence intervals of geotechnical parameters.

Machine learning models in geotechnical engineering can be improved substantially if knowledge of the underlying physics is considered during the model development. This can be performed explicitly using physics-informed neural networks. An example is provided for a monopile subject to lateral load in which the equations of pile bending can be explicitly accounted for in the loss function. The inclusion of physical considerations and prior knowledge into the loss functions can result in more robust models that provides physically consistent predictions with much less training data requirements.

The formulation of a machine learning problem depends on the intended use of the model. In time-dependent problems, the user will want to make forward-looking predictions based on a number of past observations. Recurrent Neural Networks, and more specifically models built with LSTM blocks, allow the models to learn trends from past measured behaviour. LSTM models are developed for pile driving predictions. While the driving behaviour can be accurately predicted for a small window ahead of the current pile penetration, getting accurate predictions for larger windows remains challenging. Further research and data gathering are required before such models can replace a thorough understanding of the mechanics of pile driving.

One of the most promising applications of machine lies in the combination of predictions of geotechnical behaviour from various data sources. It is common for geotechnical engineers to have both highly advanced numerical models and relatively fast but more simplified models at their disposal. The principle of multi-fidelity data fusion allows predictions from both model types to be combined to achieve higher quality estimates of the actual foundation behaviour. This is demonstrated for an application in which the stiffness of a suction caisson foundation is modelled with two different model types. The predictions of the simpler model are corrected based on the advanced numerical analyses and the results show that high-quality estimates of foundation stiffness can be obtained at a fraction of the computational cost.

To develop advanced machine learning models and apply them in practical field applications, it is crucial to have a deep understanding of their underlying principles. It is also important to subject these models to a rigorous checking process to ensure that they produce reliable and interpretable results. Moreover, it is essential to caution users against extrapolating results beyond the range of the training data. Failure to do so may result in the misuse of black-box models, which may substitute geotechnical expertise and experience, leading to unreliable predictions and decisions.

10 ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Belgian Ministry of Economic Affairs through the ETF project WINDSOIL project. The support of VLAIO through the De Blauwe Cluster SBO SOILTWIN project is also acknowledged. The second author wishes to express gratitude to the Royal Society and ETP for providing funding support for his research activities focused on the application of machine learning in offshore geotechnics.

11 REFERENCES

Amidi, A., Amidi, S. (2022). Recurrent Neural Network Cheatsheet. CS 230 Deep Learning. URL https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks#

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., others (2020). Language models are few-shot learners. Advances in neural information processing systems 33, 1877–1901.

Burd, H.J., Taborda, D.M., Zdravković, L., Abadie, C.N., Byrne, B.W., Houlsby, G.T., Gavin, K.G., Igoe, D.J., Jardine, R.J., Martin, C.M. and McAdam, R.A. (2020). PISA design model for monopiles for offshore wind turbines: application to a marine sand. Géotechnique, 70(11), 1048-1066.

Burd, H.J., Abadie, C.N., Byrne, B.W., Houlsby, G.T., Martin, C.M., McAdam, R.A., Jardine, R.J., Pedro, A.M.G., Potts, D.M., Taborda, D.M.G., Zdravković, L., Andrade, M.P. (2020). Application of the PISA design model to monopiles embedded in layered soils. Géotechnique 70, 1067–1082. https://doi.org/10.1680/jgeot.20.PISA.009

Cutajar, K., Pullin, M., Damianou, A., Lawrence, N., and González, J. (2019). Deep gaussian processes for multi-fidelity modeling. arXiv preprint arXiv:1903.07320.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., (2015). Xgboost: extreme gradient boosting. R package version 0.4-2 1, 1–4.

Chollet, F. & others, (2015). Keras. Available at: https://github.com/fchollet/keras.

Dahiya, N., Alam, S. R., Zhang, P., Zhang, S. Y., Li, T., Yezzi, A., & Nadeem, S. (2021). Multitask 3D CBCT-to-CT translation and organs-at-risk segmentation using physics-based data augmentation. Medical Physics, 48(9), 5130-5141.

Dassault Systemes (2014). Abaqus user manual. Simula Corp., Providence, RI.

Dey, S., Chakraborty, S., and Tesfamariam, S. (2021). Multi-fidelity approach for uncertainty quantification of buried pipeline response undergoing fault rupture displacements in sand. Computers and Geotechnics, 136, 104197.

Doherty, J., Gourvenec, S., Gaone, F., Pineda, J., Kelly, R., O'Loughlin, C., Cassidy, M., Sloan, S., (2018). A novel web based application for storing, managing and sharing geotechnical data, illustrated using the national soft soil field testing facility in Ballina, Australia. Computers and Geotechnics 93, 3–8.

Hegazy, Y.A., Mayne, P.W. (2006). A global statistical correlation between shear wave velocity and cone penetration data, in: Site and Geomaterial Characterization. pp. 243–248.

Hochreiter, S., Schmidhuber, J. (1997). Long short-term memory. Neural computation 9, 1735–1780.

Jain, A.K., Mao, J., Mohiuddin, K.M. (1996). Artificial neural networks: A tutorial. Computer 29, 31–44.

Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). Physics-informed machine learning. Nature Reviews Physics, 3(6), 422-440.

Kennedy, M. C., and O'Hagan, A. (2001). Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425-464.

Kreuzberger, D., Kühl, N., Hirschl, S. (2023). Machine learning operations (mlops): Overview, definition, and architecture. IEEE Access.

McKinsey & Company (2022). What are Industry 4.0, the Fourth Industrial Revolution, and 4IR? McKinsey explainers. URL https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir

Meng, X. and Karniadakis, G. E. (2020). A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. Journal of Computational Physics, 401, 109020.

Minisci, E., Vasile, M., and Liqiang, H. (2011). Robust multi-fidelity design of a micro re-entry unmanned space vehicle. Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering, 225(11), 1195-1209.

Omigbodun, A. O., Noo, F., McNitt-Gray, M., Hsu, W., and Hsieh, S. S. (2019). The effects of physics-based data augmentation on the generalizability of deep neural networks: Demonstration on nodule false-positive reduction. Medical physics, 46(10), 4563-4574.

OpenAI (2022). Fine-tuning GPT-3. URL https://platform.openai.com/docs/guides/fine-tuning

Pawar, S., San, O., Vedula, P., Rasheed, A., and Kvamsdal, T. (2022). Multi-fidelity information fusion with concatenated neural networks. Scientific Reports, 12(1), 5900.

Peherstorfer, B., Willcox, K., and Gunzburger, M. (2018). Survey of multifidelity methods in uncertainty propagation, inference, and optimization. Siam Review, 60(3), 550-591.

Perdikaris, P., Raissi, M., Damianou, A., Lawrence, N. D., and Karniadakis, G. E. (2017). Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 473(2198), 20160751.

Peuchen, J., Meijninger, B.M., Brouwer, D. (2019). North Sea as geo database. AIMS Geosci 5, 66–81.

Lu, L., Pestourie, R., Yao, W., Wang, Z., Verdugo, F., and Johnson, S. G. (2021). Physics-informed neural networks with hard constraints for inverse design. SIAM Journal on Scientific Computing, 43(6), B1105-B1132.

Raissi, M., Perdikaris, P., and Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational physics, 378, 686-707.

Rix, G.J., Stokoe, K.H. (1991). Correlation of initial tangent modulus and cone penetration resistance, in: Calibration Chamber Testing. New York: Elsevier. Presented at the ISOCCT, Potsdam

Robertson, P.K. (2009). Interpretation of cone penetration tests-a unified approach. Canadian Geotechnical Journal 46, 1337–1355.

Salvatier, J., T. V. Wiecki, & C. Fonnesbeck (2016). Probabilistic programming in python using pymc3. PeerJ Computer Science 2, e55

Scikit-learn developers (2015). Scikit-learn user guide (http://scikit-learn.org/0.16/_downloads/scikit-learn-docs.pdf).

Si, C., Gan, Z., Yang, Z., Wang, S., Wang, J., Boyd-Graber, J., Wang, L. (2022). Prompting gpt-3 to be reliable. arXiv preprint arXiv:2210.09150.

Stuyts, B. (2020). Data science applications in geo-intelligence, in: Proceedings of the Fourth International Symposium Frontiers in Offshore Geotechnics. Presented at the ISFOG2020, Austin, Texas.

Stuyts, B., Jurado, C.S., Bautista, D.G., Kheffache, A. (2022). Bayesian estimation of small-strain shear modulus from offshore CPT tests in the North Sea.

Stuyts, B., Weijtjens, W., Devriendt, C. (2022). Development of a semi-structured database for back-analysis of the foundation stiffness of offshore wind monopiles. Acta Geotechnica. https://doi.org/10.1007/s11440-022-01551-3

Stuyts, B., Weijtjens, W., Gkougkoudi-Papaioannou, M., Devriendt, C., Troch, P., Kheffache, A. (2023). Insights from in-situ pore pressure monitoring around a wind turbine monopile. Ocean Engineering 269, 113556. https://doi.org/10.1016/j.oceaneng.2022.113556

Suryasentana, S. K., Burd, H. J., Byrne, B. W., and Shonberg, A. (2022). A Winkler model for suction caisson foundations in homogeneous and non-homogeneous linear elastic soil. Géotechnique, 7and2(5), 407-423.

Zdravković, L., Jardine, R.J., Taborda, D.M., Abadias, D., Burd, H.J., Byrne, B.W., Gavin, K.G., Houlsby, G.T., Igoe, D.J., Liu, T. and Martin, C.M. (2020). Ground characterisation for PISA pile testing and analysis. Géotechnique, 70(11), 945-960.

Zhang, P., Yin, Z. Y., Jin, Y. F., Yang, J., and Sheil, B. (2022). Physics-informed multifidelity residual neural networks for hydromechanical modeling of granular soils and foundation considering internal erosion. Journal of Engineering Mechanics, 148(4), 04022015.

Zhu, Q., Liu, Z., & Yan, J. (2021). Machine learning for metal additive manufacturing: predicting temperature and melt pool fluid dynamics using physics-informed neural networks. Computational Mechanics, 67, 619-635.