Application of deep neural network combined with dynamic poroelasticity to define seismic velocities and porosity from cone penetrometer data

Héctor Marín-Moreno[1#], *Susan* Gourvenec[2], and *Jared* Charles

[1]*University of Southampton, School of Ocean and Earth Science, Waterfront Campus, European Way SO14 3ZH, UK*
[2]*University of Southampton, School of Engineering, Boldrewood Campus, Burgess Road, SO16 7QF, UK*

[#]*Corresponding author: h.marin-moreno@soton.ac.uk*

ABSTRACT

Offshore wind plays a pivotal role in enhancing Europe's energy security and achieving energy decarbonization goals. However, expediting offshore wind deployment necessitates efficient and economical site investigation surveys. To address this challenge, we introduce a novel approach utilising a deep neural network (DNN) to establish correlations between geotechnical cone penetrometer test (CPT) data and shear wave velocity (V_s) from seismic CPT. Subsequently, porosity and P-wave velocity (V_p) are derived using a V_s to bulk density correlation and a dynamic poroelastic model. The DNN is trained and tested on a dataset comprising 5284 instances of public-domain geotechnical CPT test data, including depth, tip resistance, sleeve friction, and V_s from seismic CPT. During testing, the DNN model demonstrates a mean absolute error of 55 m s^{-1} between predicted and measured V_s values. The uncertainty in V_s predictions is attributed to factors such as (i) limited training data for some soil types such as gravelly sands, (ii) intricate relationship between geotechnical CPT features and seismic properties influencing V_s, (iii) the presence of CPT features and V_s combinations that lie well outside the region from most combinations (i.e. outliers), and (iv) CPT features and V_s measurements that are averaged over different depth ranges. The derived porosity and V_p exhibit centimetre-scale resolution, facilitating improved alignment between soil unit depths inferred from geophysical and CPT data. Future work will focus on refining the DNN through filtering, pre-processing, and incorporating additional site data and input features such as pore fluid pressure and porosity measurements.

Keywords: Offshore wind; subsurface data integration; deep neural network; dynamic poroelasticity.

1. Offshore wind context

To meet the targets of the Paris Agreement, offshore wind capacity needs increasing from 40 GW in 2020 to 2000 GW by 2050 globally (IRENA, 2021; GWEC, 2021) and from 12 GW in 2020 to 60 GW by 2030, with a vision for 340 GW installed by 2050, in the EU (EU-COM, 2020).

In the report *"How to succeed in the expanding global offshore wind market"* from McKinsey (2022) the following question is implicit: *how can the world achieve such sharp increase in offshore wind deployment?* Achieving 2000 GW of installed offshore wind capacity requires more than 500000 km^2 of ocean (Putuhena et al., 2023), roughly twice the size of the UK. Currently, it takes an average of 5-10 years for an UK offshore wind farm to become operational. Six months saving from this time in a UK offshore wind farm with a capacity of 1GW, by eliminating or reducing a phase of the soil investigation, could reduce 2.2 Mt of CO_2 equivalent in emissions by getting the GW onstream sooner, relative to the same energy generated from fossil fuels (532 tCO_2e per GWh versus 6 tCO_2e from offshore wind; DESNZ, 2023; Orsted, 2023).

Offshore wind turbine foundations contribute to about 8% of the levelized cost of energy (LCOE) of an offshore wind farm. This is £280 million (assuming monopile foundations) for 1 GW wind farm with 10 MW turbines at 30 to 40 m water depth (Catapult, 2019). Dogger Bank, the biggest offshore wind farm operating in the UK, has an installed capacity of 3.6 GW supplied by 277 turbines of 14.7 MW with a cost of about £1 billion. Even a small saving of 1% in turbine foundation costs from implementing less conservative designs driven by a better understanding on geotechnical engineering parameters and their variability could generate an initial cost saving of £10 million; and potentially more significant additional savings by ensuring an optimized design for stiffness and capacity to increase reliability and fatigue life.

All foundation and anchor designs require geotechnical parameters as input, which are currently acquired through an offshore geotechnical site investigation with in-situ testing and sampling for later laboratory testing. Current methods have evolved from the offshore hydrocarbons industry and are not fit for purpose to characterise the much larger areas required for offshore wind, or provide the versatility needed for micro-siting turbines within a farm post-survey.

Acceleration of offshore wind deployment requires optimised and more digital approaches and design strategies that can bring cost-reduction from technological upscaling and minimise environmental impact (Greaves et al., 2022).

2. Background

2.1. Offshore geophysical and geotechnical site survey

Compressional, i.e. P-wave, seismic reflection 2D data is routinely acquired, and 3D data is increasing acquired, during the site selection phase of an offshore wind project. Two-dimensional shear, i.e. S-wave, data are not collected for offshore site survey, but 1D profiles can be acquired by seismic cone penetrometer testing. One-dimensional profiles of mechanical seabed resistance are acquired by a standard cone penetrometer test (CPTu), which measures tip cone resistance (q_c), sleeve friction (f_s), and generally penetration pore fluid pressure (u).

P-wave data and the associated P-wave velocity (V_p), is affected by both the response of the pore fluid and the soil skeleton, which complicates the interpretation of the data and correlation with geotechnical properties. S-wave data and the associated S-wave velocity (V_s), is primarily affected by the response of the soil skeleton and provides a more direct link to the geotechnical properties shear stiffness of the soil at small strains (G, sometimes termed dynamic shear stiffness or G_{max} in the geotechnical literature) and undrained shear strength (Vardy et al., 2017). S-wave velocity and G are key parameters for the prediction of ground deformation and soil-foundation response to cyclic loading. P- and S-wave velocity can be defined by the bulk and shear moduli of the soil at small strains (K and G) and the bulk density of the soil (ρ) through Eq. 1a and b, respectively:

$$V_p = \sqrt{\frac{K+{}^4/_3 G}{\rho}} \quad (1a)$$

$$V_s = \sqrt{\frac{G}{\rho}} \quad (1b)$$

The application of advanced techniques such as elastic full waveform inversion to (ultra) high frequency reflection seismic data can solve decimetre-scale changes in V_p (e.g. Provenzano et al., 2018). However, such data is not routinely acquired and, if it is available, such advanced techniques are not common practice due to the computational and personnel expense. Currently, smooth V_p distributions of the interval velocities derived from normal move-out correction (with metre to decametre-scale resolution) are applied for time-depth conversion.

Generally, direct measurements of V_s in seismic CPT are only collected at 10-15% of the CPT locations, so are sparse and insufficient to derive local empirical relations between geotechnical CPT data and V_s that are statistically significant. S-wave velocity can also be measured in the lab via the resonant column or bender elements, but getting high-quality, undisturbed samples can be challenging and expensive, particularly in soft soils, and only provide discrete data points. As a result, empirical relations from other locations are used to estimate V_s or G from geotechnical CPT features. These relations use cone tip resistance in its various forms (e.g. q_c, q_t, q_{net}) to derive V_s (e.g. Duan et al., 2019), as both are controlled by relative density, effective stress, and soil fabric. However, even if the best performing existing empirical relation is selected for the given site, this common solution can be sub-optimal and the error (thereon defined as the difference between predicted and true values) significant. An alternative method for deriving V_s or G and V_p from geotechnical CPT features is therefore attractive, and is explored in this paper using DNN and poroelasticity theory.

2.2. Poroelasticity

Dynamic poroelasticity theory (Gassman, 1951; Biot, 1956a, b) and grain contact theory (e.g. Walton, 1987) are routinely applied in geophysics and reservoir engineering to link the wave velocities V_p and V_s to rock properties such as porosity or pore fluid and hydrate saturation (e.g. Marin-Moreno et al., 2017).

Even though porosity clearly influences both V_p and V_s of shallow seabed soils (Fig. 1), i.e. the top 100 m of seabed, the use of these theories and related models, for shallow seabed applications comprising uncemented soils at low effective stress is rare, and few publications assessing their performance exist (e.g. Dvorkin et al., 1999; Lee, 2010). Three possible reasons are: (i) poor knowledge transfer between the geotechnical and geophysical communities; (ii) limited understanding on how uncertainties in the physical representation of these models and input physical parameters such as coordination number (the average number of grain contacts per grain) and elastic moduli of the solid grains propagate to the geotechnical properties; and (iii) the magnitude of strains and loading frequency of P and S-wave data are substantially smaller and larger respectively than those used for geotechnical design (e.g. Bazle et al., 2006), which adds uncertainty in relating geophysical to geotechnical parameters.

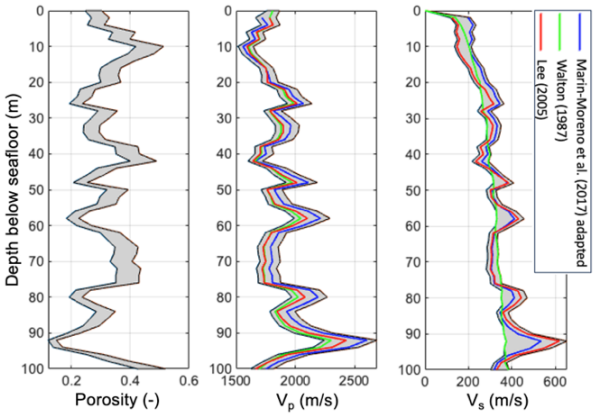

Figure 1. Synthetic results from three poroelastic analytical models assuming a clayey sand (clay content up to 20%) and 10% uncertainty in porosity (grey zone). The general trend of increasing V_p and V_s with depth is caused by the associated increase in effective stress.

Porosity (φ), which can be expressed as a void ratio, $e = \varphi/(1-\varphi)$, is a key parameter for soil type

characterisation and in geotechnical design of offshore wind fixed and floating foundations. Porosity affects the total stress through the bulk density, bulk and shear stiffness and shear strength, drainage and pore pressure dissipation through the permeability, and relative density of coarse-grained soils. Despite its importance, geotechnical design still relies on porosity data at discrete depth intervals and from borehole samples with separations between boreholes of hundreds of metres to kilometres.

Here we present an approach that combines machine learning via deep neural network (DNN) with a dynamic poroelasticity model to correlate measured CPT parameters to V_s , V_p and porosity at centimetre-scale with an uncertainty measure.

3. Method

The proposed workflow consists of three main stages:

1. Training of DNN using geotechnical CPT data and V_s from seismic CPT data, followed by validation and testing of the DNN to predict V_s at centimetric scale from geotechnical CPT data.
2. Uncertainty quantification of the predicted V_s from the DNN compared to true values using conformal prediction;
3. Estimation of porosity (φ) and V_p with uncertainty measure using the predicted V_s and its uncertainty, an empirical correlation between bulk density (ρ) and V_s, and a poroelastic model relating φ and V_s to V_p.

3.1. Training of DNN & prediction of V_s

3.1.1. DNN architecture

We use the TensorFlow ecosystem and the Keras neural network library built-in Python to implement a DNN for a regression problem (TensorFlow, 2023). Our problem involves mapping the selected CPT data input features (depth; corrected tip resistance, q_t; sleeve friction, f_s; soil behaviour type index, I_c, Eq. 2a, Robertson, 2009) to V_s (output) (Fig. 2).

$$I_c = \{[3.47 - \log(Q_t)]^2 + [\log(F_r) + 1.22]^2\}^{0.5} \quad (2a)$$

The normalized cone resistance Q_t and normalized friction ratio F_r in Eq. 2a are given by

$$Q_t = \frac{q_t - \sigma_v}{\sigma'_v} \quad (2b)$$

$$F_r = 100\left(\frac{f_s}{q_t - \sigma_v}\right) \quad (2c)$$

where σ_v and σ'_v are the vertical total and effective stress, respectively.

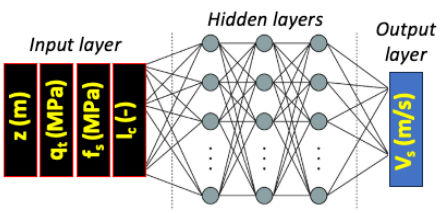

Figure 2. Schematic of the DNN architecture. Nomenclature: depth, z; corrected tip resistance, q_t; sleeve friction, f_s; soil behaviour type index, I_c; shear wave velocity, V_s.

The DNN consists of a pre-processing layer that normalizes the different input features to have a distribution around 0 with standard deviation of 1, three hidden non-linear layers with ReLU activation functions, and a linear single output layer. The hyperparameters (i) units for each of the three hidden layers (cases: 16, 32, 64 or 128 units) and (ii) the learning rate for the Adam optimization algorithm (ranging between 10^{-4} to 10^{-2}) are obtained using the Bayesian optimization algorithm within the KerasTurner (2023) hyperparameter optimization framework. We choose the mean absolute error as a loss function as it is less sensitive to outliers (here referred to CPT features and V_s combinations that lie well outside the region from most combinations) than the mean squared error.

3.1.2. DNN dataset

The CPT and seismic CPT data used to train, validate, and test the DNN was downloaded from two public-domain sources of geodata from real sites located offshore Netherlands, and in Austria and Germany (RVO, 2023; Oberhollenzer et al., 2021). Prior to training, the data were pre-processed to select only the geotechnical CPTs that had near coincident seismic CPTs (V_s data). Having identified the subset of near coincident geotechnical and seismic CPT, the set was further filtered as geotechnical CPT data is acquired every 2 cm and seismic CPT gathers V_s data every 1 m, so geotechnical CPT data closest in depth to V_s data was selected.

The final dataset consisted of 5284 instances (Fig. 3) with 1526 instances from the Dutch sector of the North Sea (RVO, 2023) and 3758 instances from onshore tests in Austria and Germany (Oberhollenzer et al., 2021), covering a wide range of soil types (Fig. 4). The dataset was randomly split into 80% for the training set, 10% for the validation set, and 10% for the test set. The validation set is used to find the optimum hyperparameters and the test set is also used as a calibration set for uncertainty quantification with conformal prediction (Section 3.2).

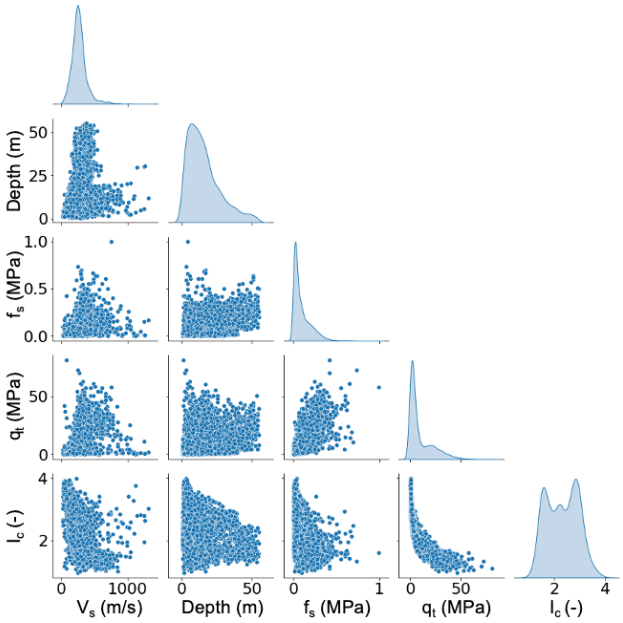

Figure 3. Scatter plots showing the joint distribution of each pair of parameters used in the DNN for the combined dataset of 5284 instances. The main diagonal shows the frequency distribution plots of each parameter.

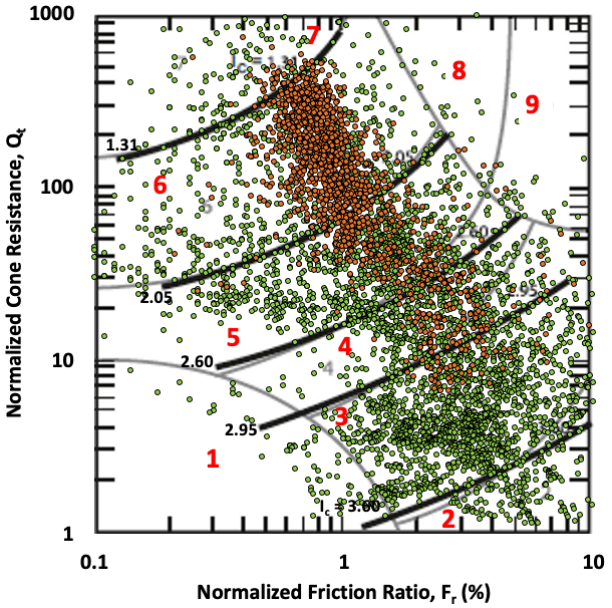

Figure 4. Soil types for the 5284 instances according to Robertson's (1990) soil type classification chart (green dots for Oberhollenzer et al., 2021; brown dots for RVO, 2023). Numbers in red indicate the different soil type regions: 1, sensitive clay and silt; 2, organic soil; 3, clay; 4, silt mixture; 5, sand mixture; 6, sand; 7, gravelly sand; 8, very stiff overconsolidated sand to clayey sand; 9, very stiff overconsolidated clay to silt.

3.2. Uncertainty quantification of V_s

We use conformal prediction to provide distribution-free uncertainty intervals that are guaranteed to contain the true value of V_s with a user-defined probability (Angelopoulus and Bates, 2022; Eq. 3). This uncertainty quantification framework is ideal for black-box machine learning applications, in our case a pre-trained DNN. For example, if a user defines a confidence level of 90% in predicting V_s from CPT features with the proposed DNN, conformance prediction ensures that the derived uncertainty intervals of a new prediction contain the true V_s in at least 90% of the cases, and so produces V_s predictions outside the intervals in 10% or less of the cases. Simplifying, this framework offers a measure of how confident a machine learning model is in a new prediction by controlling the probability of making errors based on past experience.

$$\mathbb{P} = \left(Y_{test} \in C(X_{test})\right) \geq 1 - \delta \tag{3}$$

In Eq. 3, X_{test} and Y_{test} are the test input and unknown test output, respectively, $\delta \in (0,1)$ is the user-defined error rate, $1-\delta$ is the confidence level, and C is the prediction interval which is defined as

$$C(X_{test}) = \{y : s(X_{test}, y) \leq \hat{q}\} \tag{4}$$

In Eq. 4, $s(X_{test}, y) \in \mathbb{R}$ is a user-defined score function that encodes a heuristic notion of uncertainty between the DNN-predicted value and the true value, and $\hat{q}$ is the quantile of scores defined as

$$\hat{q} = \frac{(n+1)(1-\delta)}{n} \tag{5}$$

where n is the number of data instances in a calibration dataset.

We apply what is known as split or inductive conformal prediction, which requires splitting the data into training and calibration datasets. At test time, Y_{test} is unknown and so the score function and quantile level are computed using the calibration dataset not seen during the DNN training. Here, we consider that the calibration dataset and the test set are the same.

We employ Sousa's (2022) implementation of both the naïve conformal prediction, to provide intervals with constant V_s uncertainty with depth, and of the conformalized residual fitting, to provide adaptive intervals with variable V_s uncertainty with depth.

3.3. Estimation of porosity and V_p

We use the predicted V_s and Mayne's (2007) empirical correlation between bulk density (ρ) and V_s to estimate porosity (φ) (Eq. 6). This is the only empirical correlation employed in the proposed method.

$$\rho = 10g\left\{4.17 \ln\left[V_s\left(\frac{p_a}{\sigma'_v}\right)^{0.25}\right] - 4.03\right\} \tag{6a}$$

$$\varphi = \frac{\rho_g - \rho}{\rho_f - \rho} \tag{6b}$$

where g is gravitational acceleration (9.81 m s^{-2}), p_a is atmospheric pressure in the same units as σ'_v, and we assume densities of solid grains (ρ_g) and fluid (ρ_f) of 2650 kg m^{-3} and 1030 kg m^{-3}, respectively.

Looking at Eq. 6a without previous consideration of uncertainty in the input parameters and the form of the correlation, one may be tempted to re-arrange it and estimate V_s from ρ. This should be avoided, as uncertainty in ρ will propagate exponentially to V_s uncertainty, and likely provide unrealistic results. For example, if we assume a V_s of 300 m s^{-1} with a 10% uncertainty and σ'_v of 200 kPa (approximately the σ'_v at 20 m below seafloor under hydrostatic conditions) this results in a ρ of 1900±44 kg m^{-3} (2% uncertainty). In contrast, if we assume a ρ of 1900 kg m^{-3} with a 10% uncertainty at the same effective stress of 200 kPa this results in a V_s of 300± 174 m s^{-1} (58% uncertainty).

To estimate V_p we use a poroelastic model (Lee, 2010) defined by Eqs. 7 and 8. This particular formulation was selected as it only requires one fitting parameter, the consolidation factor (α). The parameter α is related to the degree of consolidation and effective stress and its general behaviour is known, for example it becomes small as the degree of consolidation and effective stress increases. However, its exact behaviour depends on site-specific conditions and so it is effectively a free/fitting parameter (Lee, 2005).

$$K = K_g(1-\beta_p) + {\beta_p}^2\left(\frac{(\beta_p - \varphi)}{K_g} + \frac{\varphi}{K_f}\right)^{-1} \tag{7a}$$

$$G = G_g(1-\beta_s) \tag{7b}$$

In Eq .7, K and G are the bulk and shear moduli of the soil as defined in Eq. 1, K_g and K_f are the bulk modulus of the solid grains and the fluid, respectively, and G_g is the shear modulus of the solid grains. The Biot coefficient in the bulk modulus (β_p), and the coefficient (β_s) in the shear modulus can be defined as

$$\beta_p = \frac{\varphi(1+\alpha)}{(1+\alpha\cdot\varphi)} \quad (8a)$$

$$\beta_s = \frac{\varphi(1+\omega\alpha)}{(1+\omega\alpha\cdot\varphi)} \quad (8b)$$

where

$$\omega = \frac{1+2\alpha}{1+\alpha} \quad (8c)$$

It should be noted that whereas β_p has the real meaning of the Biot's coefficient, which is defined as the ratio of pore-volume change to bulk volume change at constant pressure, β_s is just a convenient way to express G in the same way as K (Lee, 2005).

Using Eq. 7 requires site-specific measurements of the effective moduli of the solid grains. Indeed, these are key parameters in the model that can vary substantially depending on the mineralogy composing the solid grains. For a given site, we can estimate a depth profile of K_g and G_g using the I_c obtained from the CPT data. To do this we propose the following relations:

$$K_g = \begin{bmatrix} K_{sand} = 37\ GPa & \text{if} & I_c < 1.31 \\ K_{mix} = K_{sand} + \left(\frac{K_{clay}-K_{sand}}{2.95-1.31}\right)\cdot \dots & & \\ \dots (I_c - 1.31) \text{ if } 1.31 < I_c < 2.95 & & \\ K_{clay} = 23\ GPa & \text{if} & I_c > 2.95 \end{bmatrix} \quad (9a)$$

$$G_g = \begin{bmatrix} G_{sand} = 44\ GPa & \text{if} & I_c < 1.31 \\ G_{mix} = G_{sand} + \left(\frac{G_{clay}-G_{sand}}{2.95-1.31}\right)\cdot \dots & & \\ \dots (I_c - 1.31) \text{ if } 1.31 < I_c < 2.95 & & \\ G_{clay} = 8\ GPa & \text{if} & I_c > 2.95 \end{bmatrix} \quad (9b)$$

where the bulk and shear moduli for sand correspond to those of quartz, and for clay to those of a mixture of clay minerals (values obtained from Mavko et al., 2009). The I_c values of 1.31 and 2.95 correspond to the limits below and above which the soil is classified as sand and clay, respectively, according to e.g. Robertson's (1990) CPT-based soil classification. The I_c values in between these limits are classified as soil mixtures of sand, silt and clay and above 3.6 are, in fact, classified as organic soil. Here we implicitly assume that an organic soil has the same solid grain moduli than clay, but Eq. 9 allows adding an explicit definition of solid grain moduli for organic soils, if required.

Combining Eq. 1b, Eq. 7b, and Eq. 8b-c and using the predicted V_s and estimated ρ, φ, and G_g , the parameter α can be obtained by solving Eq. 10. Note that only the positive solution of Eq. 10 is physically valid.

$$2\alpha^2 + (1-C)\alpha - C = 0 \quad (10a)$$

where

$$C = \frac{\left(1-\frac{V_S^2\cdot\rho}{G_g}\right)-\varphi}{\varphi\left(\frac{V_S^2\cdot\rho}{G_g}\right)} \quad (10b)$$

To calculate K with Eq. 7a we need K_f. In this case, we assume a constant value for water of 2.3 GPa but a pressure, temperature and salinity depending effective K_f (for a mixture of pore fluids) can also be calculated externally and introduced here (e.g. Marin-Moreno et al., 2017). Finally, V_p can be obtained from Eq. 1a.

4. Results

4.1. Performance of DNN

This section presents the performance of the DNN to predict V_s from geotechnical CPT data. The optimum DNN architecture produces a prediction with a mean absolute error ranging between 59±2 m s^{-1}, depending on the randomly selected training set, which also influences the optimum hyperparameters found. However, because of the size of the entire dataset, the influence of the random splitting of sets is minor as it only produces a change of ±2 m s^{-1}.

Figure 5 shows how the mean absolute error evolves with epoch for an optimized model that produces an error of 57.5 m s^{-1} in the evaluation set, after trying 50 model combinations, and of 55 m s^{-1} in the testing set (Fig. 6).

The optimized DNN model can be considered a *good model* in terms of fitting after 100 epochs, as the error in both the training and validation sets almost plateau (Fig. 5). One of the goals of the validation set is to be able to check for overfitting, i.e. if fitting in the training set improves with increasing number of epochs but it does not in the validation set. Overfitting negatively affects the generalization of the DNN to unseen data, and Fig. 5 clearly shows that it is not happening; this is a positive result. An interesting observation is that the error in the validation and testing sets are slightly lower than in the training set. This observation, although unexpected as the model learns from the training set and we should expect it performs better in this set, is likely because the three sets are drawn from the same underlaying data distribution and the size of the validation and testing sets are significantly smaller than the size of the training set (10% versus 80% of the whole dataset).

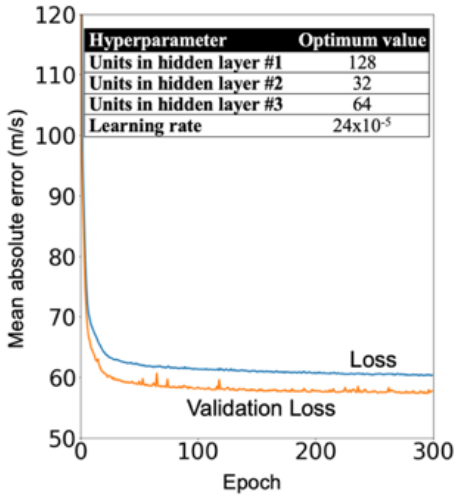

Figure 5. Mean absolute error evolution with epoch associated to the optimal hyperparameter combination shown in the inset table.

Figure 6 shows the correlation of predicted V_s with the true values measured from the seismic CPTs for the test set. Clear clustering about the 1:1 line is evident, albeit with some significant scatter. It is also evident that the model tends to underpredict the V_s for true values above 400 m s^{-1}, which correspond to sands with high q_t and gravelly sands (zones 6 & 7 in the Robertson chart shown in Fig. 4).

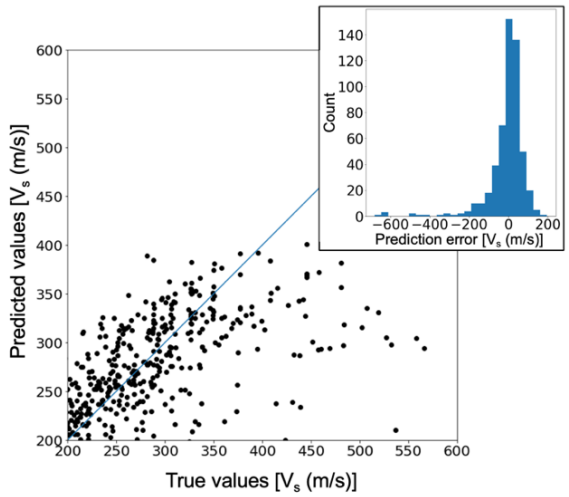

Figure 6. Predictions versus true values of V_s and associated histogram for the testing set. The mean absolute error for the test set is 55 m s^{-1}. No data points are under the inset.

4.2. Uncertainty in V_s, V_p and porosity

We illustrate the predictions of V_s, V_p and porosity for a randomly selected CPT from the Hollandse Kust West Wind Farm Zone, drawn from the RVO (2023) open-source data set. Figure 7 shows that the DNN-based prediction of V_s generally matches well the true V_s values. For this example, a 70% confidence level for the conformal prediction intervals (i.e. at least 70% of data should be contained within the intervals; see section 3.2) provides full coverage of the true V_s values. The mean absolute uncertainty for the constant V_s intervals (naïve conformal prediction) is $\pm$117 m s^{-1} and for the variable intervals (conformalized residual fitting) is $\pm$147 m s^{-1}. For this example, the shape of the I_c profile is very similar qualitatively to the V_s profile, while remarkably different to the q_t, and f_s profiles. Based on this just one example, we should be cautious about extracting a general implication for this observation. However, this aspect is also implicitly shown by the similar shape of the I_c thresholds and the contour shapes of the shear wave velocity factor or shear modulus factor when plotted in Robertson's soil type chart (Robertson, 2009).

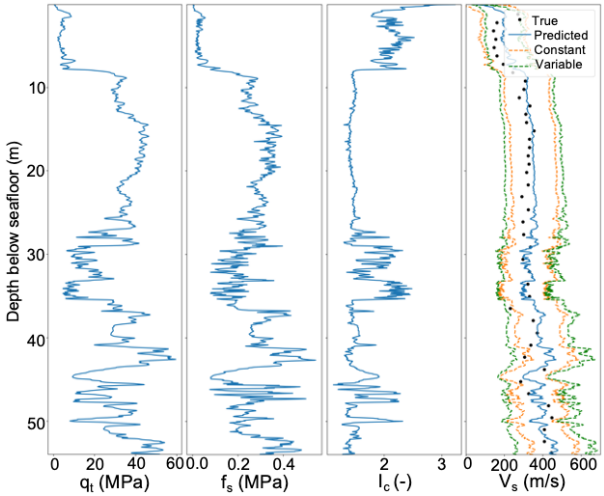

Figure 7. CPT and seismic CPT data (black dots) for test 002 in the Hollandse Kust West Wind Farm Zone and predicted V_s with constant and variable conformal intervals at 70% confidence level.

Figure 8 shows the estimated V_p and porosity and their upper and lower uncertainty bounds. These uncertainty bounds were obtained by introducing the variable upper and lower uncertainty bounds of the predicted V_s (Fig. 7) into step 3 of the workflow (Section 3.3). Uncertainty in V_s predictions drive a mean absolute uncertainty in the predicted V_p and porosity of $\pm$145 m s^{-1} and of $\pm$0.15, respectively.

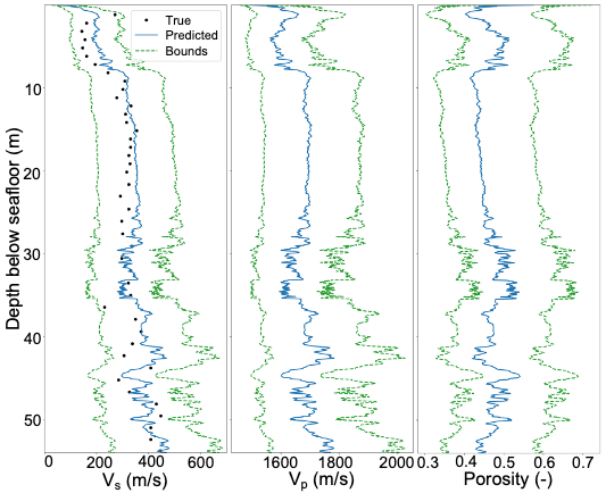

Figure 8. Seismic CPT data (black dots) for test 002 in the Hollandse Kust West Wind Farm Zone and predicted V_s, V_p and porosity with upper and lower uncertainty bounds.

5. Discussion

There is increasing interest in using machine learning algorithms for V_s prediction from CPT data (e.g. Chala and Ray, 2023 and references therein), partly due to the increase of publicly available soil investigation data that enable training and testing of different methods. Some of the studies that evaluate the performance of different machine learning algorithms disregard the actual complexity of the underlying and unknown relation between CPT features and the target V_s. This is because the models are trained and evaluated with V_s labels calculated from an imposed empirical correlation, based on the same CPT data comprising the features, instead of actual V_s data (e.g. Chala and Ray, 2023). Our focus in this work was on assessing the effectiveness of a popular machine learning algorithm such as DNN in predicting real V_s data, i.e. without a pre-imposed behaviour.

Estimating V_s and its uncertainty using DNN and conformal prediction provides an attractive pathway towards high-fidelity characterization of the seabed at centimetre-scale. However, as in all DNN applications, the accuracy of the predictions depends upon the amount and quality of the training dataset. Even though the amount of data is reasonably large and variable, the model tends to underpredict the V_s for soils with true V_s values above 400 m s^{-1}, corresponding to sand with high q_t and gravelly sand. This might be partly caused by poor data coverage for these soils, which also partly explains the wide uncertainty intervals produced with conformal prediction, even considering an arguably low confidence interval of 70% (a higher confidence interval of let us say 95% would provide wider intervals). Hence, to provide higher fidelity predictions of V_s and, most importantly,

generalization to other sites, the dataset should be extended to include a larger representation of sand with high q_t and gravelly sand (Fig. 4, upper region 6 and region 7), and very stiff overconsolidated sand to clayey sand and clay to silt (Fig. 4, regions 8 and 9) which are currently missing.

In addition to data scarcity for specific soil types, other reasons that explain the wide uncertainty intervals are: (i) the complex relationship between the CPTU features (depth, q_t, f_s, and I_c) and target V_s; (ii) the presence of outliers that have not been removed from the data; (iii) CPT features being influenced by sediment type properties over variable depth ranges different to those for V_s. To address (i), we can include other features such as pore pressure. In this work we did not include pore pressure measurements because 43% of the instances used from Oberhollenzer's et al. (2021) dataset lack them, and we wanted the DNN's input layer to have the same number of real data instances per feature. Addressing (ii) would require some a-priori filtering of data outliers by expert or engineering judgement, which in this study we consciously avoided to explore a scenario that minimizes subjectivity. It is known that the measured q_t is actually an average of the tip resistance across the tip influence zone, which is in the order of several cone diameters (decimetres to a few meters) depending on soil type. Similarly, V_s measurements from seismic CPT are an average of the shear-wave properties of soils over a metre distance in depth. Hence, to address (iii), 1 m moving average with depth for q_t, and possibly also for f_s, might be applied during the pre-processing phase to generate the dataset.

Our proposed method provides predictions of V_p and porosity that are causally linked to the CPT data and the predicted V_s through Lee's (2010) theoretical poroelastic model, and so enhance on-going efforts on data-driven only ground models (e.g. Vanneste et al., 2022). Our proposal should be used in combination with the V_p estimated from seismic data and laboratory measurements of porosity. This can help providing a better alignment between the depths of the geophysically-inferred soil units and the CPTu-inferred soil units. Unfortunately, we did not have access to the raw seismic data to evaluate the estimated V_p, and including porosity measurements is work in progress. Indeed, porosity measurements could also be included as an input feature in the DNN.

6. Conclusions

The seabed where most offshore wind turbines are and will be developed is highly heterogenous with distinct lateral and depth variations in stratigraphy, soil types and their geotechnical properties. This spatial variability can only be well-captured by combining metric to decametric resolution geophysical data, CPTu data (centimetric resolution), and data from lab tests in well-selected soil samples. The proposed workflow integrates some of these different datasets and spatial resolutions for V_s, V_p and porosity estimation with uncertainty measure. However, this is work in progress, and the DNN needs refinement including further filtering and/or pre-processing of the data used, adding new site data from other areas to avoid sparse data regions, and testing other input features, such as pore pressure and/or porosity measurements, to capture better the complex relation between V_s and CPT data. This concept can inform requirements of high-resolution geophysical surveys and enable early extraction of synthetic geotechnical parameters for engineering design. The overall ambition of the on-going work is to reduce the amount of geotechnical investigation required for a given site, and enable flexibility of windfarm layout post survey, in both cases reducing costs, increasing reliability and accelerating offshore wind deployment.

Acknowledgements

We thank RVO (2023) and Oberhollenzer et al. (2021) for making the soil investigation data publicly available. Susan Gourvenec and Jared Charles are supported by the Royal Academy of Engineering through the Chairs in Emerging Technologies Scheme.

References

Angelopoulos, A. N., and Bates, S. 2022. "A gentle introduction to conformal prediction and distribution-free uncertainty quantification", arXiv, http://doi.org/10.48550/arXiv.2107.07511

Biot, M. 1956a. "Theory of propagation of elastic waves in a fluid-saturated porous solid. I. Low-frequency range", J. Acoust. Soc. Am., vol. 28, no. 2, pp. 168-178

Biot, M. 1956b. "Theory of propagation of elastic waves in a fluid-saturated porous solid. II. Higher frequency range", J. Acoust. Soc. Am., vol. 28, no. 2, pp. 179-191

Catapult. 2019. "Guide to an offshore wind farm". Published on behalf of The Crown Estate and the Offshore Renewable Energy Catapult. Available at: https://guidetoanoffshorewindfarm.com/, accessed: 15/09/2023

Chala, A. T., and Ray, R. P. 2023. "Machine learning techniques for soil characterization using cone penetration test data", Appl. Sci., vol. 13, no. 14, https://doi.org/10.3390/app13148286

DESNZ. 2023. Department for Energy Security & Net Zero, "2022 UK greenhouse gas emissions, provisional figures". Available at: https://www.gov.uk/government/statistics/provisional-uk-greenhouse-gas-emissions-national-statistics-2022, accessed: 16/09/2023

Duan, W., Cai G., Liu, S., and Puppala, A. J. 2019. "Correlations between shear wave velocity and geotechnical parameters for Jiangsu clays of China", Pure Appl. Geophys., http://doi.org/10.1007/s00024-018-2011-x

Dvorkin, J., Prasad, M., Sakai., A., and Lavoie, D. 1999. "Elasticity of marine sediments: Rock physics modelling", Geophys. Res. Lett., vol. 26, no. 12, pp. 1781-1784 http://doi.org/10.1029/1999GL900332

EU-COM. 2020. European Commission, COM(2020) 741 final, Brussels 19/11/2020, An EU Strategy to harness the potential of offshore renewable energy for a climate neutral future

Gassmann, F. (1951). "Uber die elastizitat poroser medien" (About elasticity in porous media), Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich, vol. 96, pp. 1-23.

Greaves, D., Jin, S., Wong, P., White, D., Jeffrey, H., Scott, B., and Wigg, R. 2022. "UK perspective research landscape for offshore renewable energy and its role in delivering Net Zero", Prog. Energy, vol. 4, no. 4, http://doi.org/10.1088/2516-1083/ac8c19

GWEC. 2021. “Global Offshore wind report 2021”. Available at: https://gwec.net/global-offshore-wind-report-2021/, accessed: 15/09/2023

IRENA. 2021. “World Energy transitions outlook: 1.5◦ C pathway”. Available at: https://www.irena.org/Publications/2021/Jun/World-Energy-Transitions-Outlook, accessed: 17/09/2023

KerasTurner. 2023. “https://keras.io/keras_tuner/”, accessed: 25/08/2023

Lee, M. W. 2005. “Proposed moduli of dry rock and their application to predicting elastic velocities of sandstones”, USGS, Reston, Virginia, Scientific Investigations Reports 2005-5119, http://doi.org/10.3133/sir20055119

Lee, M. W. 2010. “Predicting S-wave velocities for unconsolidated sediments at low effective pressure”, USGS, Reston, Virginia, Scientific Investigations Reports 2010-5138, http://doi.org/10.3133/sir20105138

Marín-Moreno, H., Sahoo, S. K., and Best, A. I. 2017. “Theoretical modeling insights into elastic wave attenuation mechanisms in marine sediments with pore-filling methane hydrate”, J. Geophys. Res., vol. 12, no. 3, pp. 1835-1847, http://doi.org/10.1002/2016JB01357

Mavko, G., Mukerji, T., and Dvorkin, J. 2009. “The Rock Physics Handbook”, 2nd Edition, Cambridge University Press New York, USA, http://doi.org/10.1017/CBO9780511626753

Mayne, P. W. 2007. “In-situ test calibrations for evaluating soil parameters”, Characterization & Engineering Properties of Natural Soils, vol. 3, Taylor & Francis Group, London, Overview paper on in-situ testing, Singapore workshop, Nov 29-Dec. 01, 2006, pp. 1602-1652.

McKinsey. 2022. “How to succeed in the expanding global offshore wind market”. Available at: https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/how-to-succeed-in-the-expanding-global-offshore-wind-market, accessed: 20/09/2023

Oberhollenzer, S., Premstaller, M., Marte, R., Tschuchnigg, F., Eharter, G. H., and Marcher, T. 2021. “Cone penetration test dataset Premstaller Geotechnik”, Data in Brief, vol. 34, 106618. http://doi.org/10.1016/j.dib.2020.106618

Orsted. 2023. “https://orsted.com/en/insights/the-fact-file/what-is-the-carbon-footprint-of-offshore-wind”, accessed: 09/09/2023

Provenzano, G., Vardy, M., Henstock, T. J. 2017. “Pre-stack full waveform inversion of ultra-high-frequency marine seismic reflection data”, Geophys. J. Int., vol. 209, no. 3, pp. 1593-1611, https://doi.org/10.1093/gji/ggx114

Putuhena, H., White, D., Gourvenec, S., and Sturt, F. 2023. “Finding space for offshore wind to support net zero: A methodology to assess spatial constraints and future scenarios, illustrated by a UK case study”. Renewable Sustainable Energy Rev., vol. 182, 113358. http://doi.org/10.1016/j.rser.2023.113358

Robertson, P. K. 1990. “Soil classification using the cone penetration test”, Can. Geotech. J., vol. 27, no. 1, pp. 151-158, http://doi.org/10.11397t90-014

Robertson, P. K. 2009. “Interpretation of cone penetration tests – a unified approach”, Can. Geotech. J., vol. 46, no. 11, pp. 1337-1355, http://doi.org/10.11937T09-065

RVO. 2023. “http://offshorewind.rvo.nl”, accessed: 07-11/2023

Sousa, M. 2022. “Inductive conformal prediction: a straightforward introduction with examples in python”, arXiv, https://doi.org/10.48550/arXiv.2206.11810

TensorFlow. 2023. “https://www.tensorflow.org/tutorials/keras”, accessed: 25/08/2023

Vanneste, M., Sauvin, G., Dujardin, J.-R., Forsberg, C. F., Klinkvort, R. T., Forsberg, C. S., Hansen, R. C. 2022. “Data-driven ground models: The road to fully-integrated site characterization and design”, In: Proceedings of the 2nd Vietnam Symposium on Advances in Offshore Engineering – Lecture Notes in Civil Engineering, Ho Chi Minh City, Vietnam, pp. 3-21, http://doi.org/10.1007/978-981-16-7735-9_1

Vardy, M., Vanneste, M., Henstock, T. J., Clare, M. A., Forsberg, C. F., Provenzano, G. 2017. “State-of-the-art remote characterization of shallow marine sediments: the road to a fully integrated solution”, Near Surf. Geophys., vol. 14, no. 4, pp. 387-402, http://doi.org/10.3997/1873-0604.2017024

Walton, K. 1987. “The effective elastic moduli of a random packing of spheres”, J. Mech. Phys. Solids, vol. 35, no. 2, pp. 213-226, http://doi.org/10.1016/0022-5096(87)90036-6.