Soil variability from high-resolution S-wave full-waveform inversion: Deriving reliable cone-tip resistance from V_s for geotechnical evaluations

Eddy Revelo-Obando[1], *Ranajit* Ghose[1], and *Michael* Hicks[1]

[1]*Department of Geoscience and Engineering, Delft University of Technology, Delft, The Netherlands*
E.D.ReveloObando@tudelft.nl, r.ghose@tudelft.nl, M.A.Hicks@tudelft.nl

ABSTRACT

Capturing the spatial variability in soil is crucial for ground response analyses in the context of seismic hazard mitigation. The lateral variability in thickness and properties of the different soil layers is one of the main factors that determines the variability of the ground motion spectrum from one location to another. The absence of such lateral variability information in the subsoil in between the locations of Cone Penetration Tests (CPTs) may be compensated by the use of more densely sampled seismic data. In this research we aim to derive a shear-wave velocity field through seismic full-waveform inversion that yields a model resolution approaching that of high-resolution seismic CPT surveys. Following this, a data-driven correlation between geophysical and geotechnical information is attempted through the application of new machine-learning-based approaches.

Keywords: Full-waveform inversion; Machine learning; Shear-wave velocity.

1. Introduction

Integrating the knowledge from geophysical and geotechnical investigations is beneficial for a detailed understanding of the subsurface. Seismic methods have been widely used for earthquake seismic site response analysis. Multi-channel analysis of surface waves (MASW) is a popular method that has been used to obtain the shear-wave velocity (V_s) distribution in soil layers (Park, Xia and Miller 1999; Foti, Picozzi and Albarelle 2011). V_s is a crucial parameter in site response analysis, together with soil properties that capture the nonlinear soil behavior under seismic loading, as obtained from dynamic tests in the laboratory. MASW essentially offers a one-dimensional V_s structure derived from the surface-wave dispersion information. Besides being essentially one-dimensional, the reliability of the MASW-derived V_s rapidly decreases with depth.

In recent years, seismic full-waveform inversion (FWI) has gained the attention of the near-surface geophysical community as a powerful tool for imaging soil variability. FWI using shear waves is especially attractive for geotechnical engineering applications. However, successful use of FWI using shear waves is generally challenging due to the high cost of forward modelling of shear waves propagating in soft, low-velocity soil layers. The very small wavelengths of shear waves in low-velocity soil layers do contribute to very high resolution (Ghose 2003; Ghose and Goudswaard 2004), though modelling such wave propagation calls for a very fine spatial discretization, and therefore significant computational costs. Furthermore, factors such as the lack of a good starting model, cycle skipping due to the absence of very low frequencies in data, and attenuation of the high frequencies in soft soil, all make near-surface FWI a challenging task. To overcome these challenges, significant efforts have been made in recent years to improve FWI schemes.

Additionally, for dynamic site response analyses, one needs – in addition to soil stratigraphy – several in-situ soil properties, which are typically obtained from cone penetration tests (CPTs). Cone-tip resistance (q_c) is used to obtain the undrained shear strength of saturated cohesive soils and the friction angle of sands. CPTs provide very detailed soil variability information in the vertical direction, which is crucial in designing foundations, assessing the risk of soil liquefaction, and understanding the bearing capacity of the soil, among others (Kruiver et al. 2021). Many past studies have shown that V_s and q_c exhibit correlation with each other in the near-surface soils. Using the available SCPT database for the Groningen region of the Netherlands, we have also found such a correlation, as illustrated in Fig. 1. High q_c generally corresponds to high V_s values.

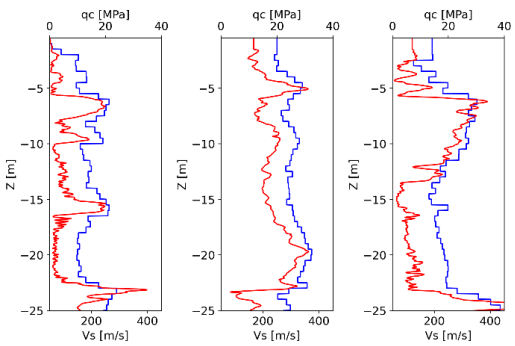

Figure 1. SCPTs from Groningen, the Netherlands (https://www.dinoloket.nl/). The blue lines denote the V_s values, and the red lines are the q_c values at the same location.

The goal of the present study is to design FWI workflows in order to retrieve 2D and 3D high-resolution V_s fields, overcoming the above-mentioned challenges, and to subsequently derive site-specific correlations between V_s and q_c through novel approaches utilizing machine learning (ML).

2. FWI of shallow pseudo-observed seismic data

FWI is a nonlinear inversion approach used in computational geophysics for subsurface imaging and elastic parameter estimation. FWI addresses an ill-posed problem, causing the inversion to fall into a local minimum in the solution space, which may differ significantly from the global minimum (Virieux and Operto 2009). Substantial research efforts have been directed toward mitigating this problem using new misfit functions and advanced optimization algorithms. Metivier et al. (2016) proposed optimal transport (OT) distance as the misfit function. The advantage of OT over traditional misfits, such as the Euclidean norm, is the convexity of the optimal transport distance with respect to the time shift between two oscillatory signals. It is known that the phase differences between the calculated and observed seismograms are one of the causes of the cycle-skipping problem. Performing the optimization process using OT generally yields a solution which is closer to the global minimum, though at extra computational cost.

We use the spectral element method to generate accurate synthetic seismograms employing the advanced python toolbox Salvus (Afanasiev et al. 2019). The model representing the shallow subsoil (Fig. 2) is derived from multiple high-quality SCPTs (depth sampling 25 centimeter) that were performed earlier up to 30 meter depth at a test site (Ghose, 2007). This model is used in viscoelastic forward modeling to generate pseudo-seismic data representing the field data. For the source time function, we use the first derivative of a Ricker wavelet with a dominant frequency of 40Hz. The frequency band in the data is 0-80 Hz, which is realistic for shear waves at soft-soil sites. We add random noise to the seismic data. For the initial V_S model for FWI, we use a highly smooth version of an actual (field-measured) SCPT-derived V_S profile from a central location at the test site. We assume that this single SCPT is available a-priori; our goal is to extract subsoil lateral variability away from this SCPT location. Initial models for density and compressional wave velocity (V_p) needed for elastic wave propagation simulation are obtained from the initial V_S model using suitable empirical relations.

We perform FWI in a multistep manner. We tune the FWI workflow to capture both lateral and vertical V_s variabilities. As an input hyperparameter, the OT misfit requires the maximum expected time shift between the modeled seismograms and the observed data. By reducing this time shift as the inversion progresses, we increase the resolution and accuracy of the inverted model. In Fig. 2 the positions of the actual SCPTs (separated by 25 meters from each other) are shown. Fig. 3 shows the initial 1-D velocity model obtained by smoothing SCPT #3 (Fig. 2). Figs. 4, 5 and 6 show the inversion results in 2-D, the difference ΔV_s between the true and inverted model, and the result in 1-D, respectively.

The inverted V_s model (Fig. 6) shows that the V_s variability in the high-density SCPT data is resolved remarkably well by the specially tuned FWI; mainly, the low-velocity zones in the upper 20 meters are very well captured; these layers are absent in the initial velocity model. These low-velocity peat and clay layers would significantly influence the ground shaking due to an earthquake. At greater depths, the resolution of the inverted model decreases partly due to high seismic attenuation. Nevertheless, the main geological features at these depths are mostly retrieved by FWI. Because V_s is three times higher than that at the shallower depths, the wavelengths are relatively large at greater depths, which adds to the loss of resolution.

Because the V_S in soft soils can be very low and highly heterogeneous, fitting seismic waveforms from a poor starting model is generally challenging due to cycle-skipping. Our results show that, with some extra computational time and proper tuning, the OT cost function in FWI is powerful in solving the cycle-skipping problem at the scale of near-surface seismic imaging.

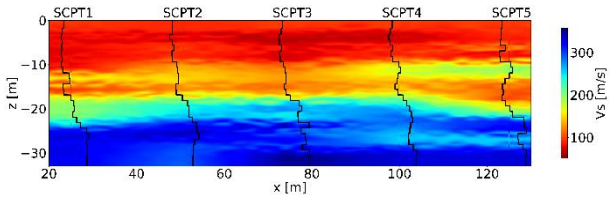

Figure 2. The "true" V_s model obtained by interpolating between 5 actual SCPTs. This model is used to derive the pseudo-observed seismic data. The horizontal separation between the SCPTs is 25 meters.

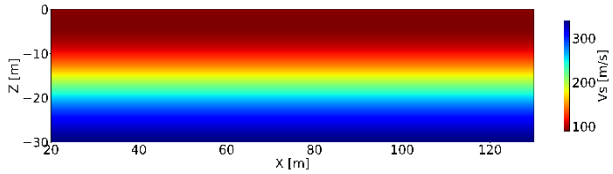

Figure 3. Initial 1-D V_S model derived by heavily smoothing V_S profile from SCPT #3. This is the input model for FWI.

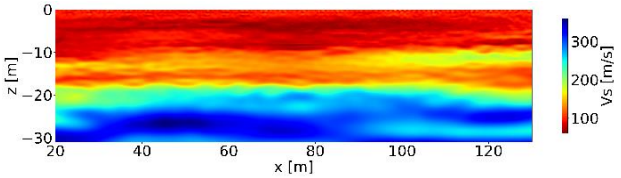

Figure 4. Inverted 2-D velocity model by FWI.

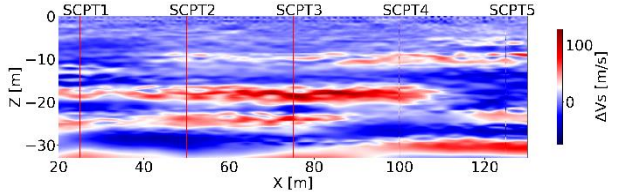

Figure 5. Difference between the true and inverted model.

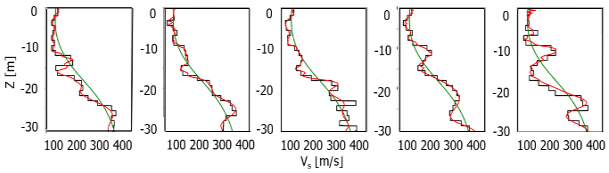

Figure 6. Inverted 1-D V_s profiles to check the accuracy and resolution of FWI to capture the lateral variability in the soil: the black lines denote the true V_s values at the locations of the SCPTs, the green lines are the initial V_s values (same at all locations), and the red lines are the inverted V_s values, using the pseudo-observed seismic data.

3. Deriving q_c from V_s

In the past, empirical correlations between V_s and q_c were found in numerous field studies. Such correlations also considered the effect of factors such as soil density, soil type, overburden pressure, and effective stress. These earlier correlations were mostly derived through regression analysis. More recently, pattern recognition and machine learning techniques have been used in determining such correlation.

Having explored the potential of elastic FWI to capture reliably the fine-scale V_s variability in the near surface soil layers as discussed above, in the next step we make use of the SCPT database from Groningen, in order to investigate the feasibility of predicting q_c from V_s through the special adaptation of machine learning (ML) approaches. The utilized database consists of 45 irregularly distributed SCPTs in Groningen. Using this information, we create a new database containing high-quality V_s profiles, spatial coordinates of the SCPTs, and relevant geological information. Geological information is obtained from TNO's 3-D GeoTOP model for the Dutch subsurface. The information is organized in vectors, each one containing the name of the SCPT, V_s, X and Y coordinates, the maximum depth to which measurements were conducted, and the geology expressed as a Boolean vector to indicate at each depth which lithology is more likely; e.g., clay, sandy-clay, gravel, etc. For each depth level, we define the feature vector $\boldsymbol{v}_f$ as:

$$\boldsymbol{v}_f = [V_s, X, Y, h, L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8]^T. \quad (1)$$

In Equation (1), V_s is the shear-wave velocity at a specific depth level, X and Y are the spatial coordinates of the SCPT, h is the actual depth level, T denotes transpose, and the L features correspond to the most likely lithology at that depth level. These features are expressed as one-hot vectors. This means that from the 8 possible values at each depth level, one element takes on the value of 1, and the others are set to 0. This applies for all the SCPTs at every depth level defined in the data.

This new dataset serves as the training dataset for ML. The target of the prediction is q_c. A common pre-processing step before splitting the data for training, validation, and ML tests is to normalize the values so as to have the input vectors on a standard scale and to facilitate the training of a chosen ML algorithm. Moreover, we exclude some SCPTs that are judged as outliers. Fig. 7 illustrates the locations of the SCPTs in the province of Groningen. The average distance between the SCPTs exceeds 200 m.

We adapt three different ML techniques: Support Vector Regressor or SVR (Boser, Guyon, and Vapnik 1992), Random Forest Regressor or RFR (Breiman 2001), Extreme Gradient Boosting or XG Boost (Chen and Guestrin 2016). These algorithms are tested using the Python toolbox Scikit Learn. GroupKfold cross-validation was used to avoid overfitting in the training process. A total of 5 Kfolds were used during the validation stage, and 3 SCPTs were set aside for testing. We conduct a random search to determine the best hyperparameters leading to each technique's highest prediction accuracy. For this, we select a sufficiently large range of possible hyperparameters for each technique. Then we train the data and select the hyperparameters that yield the best possible predictions.

Figs. 8, 9, and 10 show the results of the q_c predictions at locations that are not used in the training. The label on top of each figure denotes the SCPT used for the prediction. The locations of the SCPTs used for training and prediction are shown in Fig 11.

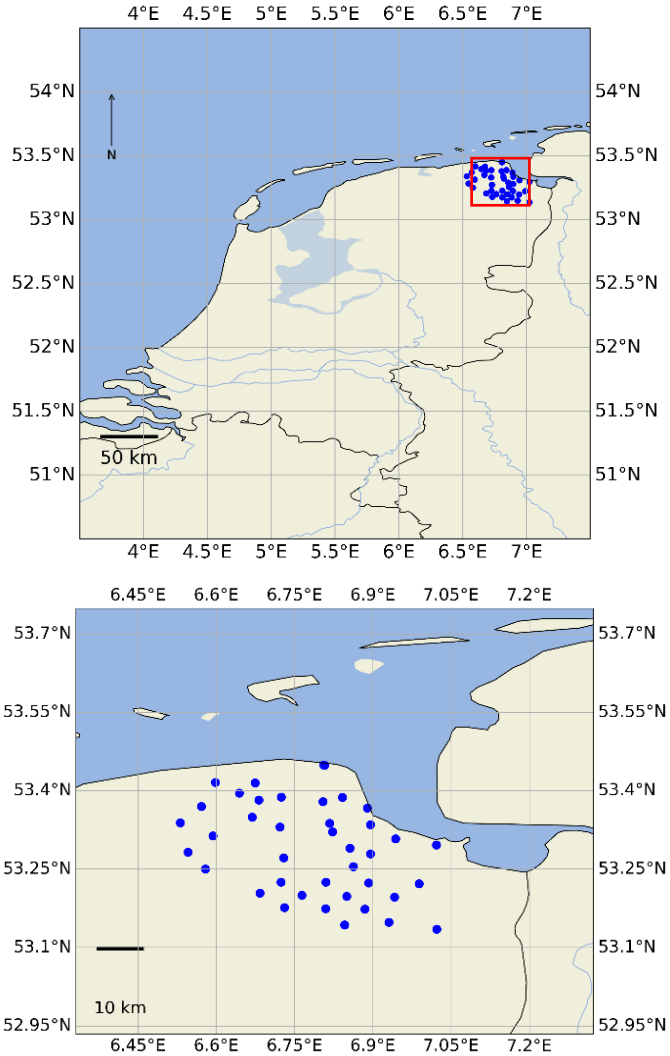

Figure 7. SCPT database from Groningen. Top: locations of SCPTs in the northern tip of the Netherlands. Bottom: zoomed-in area in Groningen, marked in the top figure.

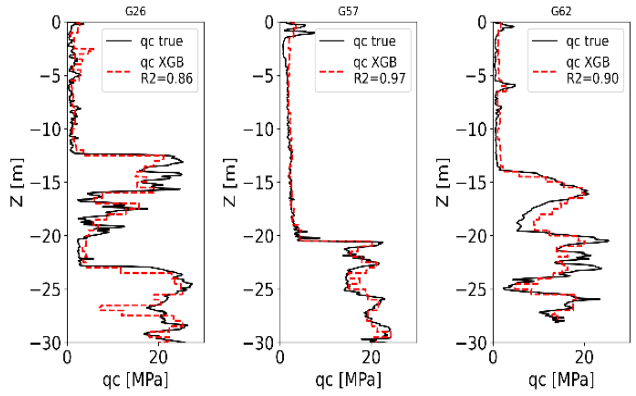

Figure 8. True and predicted q_c using XG Boost algorithm.

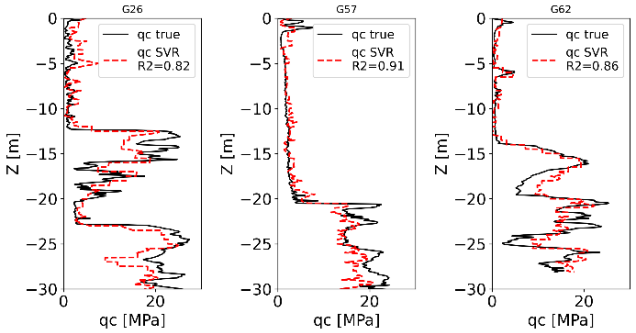

Figure 9. True and predicted q_c using the RFR algorithm.

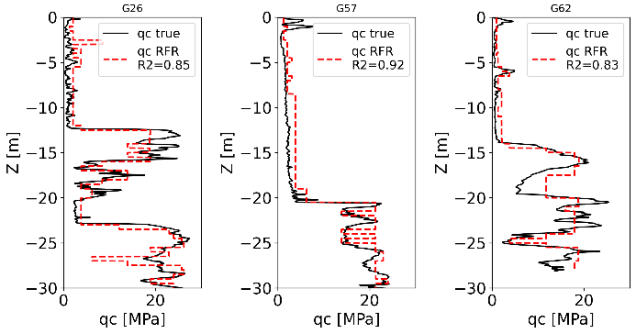

Figure 10. True and predicted q_c using the SVR algorithm.

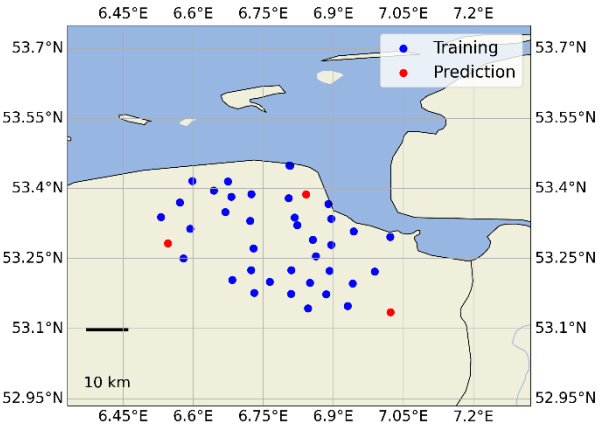

Figure 11. SCPT locations for training and prediction in the northern Netherlands. Blue dots - SCPTs used for training. Red dots - SCPTs used for prediction (Figs. 7-9).

At the locations (red dots) shown in Fig. 10, we have successfully predicted q_c with reasonable accuracy. The coefficient R^2 is used as a measure of accuracy. We find that XG Boost offers the best results among the three tested ML methods. XG Boost requires less computation time than SVR. It is important to note that, in spite of the considerable distance between the SCPTs, the very large size of the region in which the SCPTs are located and the ML algorithms are trained, our adaptation of the ML techniques offers a good prediction of q_c. If such a prediction is performed in a site-specific manner, we anticipate the accuracy to be higher.

4. Conclusions

In this research, we have investigated the possibility of adapting advanced FWI and ML approaches to capture lateral variability in the near-surface soil layers and in soil properties that affect local seismic site response. Shear-wave seismic data can be inverted to derive a V_s field with a resolution that approaches that of SCPTs. This high resolution and accuracy are necessary to translate, in the next step, V_s to q_c. Special adaptations of FWI have addressed the computational and technical challenges. Moreover, appropriate ML techniques can be designed in order to achieve reasonable q_c predictions from the V_s. In the next phase of this research, we will test the developed methodologies on field-seismic and CPT databases from Groningen, where induced seismicity is a major concern.

Acknowledgments

The authors are grateful for the financial support provided by the Dutch Research Council (NWO) under the DeepNL programme (project 3DSOIL).

References

Afanasiev, M., Boehm, C., van Driel, M., Krischer L., Rietmann, Max., May, D. A., Knepley, M. G., and Fichtner, A. "Modular and flexible spectral-element waveform modelling in two and three dimensions." Geophysical Journal International, 216(3), pp. 1675–1692, 2019. https://doi.org/10.1093/gji/ggy469

Boser, E., Guyon, I. M., and Vapnik, V. N. "A training algorithm for optimal margin classifiers." In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 1992, pp.144–152. https://doi.org/10.1145/130385.130401

Breiman, L. Random Forests. Machine Learning, 45, pp. 5–32, 2001. https://doi.org/10.1023/A:1010933404324

Chen, T., and Guestrin, C. "XGBoost: A Scalable Tree Boosting System." In Proceedings of the 22nd ACM SIGKDD International Conference of Knowledge Discovery and Data Mining, pp. 785–794, 2016. https://doi.org/10.1145/2939672.2939785

Foti, S., Picozzi, M., and Albarelle, D. "Application of surface-wave methods for seismic site characterization", Surveys in Geophysics, 32(6), pp. 777–825, 2011. https://doi.org/10.1007/s10712-011-9134-2

Ghose, R. "High-frequency shear-wave reflections to monitor lateral variations in soil, supplementing downhole geotechnical tests." (Re)claiming the Underground Space, Saveur (Ed.), Swets & Zeitlinger, pp. 827-833, 2003.

Ghose, R., and Goudswaard, J. "Integrating S-wave seismic-reflection data and cone-penetration-test data to derive laterally varying geotechnical information." Geophysics, 69(2), pp. 440-459, 2004. http://dx.doi.org/10.1190/1.1707064

Ghose, R. "Integrating shear-wave seismic and cone penetration testing to derive laterally varying geotechnical information." In Project Report DAR.5761, p. 85, Dutch Technology Foundation (STW), 2007.

Kruiver, P.P., de Lange, G., Kloosterman, F., Korff, M., van Elk, J., and Doornhof, D. "Rigorous test of the performance of shear-wave velocity correlations derived from CPT soundings: A case study for Groningen, the Netherlands." Soil Dynamics and Earthquake Engineering, 140, pp. 106471, 2021. https://doi.org/10.1016/j.soildyn.2020.106471

Métivier, L., Brossier, R., Mérigot Q., Oudet, E., and Virieux, J. "Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion." Geophysical Journal International, 205(1), pp. 345–377, 2016. https://doi.org/10.1093/gji/ggw014

Park, C., Xia, J., and Miller, R. “Multichannel analysis of surface waves (MASW)”, Geophysics, 64(3), 1999. https://doi.org/10.1190/1.1444590

TNO Geological Survey of The Netherlands. “DINOloket: Data and information on the Dutch Subsurface”, [online] Available at: [https://www.dinoloket.nl/], accessed: 13/02/2024.

Virieux, J., and Operto, S. "An overview of full-waveform inversion in exploration geophysics." Geophysics, 74(6), pp. WCC1-WCC26, 2009. https://doi.org/10.1190/1.3238367