Cone Penetration Testing 2022 – Gottardi & Tonni (eds)

A data-driven approach to predict shear wave velocity from CPTu measurements

I. Entezari & J. Sharp
ConeTec Group, Burnaby, British Columbia, Canada

P.W. Mayne
Georgia Institute of Technology, Atlanta, Georgia, USA

ABSTRACT: The use of machine learning modelling to predict shear wave velocity (V_S) from piezocone penetration tests (CPTu) is presented. A large dataset of paired V_S-CPTu data (n = 104,054) compiled from seismic piezocone (SCPTu) soundings completed in a wide variety of soil types with various stress histories and geological environments was used to develop machine learning models to directly estimate V_S from CPTu data. The impact of soil microstructure on the results was investigated and separate models were developed to predict V_S in cemented and uncemented soils. The results of machine learning models outperformed the existing widely used CPT-based relationships to predict V_S.

1 INTRODUCTION

Shear wave velocity (V_S) is an important property of geomaterials and is widely used to evaluate the dynamic and elastic properties of soils in geotechnical design. V_S measurements provide the fundamental stiffness of the ground in terms of the small-strain shear modulus (G_o), specifically: $G_o = \rho V_S^2$, where $\rho = \gamma_t/g_a$ = soil total mass density, γ_t = soil total unit weight, and g_a = gravitational acceleration constant.

V_S measurements can be obtained using a variety of test methods. The value of V_S can be measured in the laboratory using high quality undisturbed samples and special equipment (resonant column, bender elements), which is costly and restricted to a limited number of samples. In-situ measurements of V_S are preferable to preserve site-specific conditions and minimize errors due to sampling disturbance and stress release.

In-situ measurements of V_S can be obtained through downhole and crosshole tests, seismic piezocone tests (SCPTu), spectral analysis of surface waves (SASW), and multichannel analysis of surface waves (MASW). SCPTu method is often preferred as it is a rapid and cost-effective technique to measure in-situ wave velocities in conjunction with CPTu parameters, including cone tip resistance (q_t), sleeve friction (f_s), and dynamic porewater pressure (u_2) in a single direct push sounding.

Although performing site-specific testing is the preferred method to determine shear wave velocity, several empirical relationships have been developed to estimate V_S from the basic CPTu for lower risk projects. When actual measurements of V_S are not practical, estimates can still provide useful additional information. Existing empirical CPT relationships for V_S have been developed using statistical approaches. This paper explores the use of a data-driven approach via machine learning modelling to predict V_S from CPTu. Machine learning requires little or no priori assumptions to be considered and thus are more flexible than statistical models.

The development dataset used in this study is comprised of V_S-CPTu data pairs from ConeTec SCPTu soundings collected from 2017 to early 2021. The soundings have been completed in a wide variety of soil types with various stress histories and are from geological environments around the world. The dataset is tested with a random forest algorithm to develop a model for the prediction of V_S from CPTu data. The results of the machine learning models are compared to empirical equations proposed by Mayne (2006) and Robertson (2009). Furthermore, the impacts of soil microstructure and cementation on estimated V_S results are discussed and separate models are developed for the categories of uncemented and cemented soils.

2 BACKGROUND

2.1 *Seismic Piezocone Tests (SCPTu)*

The SCPTu is similar to the CPTu probe with the addition of one or more geophones or accelerometers located behind the cone tip. As shown in Figure 1, the equipment required to perform SCPTu includes the

DOI: 10.1201/9781003308829-51

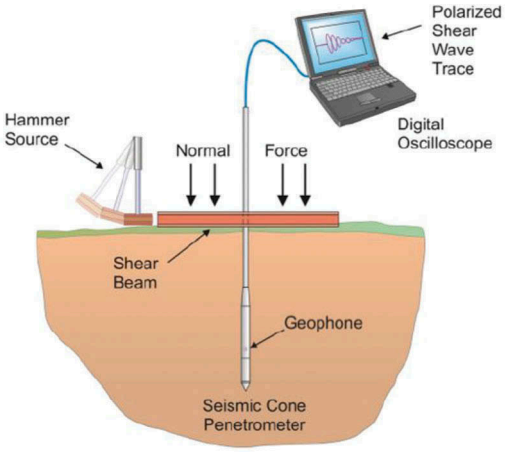

Figure 1. Schematic of seismic CPTu equipment.

seismic source on the ground surface, seismic sensors behind the cone probe, a data-acquisition system, and a data recording trigger circuit (Styler et al. 2016).

The seismic testing is conducted at selected depth intervals (typically every 1m), while the penetrometer is pushed into the ground. The shear waves are generated by striking a horizontal beam pressed firmly against the ground. Paired leftand right-strikes are used to define either the first arrival time of shear waves, or first crossover, or both. V_S is calculated using the difference in arrival times of the shear wave traces between the source and geophone at two successive depths. The SCPTu has several advantages including correlated V_S and CPTu results, the capability to do deep tests, the ability to measure compression wave velocity (V_P), and availability of soil property interpretations based on small-strain rigidity index ($I_G = G_o/q_{net}$, where q_{net} is cone net tip resistance).

2.2 *Existing CPT relationships*

Relationships between CPT data and V_S have been studied previously by various researchers. Hegazy & Mayne (1995) developed various expressions to estimate V_S using parameters including tip resistance (q_c), sleeve friction (f_s), vertical effective stress (σ'_{vo}), and in-situ void ratio (e) for Quaternary clays, sands, and mixed soils. Various relationships have also been proposed by Piratheepan (2002) for the estimation of V_S based on tip resistance (q_c), sleeve friction (f_s), vertical effective stress (σ'_{vo}), depth (z), and soil behaviour type index (I_c) for Holocene clays, sands, and other soils. Mayne (2006) showed a relationship where the V_S is a function of the sleeve friction (f_s) for Quaternary soils. Another correlation developed by Andrus et al. (2007) for Holocene and Pleistocene soils is based on tip resistance (q_t), depth (z), soil behaviour type index (I_c), and a time factor depending on the soil age. Robertson (2009) also developed a generalized soil relationship where V_S is a function of net tip resistance ($q_{net} = q_t - \sigma_{vo}$), total vertical stress (σ_{vo}), atmospheric pressure (σ_{atm}), and soil behaviour type index (I_c). An overview on some of the CPT relationships to predict V_S has been provided in Wair et al. (2012).

In this study, the estimated V_S results from the machine learning models are compared to the results obtained by the empirical expressions proposed by Mayne (2006) and Robertson (2009), shown in Equations 1 and 2, respectively:

$$V_s = 118.8 \log(f_s) + 18.5 \quad (1)$$

$$V_s = \left[\left(10^{0.55 I_c + 1.68}\right)(q_t - \sigma_{vo})/\sigma_{atm}\right]^{0.5} \quad (2)$$

where V_S is in m/s in both equations, f_s is in kPa in Eq. 1, and σ_{atm} is in same units as q_t and σ_{vo} in Eq. 2.

2.3 *Impact of soil microstructure*

The existing empirical correlations developed for interpretation of CPT results have been generally developed using silica-based uncemented soils with little or no microstructure (Robertson 2016). Therefore, caution should be exercised when CPT based relationships are used in soils with microstructure. According to Robertson (2016), the empirical parameter, K^*_G, can be used to determine whether soils are cemented or not. K^*_G is calculated as $(G_o/q_{net})(Q_{tn})^{0.75}$ (Robertson 2016), where Q_{tn} is the normalized tip resistance. Soils with K^*_G of less than 330 are likely young and uncemented with little or no microstructure, while soils with K^*_G of greater than 330 can be classified as cemented and microstructured soils.

The cemented versus uncemented soils are considered in this study for the evaluation of the performance of the machine learning model. Furthermore, individual models are developed specifically for uncemented and cemented soils.

3 DESCRIPTION OF DATASET

To investigate the potential of a data-driven approach to estimate V_S from CPTu data, a dataset of paired V_S-CPTu data was compiled using ConeTec's geospatial database. The database was queried to find the relevant information that resulted in 14,855 SCPTu tests worldwide with more than 248,500 V_S-CPTu data pairs. For this study, soundings collected after 2016 were selected in order to only utilize modern tests with increased quality control. Procedural changes in ConeTec's SCPTu methodology yielded slightly higher accuracy in V_S measurements due to signal enhancement and signal

stacking after this date (Styler & Weemees 2016). Consequently, the dataset was reduced to 104,809 V_S-CPTu data pairs from 7171 independent SCPTu soundings worldwide. Most of the soundings are from North America, with additional contributions from various sites in South America, Australia, Europe, and Asia. To pair the CPTu parameters with V_S measurements at a given depth, the median of CPTu parameters over a window size equal to the V_S depth interval was calculated. Only depth intervals equal to or less than 1 m were considered to minimize variations due to potential soil heterogeneity. The CPTu parameters paired with V_S included corrected tip resistance (q_t), sleeve friction (f_s), porewater pressure (u_2) and depth (z) at each V_S measurement. Additional parameters including normalized tip resistance (Q_{tn}), normalized friction ratio (F_r), normalized porewater pressure (B_q), net tip resistance (q_{net}), total stress (σ_{vo}), and effective stress (σ'_{vo}) as well as small-strain shear modulus (G_o), small-strain rigidity index (I_G), and K^*_G were also calculated.

Calculation of a number of these parameters required the soil unit weight (γ). The machine learning model based on corrected tip resistance (q_t), sleeve friction (f_s), porewater pressure (u_2) and depth (z) developed by Entezari et al. (2021) was used to estimate unit weight at each depth. The measured equilibrium pore pressure profile of each SCPTu sounding, combined with the estimated unit weight profile, was used to determine in-situ vertical stresses.

Data points with net tip resistance (q_{net}) of less than 100 kPa were screened out to remove fluid-like tailings from the dataset. Also, data points with sleeve friction (f_s) of less than 1 kPa were screened out in order to remove data where the soil-sleeve friction was less than internal o-ring friction. The final dataset used included 104,054 V_S-CPTu data pairs. Table 1 lists the summary statistics of the paired dataset.

Table 1. Summary statistics of the V_S-CPTu dataset.

	Min	Max	Mean
V_S (m/s)	9	1000	251
q_t (MPa)	0.1	94.1	8.4
f_s (kPa)	1.0	1577	117.6
u_2 (kPa)	-87.2	5489	245.0
z (m)	0.3	129.6	17.3
σ'_{vo} (kPa)	0.1	2185	215.4

Total number of data pairs = 104,054.

3.1 *Soils with microstructure*

The plot of normalized tip resistance (Q_{tn}) versus small-strain rigidity index (I_G) for the dataset is shown in Figure 2. Accordingly, 63,740 data points fall in the young and uncemented soils category (soils with little or no microstructure), where K^*_G is less than 330 (green points in Figure 2). Another 40,314 data points fall in the cemented soils category (soils with microstructure), where K^*_G is greater than 330 (blue points in Figure 2).

3.2 *Training and test datasets*

The dataset was split into training and test sets. The training set was used to calibrate the model whereas the test set was used to evaluate the model performance. The data collected from 2017 to 2019 was used as the training set and data collected in 2020 and early 2021 provided the test set. This allows for an unbiased performance evaluation of the model (a blind test) where the potential errors due to variation in stress histories and geological environments are taken into account. The number of paired V_S-CPTu data points for the training and test sets are listed in Table 2.

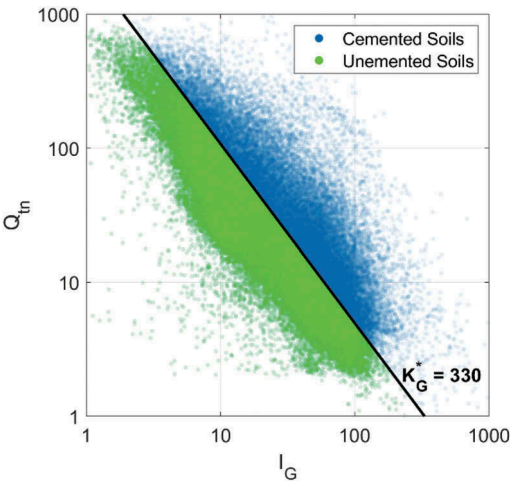

Figure 2. The dataset plotted in the Q_{tn}-I_G chart.

Table 2. Number of data pairs in the training and test sets.

	All	Uncemented	Cemented
Training set	73,010	45,386	27,624
Test set	31,044	18,354	12,690

4 MACHINE LEARNING MODELLING

Machine learning models acquire information from prior data, allowing the computers to discover predictive rules applicable for future data. Machine learning models are generally data-hungry and need large

datasets for training. In general, as more data become available, the more accurate and robust the predictions become. Machine learning is widely used in numerous disciplines and has gained interest in geotechnical engineering. Example applications of machine learning for CPT interpretations can be found in Erzin & Ecemis (2016), Reale et al. (2018), Wang et al. (2019), Erharter et al. (2021), Rauter & Tschuchnigg (2021), and Entezari et al. (2020, 2021).

In this study, the random forest algorithm (Breiman 2001) was employed to calibrate CPTu data to V_S measurements. It is one of the most widely used machine learning algorithms for classification and regression tasks. Random forest is an ensemble of several decision trees and thus overcomes the shortcomings of traditional decision trees, predominantly overfitting. The models here were trained using four input parameters including corrected tip resistance (q_t), dynamic porewater pressure (u_2), sleeve friction (f_s), and depth (z).

4.1 *Performance evaluation*

The performance of the random forest models is evaluated using the properties of the cumulative distribution function (CDF) of errors on the test set. The error is calculated as the discrepancy between the measured V_S from SCPTu and predicted V_S from the random forest models. The 50th percentile in the CDF is taken as the bias of the prediction. Assuming the errors follow a normal distribution, the CDF values at 15.9% and 84.1% correspond to ±1 standard deviation. The average of the two CDF values at 15.9% and 84.1% is considered as the overall error of the model.

To compare the performance of the machine learning models to existing CPTu expressions, similar performance evaluation is performed on the test set using the predicted V_S obtained from the equations proposed by Mayne (2006) and Robertson (2009).

5 RESULTS

5.1 *All-soils model*

An all-soil model was developed using the random forest model trained with all data points in the training set. The relationship between measured V_S from SCPTu and the estimated V_S from the random forest model is shown in Figure 3. This relationship is shown for the test set. The R^2 of the model on the test set was observed to be 0.58. The error analysis using CDF of errors showed that the bias and error of the estimated results are –8.5 and 49.5 m/s, respectively. The bias –8.5 m/s means that random forest model overestimates the measured V_S by 8.5 m/s overall. The error of 49.47 m/s means that 68.2% of the estimated V_S values fall within ±49.5 m/s of the measured V_S from SCPTu testing.

The performance of the model was also assessed on uncemented and cemented soils. When only uncemented soils were considered in the test set, the bias and error of the estimated V_S results are –23.4 and ±34.8 m/s, respectively. For cemented soils, the bias and error of the estimated V_S were observed to be 27.2 and ±62.8 m/s, respectively.

5.2 *Uncemented and cemented soils models*

Using the uncemented and cemented soil categories in the training set, two separate models were developed for the estimation of V_S in these types of soils. Figure 4 shows the relationship between the SCPTu measured and random forest predicted V_S for the fraction of the test set in uncemented soils. As can be seen, the correlation between estimated and measured V_S significantly improved compared to the all-soils model shown in Figure 3 (R^2 of 0.79 compared to 0.58). The bias and error of the estimated results were observed to be 0.6 and 28.2 m/s, respectively.

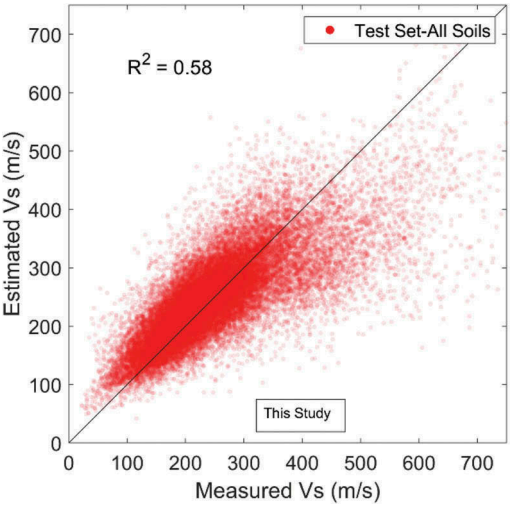

Figure 3. Relationship between measured and random forest estimated V_S on the test set using all soils.

Evidently, the random forest model is far better able to model the relationship between V_S and CPTu parameters in uncemented soils, compared to the all-soil scenario.

The results of the random forest model developed for cemented soils are also shown in Figure 4. The bias and error were observed to be -12.3 and 54.1 m/s, respectively. Compared to the all-soil model, this model performs better on cemented soils, but the bias and error are still high. This is presumably because microstructure can have a variety of impacts on CPTu parameters. Thus, the learnt relationship between V_S and CPTu parameters in

cemented soils of the training set may not be applicable on the cemented soils of the test set.

5.3 *Existing relationships*

Figures 5 shows the relationships between the estimated V_S calculated using the methods of Mayne (2006) and Robertson (2009) with the measured V_S using SCPTu on the test set. For the Mayne (2006) model, the bias and error were observed to be 12 and 68.6 m/s, respectively, when error assessment was done on all soils. When only uncemented soils were considered, the bias and error were dropped to -7.2 and 52.5 m/s, respectively. The bias and error were calculated to be 51.1 and 82.8 m/s, respectively, on cemented soils.

In case of Robertson (2009) model, the bias and error were 21.5 and 64.3 m/s, respectively, on all soils in the test set. The bias and error were observed to be -6.1 and 50.3 m/s, respectively, for uncemented soils, compared to 69.2 and 57.8 m/s for cemented soils. Overall, it can be seen that these expressions perform better on uncemented soils, as expected. A summary of model performances is presented in Table 3. It should be noted that no limits were applied to the two existing methods because the intent was to compare the results to those obtained from the random forest models developed using a wide range of soil types. Limiting the range of applicable data to be used in the existing methods would be prudent and may result in a better average correlation and error.

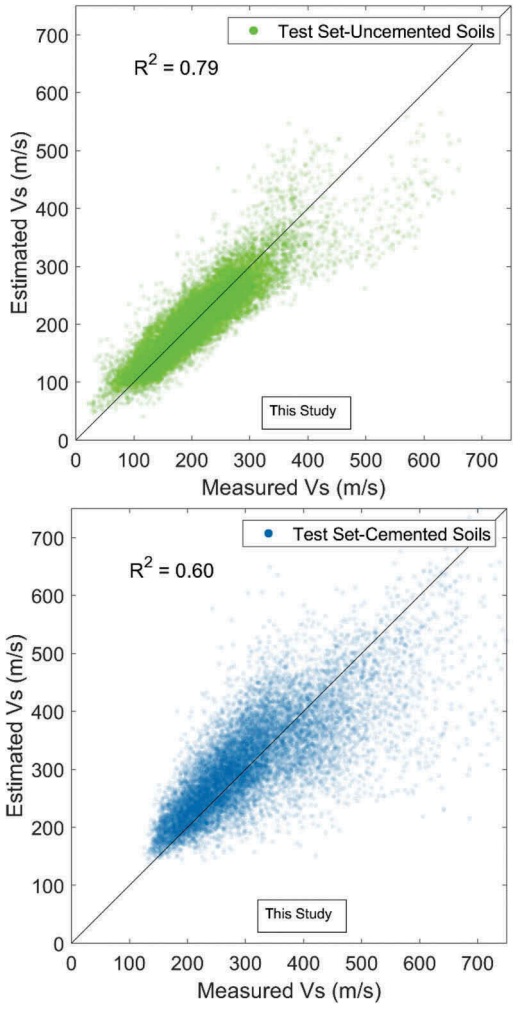

Figure 4. Relationships between measured and random forest estimated V_S of the uncemented (top) and cemented soils (bottom) in the test set when separate models were trained using uncemented and cemented soils in the training set.

Table 3. Performance of different models.

	Bias±Error (m/s)		
Model	All Soils	Uncemented	Cemented
RF-All Soils	-8.5±49.5	-23.4±34.8	27.2±62.8
RF-Uncemented	NA	0.6±28.2	NA
RF-Cemented	NA	NA	-12.3±54.1
Mayne (2006)	12.0 ±68.6	-7.2±52.5	51.1±82.8
Robertson (2009)	21.5 ±64.3	-6.1±50.3	69.2±57.8

5.4 *Example SCPTu Vs profile*

An example SCPTu profile of V_S estimated using the random forest models developed in this study is shown in Figure 6. The estimated V_S values from the expressions of Mayne (2006) and Robertson (2009), as well as the measured V_S profile, are displayed along with the results of this study. The SCPTu sounding is from 2020 and is thus part of the test set. As evident, both all-soils and uncemented models are in agreement with the measured V_S. The uncemented model, however, shows less fluctuations and a better performance compared to the all-soils model. Both of these models appear to outperform the Mayne (2006) and Robertson (2009) models. The analysis of K^*_G revealed that the soils are uncemented for the whole profile except for depth ranges between 3.5-9 m and 15-17.5 m.

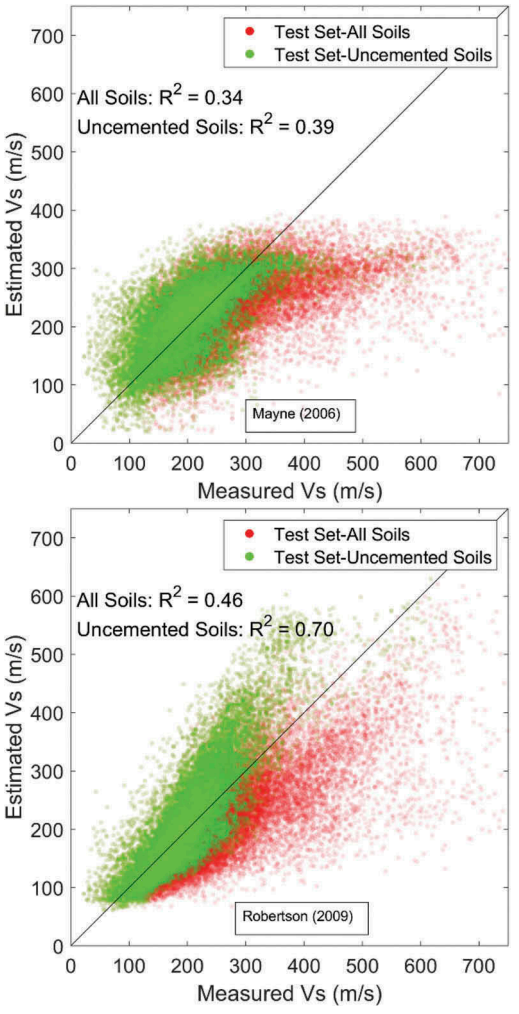

Figure 5. Relationships between measured vs. estimated V_S using Mayne (2006) (top) and Robertson (2009) (bottom) methods.

6 DISCUSSION

In practice, a priori information on the soil microstructure is required in order to be able to employ soil-specific models developed in this study to estimate V_S (uncemented and cemented soils models). Determining the soils categories based on K^*_G is not practical without knowing V_S. Therefore, information on soil categories should be available from other sources such as previous SCPTu testing in the region under investigation or information on the geology of soils.

When such information is not available, the results of this study showed that the developed all-soils model performs better than the Mayne (2006) and Robertson (2009) models when CPTu is pushed in regions with both cemented and uncemented soils.

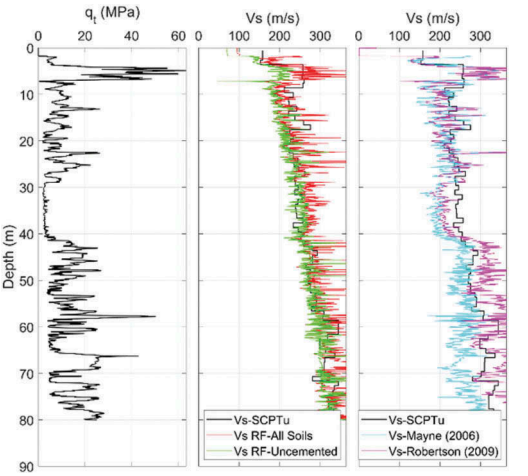

Figure 6. Example profile of Vs estimated from models developed in this study and existing relationships.

When a priori information on soil categories is available, the soil-specific models developed in this study could lead to better results than the all-soils model and the equations proposed by Mayne (2006) and Robertson (2009).

In future work, the dataset compiled in this study will be used to investigate the potential of machine learning algorithms to classify and identify cemented and uncemented soils from CPTu parameters. However, K^*_G of 330 as a threshold to distinguish cemented from uncemented soils has been determined empirically and may not be an absolute metric.

In addition to the models developed and presented in this paper, random forest models were trained by adding normalized tip resistance (Q_{tn}), normalized friction ratio (F_r), normalized porewater pressure (B_q), and effective stress (σ'_{vo}) to the input variables, but no significant improvements were observed.

7 CONCLUSIONS

Machine learning models using a random forest algorithm were developed to directly predict V_S from CPTu data. A dataset of paired V_S-CPTu data compiled from 7171 SCPTu soundings completed at various sites with a wide variety of soil types, stress histories, and geological environments was used to develop machine learning models. Results showed that the all-soils model developed using random forest algorithm can estimate V_S with ±49.5 m/s error. The model developed for uncemented soils showed a significant improvement and could predict V_S with ±28.2 m/s error. The model developed for cemented soils achieved an accuracy of ±54.1 m/s. All the developed machine learning models outperformed the studied existing relationships from literature. Although actual measurement of V_S is always

preferrable, it appears to be more crucial when dealing with soils that have microstructure. The models developed are from a very large dataset compiled from SCPTu soundings from various geological regions and are therefore considered to be robust, however engineering judgement should always be exercised when using any empirical statistics or models.

REFERENCES

Andrus, R.D., Mohanan, N.P., Piratheepan, P., Ellis, B.S., and Holzer, T.L. 2007. Predicting shear wave velocity from cone penetration resistance, *Proc. 4th Intl. Conf. on Earthquake Geotech. Engrg*., Thessaloniki, Greece.

Breiman, L. 2001. Random forests. *Machine Learning*, 45(1): 5–32.

Entezari, I., McGowan, D., and Sharp, J. 2020. Tailings characterization using cone penetration testing and machine learning, *Proc. Tailings & Mine Wastes 2020*, University of British Columbia, Vancouver, 695–704.

Entezari, I., Sharp, J., and Mayne, P.W. 2021. Soil unit weight estimation using the cone penetration test and machine learning, *Proc. GeoNiagara 2021*, Niagara Falls, Canada.

Erharter, G.H., Oberhollenzer, S., Fankhauser, A., Marte, R., and Marcher, T. 2021. Learning decision boundaries for cone penetration test classification, *Computer-Aided. Civil & Infrastructure Eng*. 1: 1–15.

Erzin, Y. & Ecemis, N. 2016. The use of neural networks for the prediction of cone penetration resistance of silty sands. *Neural Comput. Appl*. 28: 727–736.

Hegazy, Y.A. & Mayne P.W. 1995. Statistical correlations between VS and cone penetration data for different soil types. *Proc. CPT '95*, Linkoping, Sweden, Vol. 2: 173–178.

Mayne, P.W. 2006. In-situ test calibrations for evaluating soil parameters. *Proc. Characterization and Engineering Properties of Natural Soils II*, Singapore, Vol. 3: 1601–1652.

Piratheepan, P. 2002. Estimating shear wave velocity from SPT and CPT data. *MSc Thesis*, Clemson University.

Rauter, S. & Tschuchnigg, F. 2021. CPT data interpretation employing different machine learning techniques. *Geosci. J*. 11(7), 265.

Reale, C., Gavin, K., Librić, L., Jurić-Kaćunić, D. 2018. Automatic classification of fine-grained soils using CPT measurements and Artificial Neural Networks. *Adv. Eng. Inform*. 36: 207–215.

Robertson, P.K. 2009. Interpretation of cone penetration tests – a unified approach, *Can. Geotech. J*. 46 (11):1337–1355.

Robertson, P.K. 2016. Cone penetration test (CPT)-based soil behaviour type (SBT) classification system-an update. *Can. Geotech. J*. 53: 1910–1927.

Styler, M.A. & Weemees, I. 2016. Quantifying and reducing uncertainty in down-hole shear wave velocities using signal stacking. *Proc. ISC'5, Gold Coast, Australia*.

Styler M.A., Weemees, I., Mayne, P.W. 2016. Experience and observations from 35 years of seismic cone penetration testing (SCPTu), *Proc. GeoVancouver 2016*: www.cgs.ca

Wair, B.R., DeJong, J.T., and Shantz, T. 2012. *Guidelines for Estimation of Shear Wave Velocity.* PEER Rept. 2012/08, Pacific Earthquake Engineering Research Center, Berkeley, CA: 95 p.

Wang, H., Wang, X., Wellmann, J.F., Liang, R.Y. 2019. A Bayesian unsupervised learning approach for identifying soil stratification using cone penetration data. *Can. Geotech. J*. 56: 1184–1205.