Civil Engineering Infrastructures Journal xx,xx
DOI: 10.22059/CEIJ.2024.366170.1969

Evaluation of Uncertainty in Shear-Wave Velocity Based on CPT Records Using the Robust Optimization Method

Hossein MolaAbasi[1], Mohammad Mahdi Hajitaheriha[2*], Maziar Salahi[3], Jiliang Li[4], Omolbanin Ataee

[1] Assistant Professor, Department of Civil Engineering, Gonbad University, Gonbad, Golestan, Iran, hma@gonbad.ac.ir

[2*]PhD. Candidate, Department of Civil Engineering, Memorial University of Newfoundland, A1B 3X5, St. John's, Newfoundland, NL, Canada, mhajitaherih@mun.ca

[3]Professor, Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Guilan, Rasht, Iran, salahim@guilan.ac.ir

[4]Assistant Professor, Department of Civil Engineering and Construction Management, California Baptist University, California, U.S, jli@calbaptist.edu

[5] Assistant Professor, Department of Geography and Urban Planning, University of Mazandaran, Babolsar, Mazandaran, Iran. o.ataee@umz.ac.ir

Received: 03/10/2023
Revised: 16/01/2024
Accepted: 12/02/2024

Abstract

Shear-wave velocity (V_s) is used to evaluate the soil shear modulus and classify the soil type in pseudo-static analysis. Empirical correlations are developed to relate V_s and cone penetration test (*CPT*) records. However, uncertainty in the input parameter measurements is always a major concern. Therefore, the current research employs a novel method based on robust optimization to study the effect of such uncertainties. To measure the merits of the suggested method, 407 records were collected and categorized for several soil types. The identification procedure employed in this investigation is based on the robust model of least squares, solved using the interior point technique for second-order cone problems. The uncertainty definition is examined against correlation coefficients for empirical models, and optimum values are determined based on the Frobenius norm of the data points. A diagram for calculating the shear wave velocity considering uncertainties was also presented. This study suggests that the robust method is the best pattern recognition tool for uncertain datasets compared to previous statistical models. Other power models also have good accuracy compared to the polynomial model, but when uncertainty is taken into account, the accuracy of the other models is lower compared to the polynomial model.

Keywords: *CPT*; Polynomial model; Robust optimization; Shear-wave velocity; Uncertainty.

1. Introduction

The shear wave velocity (V_s) induced shear modulus is a main geotechnical property corresponding to small strains which is of importance in geotechnical research. Due to the constraints in gathering undisturbed samples, particularly in granular soils, in situ seismic tests, in place of laboratory measurements are the best possible direct tests in achieving the V_s. To establish the V_s profile, the surface wave velocity assessment, as well as down-hole and cross-hole techniques can be conducted (Eslami et al., 2020; Mayne, 2007; Robertson, 2009). However, because of the limits of the noise level and space constraints in urban areas, seismic in situ investigations are not usually possible or appropriate. Therefore, it is convenient to estimate V_s indirectly by other common in situ tests, such as the standard penetration test (*SPT*) for compacted soils and the cone penetration test (*CPT*) for soft soils. Among in situ tests, *CPT* is a more versatile and reliable test that is used in geotechnical site investigations (Anagnostopoulos et al., 2003). Zhang et al. (2021) proposed a multilayer fully connected network (*ML-FCN*) to optimize the training of the deep neural network (*DNN*) using the V_s and

SPT datasets. Zhao et al. (2021) developed a new *PSO-KELM* hybrid machine learning model to evaluate soil liquefaction potential and explore nonlinear relationships *between cyclic resistance ratio (CRR), CPT, and* V_s measurements. Wang et al. (2022) assessed thirteen alluvium sites in Taipei Basin, and measurements of shear wave velocity were concurrently obtained using the five common seismic methods. Measurement discrepancies were quantified through statistical analysis of the data, offering guidance for method selection. Yang et al. (2023) integrated *CPT-* V_s data to create a simplified probabilistic assessment for liquefaction potential.

Chala & Ray (2023) employed machine learning (*ML*) algorithms to predict V_s from cone penetration test (*CPT*) data, including Random Forests (*RFs*), Support Vector Machine (*SVM*), Decision Trees (*DT*), and eXtreme Gradient Boosting (*XGBoost*). Zhou et al. (2022) presented two SVM models optimized with genetic algorithm (*GA*) and grey wolf optimizer (*GWO*) to predict soil liquefaction potential, validated using CPT, SPT, and VS test data with varying input variables. Several correlations and mathematical methods were suggested for estimating the V_s based on *CPT* records for loose sand, silt, clay, and all other soil types (Comina et al., 2022; Jakka et al., 2022; Meng & Pei, 2023; Zhao et al., 2022). Wang et al. (2022) investigated 13 alluvial sites in Taipei Basin and measured the velocity profiles employing each of the five most common seismic methods. Using test data and statistical models, differences in seismic methods were quantified as calibrated measurement uncertainties, which can be used as a reference for selecting an appropriate method to measure shear wave velocity. Zhai, et al. (2024) aimed to develop a Bayesian framework that considered both in-situ test data (*SPT, CPT*) and prior information, to determine the probabilistic characteristics of V_s while accounting for transformation uncertainty. The study found out that the model which includes two in-situ tests accurately predicts shear wave velocity.

Using different travel times (i.e., first arrival picks, peaks and troughs picks, crossover picks, and the peak response of the cross-correlation function) and different velocity analysis methods (i.e., pseudo-interval, true-interval, corrected vertical travel time slope-based, and raytracing), Stolte, et al. (2020) developed a number of V_s profiles. Gilder et al. (2021) presented *CPTu* data related to the Kathmandu valley sediments and employed the established *CPTu* interpretation procedures to assess the in-situ soil properties. Previously, for the assessment of variability and seismic response of the subsoils in Kathmandu, *SPT* data and limited shear wave velocity measurements were predominantly used. This study provided further data to supplement the existing SAFER/GEO-591 database, new shear wave velocity measurements, and initial estimates of *CRR* at the visited sites. It was concluded that liquefaction assessment mainly due to the presence of saturated silts in the valley demands a more detailed methodology.

Wang et al. (2022) examined the performance of shear wave velocity-standard penetration test (V_s -*SPT*) and shear wave velocity-piezocone penetration test (V_s -*CPTU*) models and found that V_S has a strong correlation with the depth of soil (*D*) but weak correlation with *SPT-N* or *CPT*-q_c. This indicats that most of the transformation models in previous studies are not suitable to these sites as they disregard soil depth in their formulation. In an attempt to confirm if the assumption about the consistency of depth is true for V_s data using models created in the *CSR* framework, Wang et al. (2022) created two models to assess the chances of liquefaction. These models use the V_s -based probabilistic approach and consider the uncertainty of measurements. By assuming consistency in depth, it was found that the performance of the suggested models is similar to two commonly used models based on the *CSR* framework, and better than the Chinese code model.

Mohammadikish et al. (2023) utilized two different approaches, namely, the whole data strategy and partial distance strategy, in their study. Bayat et al. (2023) introduced an analytical approach to optimize the compaction pattern and dynamic compaction variables regarding regular constraints. They employed a metaheuristic approach (Genetic Algorithm) to find global optimum. Results indicated that the maximum allowed values of tamper mass and the number of tamper drops were required to minimize compaction energy.

In this study, researchers examined the effectiveness of fuzzy c-means clustering in analyzing incomplete data to evaluate the probabilistic liquefaction. The data used for this analysis included cone penetration test (*CPT*) and shear wave velocity (V_s) field data. This method compared the traditional deterministic and probabilistic liquefaction evaluation approaches, and it was found the fuzzy c-means clustering model demonstrated a similar predictive capability compared to other methods. Thus, it is considered reliable for evaluating the liquefaction possibility. The main variables as input parameters are cone resistance (q_c), sleeve friction (f_s), and overburden pressure in effective form (σ'_{V0}) in the correlation functions, such as linear, logarithm, power, and polynomial functions. Some of the important correlations are presented in Table 1.

Table 1. A list of the proposed correlations between V_s and *CPT*

Functional Form	Proposed Correlation (m/s)	Eq.	Author(s)	Soil Type	Units	
					q_c	f_s
$V_S = a_1 + a_2 q_c$	V_S =154+0.64q_C	*1*	Barrow, 1983	All	kgf/cm^2	-
	V_S=134+0.52q_C	*2*	Sykora, 1983	Sand	kgf/cm^2	-
	V_S =160+0.9q_C	*3*	Iyisan & Ansal, 1993	All	kgf/cm^2	-
	V_S =218+0.70q_C	*4*	Iyisan & Ansal, 1993	Sand	kgf/cm^2	-
$V_S = a_1 + a_2 \ln(q_c)$	V_S = 109.29+52.674ln(q_C)	*5-1*	Tun, 2003	All	MPa	MPa
	V_S =109.29+52.674ln(q_C)	*5-2*	Tun & Ayday, 2018	All	MPa	MPa
$V_S = a_1 + a_2 \log(f_s)$	V_S =18.5+118.8log(f_S)	*6*	Mayne, 2006	All	-	kPa
$V_S = a_1 (q_c)^{a_2}$	V_S =54.8$(q_C)^{0.29}$	*7-1*	Sykora, 1983	Sand	kgf/cm^2	-
	V_S =45$(q_C)^{0.41}$	*7-2*	Iyisan & Ansal, 1993	All	kgf/cm^2	-
	V_S =1.75$(q_C)^{0.627}$	*7-3*	Mayne & Rix, 1995	Clay	kPa	kPa
	V_S =55.3$(q_C)^{0.377}$	*7-4*	Iyisan & Ansal, 1993	Clay	kgf/cm^2	-
	V_S =211$(q_C)^{0.23}$	*7-5*	Madiai & Simoni, 2004	All	MPa	MPa
$V_S = a_1 (q_c)^{a_2} {f_s}^{a_2}$	V_S =12.02$(q_C)^{0.319}(f_S)^{-0.0466}$	*8-1*	Hegazy & Mayne, 1995	Sand	kPa	kPa
	V_S =155$(q_C)^{0.29}(f_S)^{-0.10}$	*8-2*	Madiai & Simoni, 2004	All	MPa	MPa
$V_S = a_1 (q_c/Pa)^{a_2} + a_3$	V_S =50[$(q_C/p_a)^{0.43}$ - 3]	*9*	Paoletti et al., 2010	Sand	kPa	-
$V_S = a_1 (q_c)^{a_2} (f_s)^{a_2} (\sigma_v)^{a_3}$	V_S =359$(q_C)^{0.119}(f_S)^{0.1}(\sigma_v')^{0.204}$	*10*	Kruiver et al., 2021	All	MPa	MPa
$V_S = a_1 (q_c)^{a_2} (f_s)^{a_2} (Z)^{a_3}$	$V_S = 18.4(q_c)^{0.144}(f_s)^{0.0832}(Z)^{0.278}$	*11*	McGann et al., 2015b	All	kPa	kPa
$V_S = a_1 + a_2 q_c + a_3 f_s + a_4 {q_c}^2 + a_5 {f_s}^2 + a_6 (q_c f_s)$	V_S (m/s) = 100[1.36 - 0.35f_s + 0.15q_c - 0.05f_s^2 - 0.018q_c^2 + 0.39$(f_s)(q_c)$]	*12-1*	Mola-Abasi et al., 2015	Clay	MPa	MPa
	V_S (m/s) = 100[1.73 + 2.74f_s + 0.03q_c - 4.015f_s^2 - 0.00026q_c^2 + 0.007$(f_s)(q_c)$]	*12-2*		Sand	MPa	MPa
	V_S (m/s) = 100[1.47 + 2.07f_s + 0.10q_c + 9.50f_s^2 - 0.0023q_c^2 - 0.034$(f_s)(q_c)$]	*12-3*		Mixed	MPa	MPa
	V_S (m/s) = 100[1.40 + 1.59f_s + 0.09q_c - 1.33f_s^2 - 0.002q_c^2 + 0.05$(f_s)(q_c)$]	*12-4*		All	MPa	MPa

Pa = atmospheric pressure (100 kPa)

In Table 1. a_i's are constant coefficients of the model and each researcher has estimated the coefficients using the data related to the study case. Such empirical correlations (Table 1) are generally based on statistical regression analyses with notable modeling drawbacks. For example, inaccuracies enter into the field measurements of V_s in case histories similar to all other natural phenomena measurements (Ghose, 2004). Such inaccuracies may exist in other influencing parameters and can cause deviation. Therefore, proposing a formula capable of dealing with the uncertainties and inaccuracies in input parameters is required.

Accordingly, a novel optimization method is proposed in the present study to overcome the disadvantages of the previous empirical correlations by considering the uncertainties. The other main aim of this paper is to validate the previous models for V_s based on the *CPT* parameters, q_c, and f_s, via a database and to quantify the effect of uncertainties on the evaluation of the correlation parameters using the robust optimization model. The robust optimization model is the robust counterpart of the least-squares model that is reformulated as a Second-Order Cone Program (*SOCP*) in which possible uncertainties can be reasonably adjusted. The *SOCP* has been widely used in optimaization and could be applied in predicting V_s as an advancement in terms of assessment compared to previous regression methods. Also, *SOCP* considers the variation of inaccuracies and uncertainties.

The novelty of this article is the consideration of data uncertainty, which was not taken into account in previous similar articles. Therefore, this article presents an uncertainty-tolerant model for use by researchers. In other words, a model that has very low uncertainty in the parameters of its influence on the result output. It is worth noting that given the uncertainty of the data, the coefficients of the models are also revised.

The rest of the paper is organized as follows: Section 2 analyzes the uncertainty and robust optimization model. Section 3 presents the database. In Sections 4 and 5, the modeling process and main results obtained from the present study are summarized, respectively.

2. Review of the Robust Optimization Framework

Robust optimization is a modeling method where significant uncertainties are presented. The modeler aims to discover ideal solutions for the worst-case scenario of uncertainties in a specific database (Ben-Tal et al., 2009). In real-world applications, it is common to assume that the data is available with a certain level of uncertainty. Thus, classical algorithms may not meet the expectations of the modeler. Robust optimization is a framework for dealing with such situations. It should be noted that the robust approach normally involves computational complexity, such as, inaccuracies entered in the field measurements of V_s or depth (precision in measurement). Such inaccuracies may also exist in other influential parameters, which may cause prediction uncertainties. Suppose that such input-induced variations are considered boundaries of the central data point (Figure 1), and the true data point exists at every point within this boundary. In this case, robust optimization aims to minimize the maximum error with a certain level of uncertainty (Kalantary, 2013). Such input-induced deviations are considered the boundaries of the central data point (Figure 1).

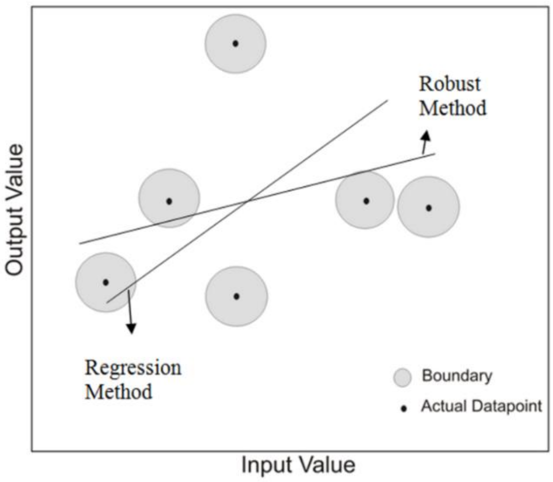

Figure 1. Robust and regression methods.

Evaluating constant coefficients (a_i) of empirical equations 1 to 12 (a_1 *to* a_6) using regression analysis can be formulated as follows:

$$Ax = b, A \in R^{m\times n}, b \in R^{m\times 1} \ and \ x \in R^{n\times 1} \tag{13}$$

where A and b are data matrices, and x is the vector of variables. Also, m and n are the number of case histories and input parameters, respectively, that show the over-determined set of equations in the case of (m>n). All models are first examined using a compiled database and simple least-squares regression analysis. The classic method for solving the least squares problem is as follows:

$$min_{x\in R^n} \|(Ax - b)\|^2 \tag{14}$$

Next, the robust least-squares model is presented in Equation (15). If the level of uncertainty in the databases is known and is equal to ρ, the robust model for minimizing the worst-case residual is as follows (Alizadeh & Goldfarb, 2003):

$$min_x max_{\|[E,r\|_F\le\rho} \|(A + E)x - (b + r)\|^2 \tag{15}$$

Where, ρ, E, and r are the uncertainties in A and b, respectively, and the norm of the matrix, $\| \ \|_F$ is the Frobenius norm (Golub & Van Loan, 2013). Equation (15) in its current form cannot be solved. However, it can be written in *SOCP* form (Sturm, 2002). First, for a given x

$$r(A, b, x) \xrightarrow{def} max \{\|(A + E)x - (b + r)\|| \mid \ \|E, r\|_F \le \rho\} \tag{16}$$

From the triangular inequality we have

$$\|(A + E)x - (b + r)\| \le \|Ax - b\| + \left\|(E, -r)\begin{pmatrix} x \\ 1 \end{pmatrix}\right\| \tag{17}$$

Moreover,

$$\left\|(E, -r)\begin{pmatrix} x \\ 1 \end{pmatrix}\right\| \le \|E, r\|_F \left\|\begin{pmatrix} x \\ 1 \end{pmatrix}\right\| \le \rho \left\|\begin{pmatrix} x \\ 1 \end{pmatrix}\right\| \tag{18}$$

But for the choice $(E, -r) = uv^t$, where

$$u = \begin{cases} \rho\dfrac{Ax - b}{\|Ax - b\|}, if \ Ax - b \ne 0 \\ any \ vector \ \in R^m \ of \ norm \ \rho, otherwise \end{cases} \quad and \ v = \frac{\binom{x}{1}}{\left\|\binom{x}{1}\right\|} \tag{19}$$

we have

$$\left\|(E,-r)\binom{x}{1}\right\| = \|u\| \times \left\|\binom{x}{1}\right\| = \rho \left\|\binom{x}{1}\right\| \tag{20}$$

Therefore,

$$r(A,b,x) = \|Ax-b\| + \rho \left\|\binom{x}{1}\right\| \tag{21}$$

Thus, min $r(A, b, x)$ is equivalent to the following *SOCP*:

$$\min(t+\rho s) \tag{22}$$

$$\|Ax-b\| \le t,$$

$$\sqrt{1+\|x\|^2} \le s$$

The problem of Equation (22) is a *SOCP* that can be solved using efficient interior point-based software packages such as SeDuMi (Sturm, 1999). It is solved for different values of uncertainty parameters, ρ. To account for the uncertainty, a new parameter of the following form is introduced:

$$\text{Uncertainty}\ \ (\%) = \frac{\rho}{2\|\text{DATA}\|_{\text{fro}}} \times 100 \tag{23}$$

The uncertainty parameter introduced in Equation (23) means that each data point can have an extreme uncertainty up to half of its value. In other words, data are assumed to be in the form of:

$$Data\ point = actual\ value \pm uncertainty = actual\ value \pm \left(\frac{actual\ value}{2}\right) \tag{24}$$

One of the advantages of this method is to find the model coefficients and find a logical relationship between the uncertainty and the predicted value. In other words, by knowing the measurement accuracy associated with the data and determining the corresponding uncertainty, the corresponding uncertainty coefficient can be determined and the model output value can be predicted from the corresponding graphs. Of course, the models must be simple and straightforward enough so that they can be converted into linear and matrix models by changing parameters or mapping.

3. Database Compilation

This study utilizes the data from a project in Eskisehir, Turkey. From a combination of *CPT* logs and V_s profiles extracted from the *SCPT* data, 407 triple data (q_c, f_s, and V_s) were obtained from 37 sites within the specified range where the V_s values come from approximately the same depth ranges. The database was divided into four soil types. Figure 2 shows the *CPT* data as a sample, and Figure 3 shows the scatter of q_c, *fs*, and V_s as a function of depth. Also, in Table 2, a sample of the database is presented. Two references, (Mola-Abasi et al., 2015) and (Tun, 2003) introduced more information about the site investigation and data process. In the mentioned articles, detailed explanations of the data and physical-mechanical conditions were given, and since this is not part of the main discussion of the article, the readers' attention was drawn to these articles for further research.

Table 2. A series of *CPT* and V_s case records in this study

Soil type	***q_c* (MPa)**	***f_s* (MPa)**	***V_s***
Clay	1.000	0.020	137.40
	2.000	0.100	167.30
	1.000	0.030	185.80
Sand	26.000	0.200	279.40
	16.000	0.100	233.70
	20.000	0.050	250.50
Mix	40.000	0.100	185.80
	1.000	0.010	153.60
	2.000	0.030	172.00

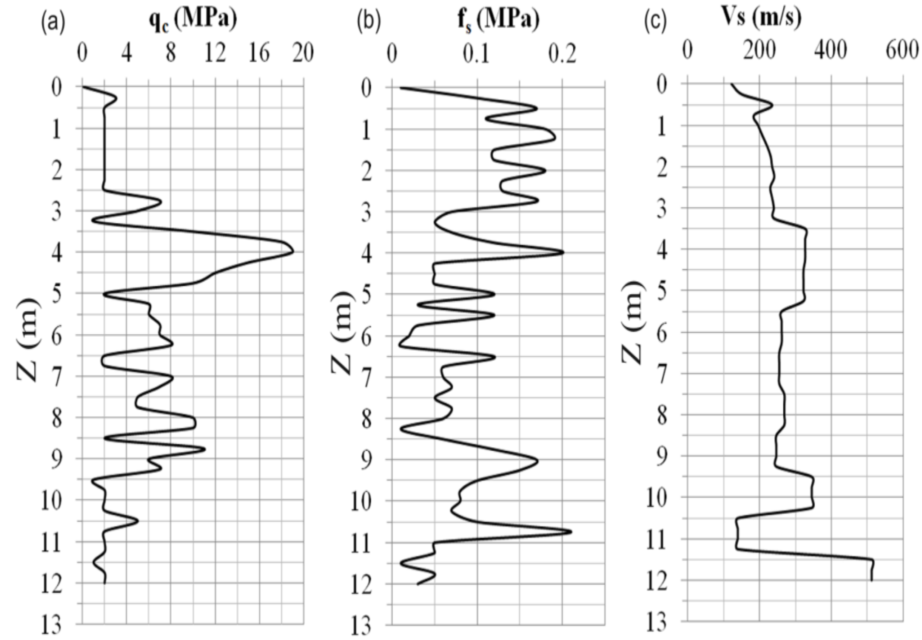

Figure 2. An example of a *CPT* and V_s record a) q_c, b) f_s and c) V_s

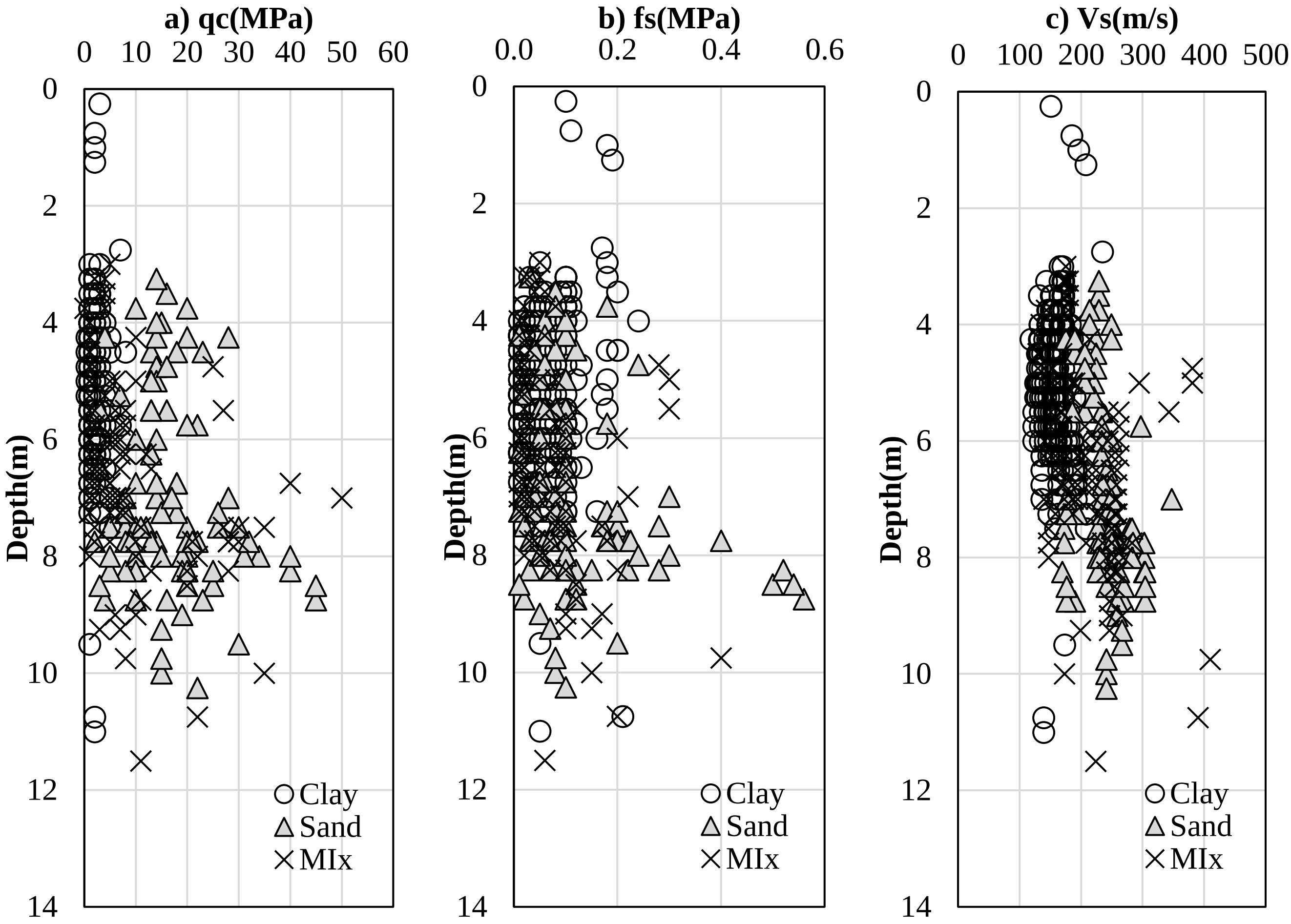

Figure 3. Depth distribution of measurement values: a) q_c, b) f_s and c) V_s for clay, sand, and mix

4. Modeling using Robust Optimization Method

Six correlation models are considered to investigate how parameter uncertainty affects the prediction of V_s. These models are similar to the great majority of those listed in Table 1, in the sense that the V_s is assumed to be independent of any soil parameter, except q_c and f_s. The six equations of the models are as follows:

$$V_S = a_1 + a_2 q_c \tag{25}$$

$$V_S = a_1 + a_2 \, ln(\, q_c) \tag{26}$$

$$V_S = a_1 + a_2 \, log(\, f_s) \tag{27}$$

$$V_S = a_1 (q_c)^{a_2} \tag{28}$$

$$V_S = a_1 (q_c)^{a_2} f_s{}^{a_2} \tag{29}$$

$$V_S = a_1 + a_2 q_c + a_3 f_s + a_4 q_c{}^2 + a_5 f_s{}^2 + a_6 (q_c f_s) \tag{30}$$

To evaluate the model's performance in this study, several performance indices, including the absolute fraction of variance (R^2) as defined in Equation (31); the root mean square error (*RMSE*) as determined by Equation (32); the mean absolute percentage error (*MAPE*) as calculated using Equation (33); and the mean absolute deviation (*MAD*) as given by Equation (34), were calculated as follows:

$$\mathrm{R}^2 = 1 - \left[\frac{\sum_{\mathrm{i}=0}^{\mathrm{M}} (\mathrm{Y(i)_c} - \mathrm{Y(i)_o})^2}{\sum_{\mathrm{i}=1}^{\mathrm{M}} (\mathrm{Y(i)_o})^2} \right] \tag{31}$$

$$RMSE = \sqrt{\frac{1}{M} \sum_{1}^{M} (\mathrm{Y(i)_o} - \mathrm{Y(i)}_c)^2} \tag{32}$$

$$MAPE = \frac{\sum_1^M |\mathrm{Y(i)_o} - \mathrm{Y(i)_c}|}{\sum_1^M C_{mi}} \times 100 \tag{33}$$

$$MAD = \frac{\sum_1^M |\mathrm{Y(i)_o} - \mathrm{Y(i)_c}|}{M} \tag{34}$$

Where, *M* is the number of data, and Y_c and Y_o are calculated and observed values, respectively. The optimal model performance will be achieved by lower *RMSE*, *MAPE,* and *MAD* values. The limitation of R^2 is between 0 to 1 and increasing R^2 leads to higher model accuracy.

5. Results

The linear regression results for the models were presented below. As shown in Table 3, model 6 and model 4 have good accuracy, but based on other statistical parameters, model 6 is slightly more accurate than model 4. One of the topics discussed in this study is investigating the effects of uncertainty, and an important question is whether models 4 and 6 (which have sufficient accuracy) have the same accuracy when the data is uncertain. In the following, the question was answered by examining the uncertainty.

Table 3. The regression results for the models

	soil type	**a1**	**a2**	**a3**	**a4**	**a5**	**a6**	**R2**	**RMSE**	**MAPE**	**MAD**
Model 1	Clay	5.59	150.13	-	-	-	-	0.9687	22.70	9.875	17.7
	Sand	4.59	209.28	-	-	-	-	0.8235	37.65	12.67	28.4
	Mixed	3.82	200.05	-	-	-	-	0.8580	34.10	12.01	25.9
	All	5.84	171.05	-	-	-	-	0.8550	34.41	12.07	26.1
Model 2	Clay	153.44	15.38	-	-	-	-	0.9695	22.62	9.86	17.7
	Sand	118.44	63.28	-	-	-	-	0.8208	37.92	12.72	28.6
	Mixed	167.72	44.12	-	-	-	-	0.8804	31.79	11.57	24.2
	All	148.77	50.25	-	-	-	-	0.8656	33.32	11.86	25.3
Model 3	Clay	175.66	11.42	-	-	-	-	0.9676	22.82	9.898	17.8
	Sand	381.57	89.93	-	-	-	-	0.8179	38.23	12.78	28.8
	Mixed	407.92	135.09	-	-	-	-	0.8875	31.06	11.44	23.7
	All	339.56	98.86	-	-	-	-	0.8313	36.85	12.52	27.8
Model 4	Clay	152.80	0.48	-	-	-	-	0.9988	19.61	9.297	15.5
	Sand	237.63	0.84	-	-	-	-	0.9958	19.92	9.355	15.7
	Mixed	175.10	2.52	-	-	-	-	0.9965	19.84	9.34	15.7
	All	171.33	1.71	-	-	-	-	0.9958	19.91	9.354	15.7
Model 5	Clay	151.60	0.09	0.00	-	-	-	0.9701	22.56	9.85	17.6
	Sand	227.67	0.13	0.08	-	-	-	0.8266	37.33	12.61	28.2
	Mixed	266.12	0.13	0.13	-	-	-	0.9151	28.22	10.91	21.7
	All	172.75	0.19	0.04	-	-	-	0.8742	32.43	11.69	24.7
Model 6	Clay	139.59	16.86	-77.19	-2.29	42.38	36.68	0.9988	19.60	9.296	15.5
	Sand	168.71	6.31	423.58	-0.06	-661.30	-0.98	0.9960	19.89	9.35	15.7
	Mixed	126.81	11.41	989.63	-0.32	-1237.26	2.90	0.9971	19.78	9.329	15.7
	All	132.99	12.96	284.00	-0.30	-399.72	7.47	0.9971	19.77	9.328	15.6

The Frobenius norm of data matrices changes due to the measurements determined to evaluate the uncertainty in V_s predictions. The variation of the coefficient (R^2) is evaluated using the uncertainties of the four soil types. The results are summarized in Figures 4 and 5. Figure 4 shows that as the uncertainty increases, the variability decreases and approaches a stable value.

It is important to note here that if the uncertainty is set to zero (for example, Figure 4f of this study), this method reduces to an ordinary least squares multiple regression technique with the same coefficients (Table 2). Figure 4 constitutes the sensitivity analysis of each coefficient of the variables related to the uncertainty level. In other words, it was proved that these respective variables had greater sensitivity for the coefficients with greater variations. Figure 5 shows the variation of R^2 versus uncertainties for each model of interest. As shown in Figure 5, the best-matched correlation amongst the others was in the form of Equation (28) (Figure 5d when uncertainty is zero). However, the least error in the estimation of V_s was obtained by the polynomial models. As illustrated in Figure 5, the models mentioned above can make fairly accurate predictions when uncertainties are not taken into account, but they are not as reliable when dealing with high levels of uncertainty. In each sub-set, the proposed method achieved more accurate predictions assuming uncertainty. As presented in Equation (30), the polynomial model provided a better fit to the observed data than the others. It can be easily seen that although the uncertainty level was almost 100 %, the value of R^2 was greater than 0.9.

For a more detailed explanation of graphs 4 and 5, it should be borne in mind that by using this method and taking data uncertainty into account, model coefficients are obtained, and the search for a logical relationship between the uncertainty and the predicted value reveals this term. In other words, by knowing the measurement accuracy associated with the data and determining the corresponding uncertainty, the corresponding uncertainty coefficient can be determined and the model output value can be predicted from the corresponding graphs. Of course, it is is important to mention that all models are converted into linear and matrix models by changing the power with the logarithm.

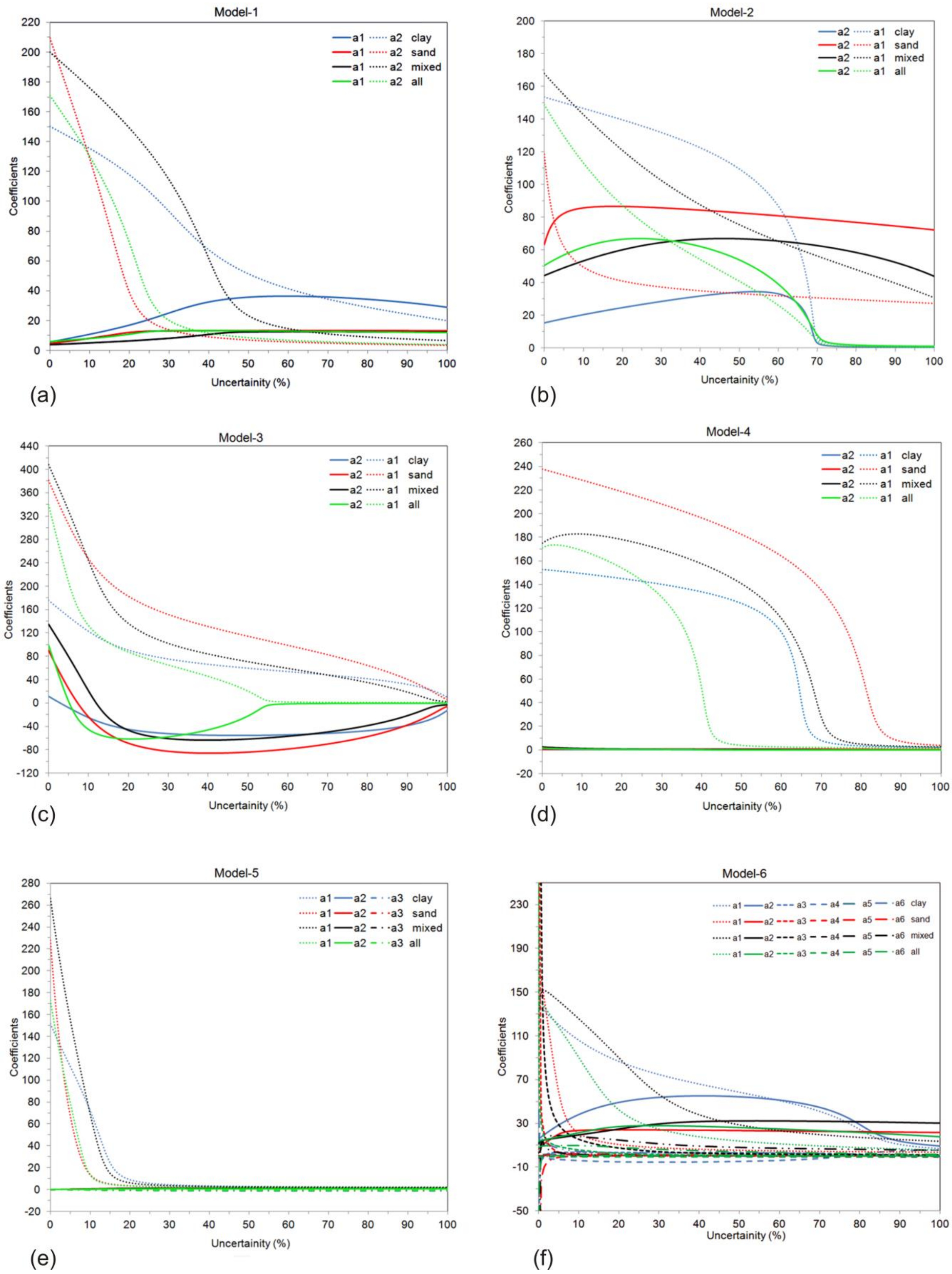

Figure 4. Variations of coefficients versus uncertainties for different models and different soil types: a) Model-1(Eq. 25), b) Model-2(Eq. 26), c) Model-3(Eq. 27), d) Model-4(Eq. 28), e) Model-5(Eq. 29) and f) Model-6(Eq. 30)

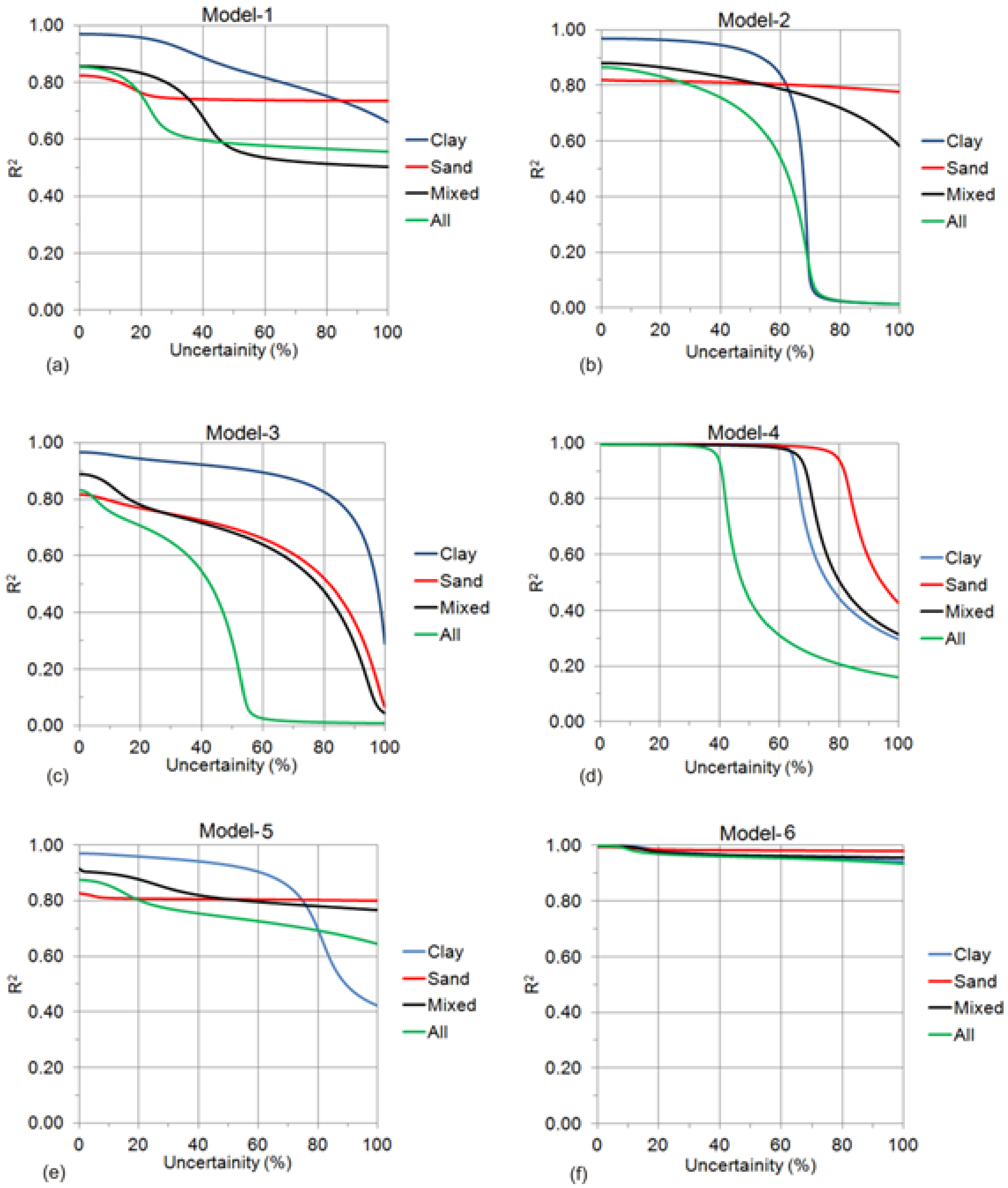

Figure 5. Variations of R^2 versus uncertainties for different models: a) Model-1(Eq. 25), b) Model-2(Eq. 26), c) Model-3(Eq. 27), d) Model-4(Eq. 28), e) Model-5(Eq. 29) and f) Model-6(Eq. 30)

6. Conclusions

Shear wave velocity is an essential engineering tool required to define the dynamic properties of soils and, preferably, can be determined indirectly by *CPT*. However, the inaccuracies in measuring or estimating the influencing parameters have consistently been a significant issue. Therefore, different statistical methods have been introduced to mitigate the impact of these inaccuracies on predicting future events by using the robust optimization model. Six correlation models were considered to investigate the impact of parameter uncertainty on the prediction of V_s. These models were similar to the great majority of those listed in Table 1 in that the V_s was assumed to be independent of any soil parameters, except for q_c and f_s. Other geotechnical soil properties can be added to obtain a better correlation, such as relative density, void ratio, porosity, and unit weight. However, this study aimed to present the simplest correlation. The databases were evaluated by dividing them into four different groups, namely, clay, sand, mixed, and all soils. Six empirical correlation models, which different researchers introduced, were evaluated together with the proposed model. Among the previously proposed equations, the equations proposed in the form of Equation (28) gave the highest R^2 value. A robust optimization technique was developed to assess the impact of uncertainty of each model parameter, independently of the analysis results. This is an advance compared to limited stochastic approaches that consider parameter variations individually. A new parameter was introduced to represent the level of uncertainty in the data. Statistical comparison of the models showed that the accuracy of the model based on Equation (28) is generally close to the polynomial model at very small uncertainty values. However, when data uncertainty is high (especially for the parameters mentioned above), the new polynomial model performes better. All the results obtained in this study showed that such correlations resulting from local records should not directly be used for V_s. The polynomial proposed relationships could be used for measuring V_s. We recommend reanalyzing the presented model using data from other regions. It is proposed to adapt the methods of this work to new open model data.

References:

Alizadeh, F., Goldfarb, D. (2003). “Second-order cone programming”, *Mathematical Programming*, 95(1), 3–51, http://dx.doi.org/10.1007/s10107-002-0339-5.

Anagnostopoulos, A., Koukis, G., Sabatakakis, N., Tsiambaos, G. (2003). “Empirical correlations of soil parameters based on cone penetration tests (CPT) for Greek soils”, *Geotechnical and Geological Engineering*, 21, 377-387, http://dx.doi.org/10.1023/B:GEGE.0000006064.47819.1a.

Barrow, B.L., Stokoe, K.E.II. (1983). “Field investigation of liquefaction sites in Northern California”, Geotechnical Engineering Thesis, GT 83-1, *Civil Engineering Department, University of Texas at Austin*, 212pp.

Bayat, M., Saadat, M., & Hojati, A. (2023). “Optimization of Dynamic Compaction Procedure for Sandy Soils”, *Civil Engineering Infrastructures Journal*. http://dx.doi.org/ 10.22059/CEIJ.2023.351287.1889.

Ben-Tal, A., El Ghaoui, L., and Nemirovski, A. (2009). “Robust optimization”, *Princeton university press*, 28, http://dx.doi.org/10.1515/9781400831050.

Chala, A. and Ray, R. (2023). “Machine Learning Techniques for Soil Characterization Using Cone Penetration Test Data”, *Applied Sciences Journal*, 13(14), 8286, http://dx.doi.org/10.3390/app13148286.

Comina, C., Foti, S., Passeri, F., Socco, L.V. (2022). “Time-weighted average shear wave velocity profiles from surface wave tests through a wavelength-depth transformation”, *Soil Dynamics and Earthquake Engineering*, 158, http://dx.doi.org/10.1016/j.soildyn.2022.107262.

Eslami, A., Akbarimehr, D., Aflaki, E., Hajitaheriha, M.M. (2020). “Geotechnical site characterization of the Lake Urmia super-soft sediments using laboratory and CPTu records”, *Marine Georesources & Geotechnology*, 38, http://dx.doi.org/10.1080/1064119X.2019.1672121.

Ghose, R., (2004). “Model-based integration of seismic and CPT data to derive soil parameters” , *Proc. 10th European Meeting of Environmental and Engineering Geophysics*, B019, http://dx.doi.org/10.3997/2214-4609-pdb.10.B019.

Gilder, C. E., Pokhrel, R. M., De Luca, F., & Vardanega, P. J. (2021). “Insights from CPTu and seismic cone penetration testing in the Kathmandu Valley, Nepal", *Frontiers in Built Environment*, 7, 646009, http://dx.doi.org/10.3389/fbuil.2021.646009.

Golub, G. H. and Van Loan, C. F. (2013). "Matrix Computations", *3rd ed. Baltimore, MD: Johns Hopkins*, http://dx.doi.org/10.56021/9781421407944.

Hegazy Y.A., Mayne, P.W. (1995). "Statistical correlations between Vs and cone penetration data for different soil types", *Proceedings of the International Symposium on Cone Penetration Testing (CPT'95)*, Linkoping, Sweden, 4-5 October 1995, Swedish Geotechnical Society 2, 173–178.

Iyisan, R., Ansal, A. (1993). "Determination of dynamic soil properties by borehole seismic methods" , *2th National Earthquake Engineering Conference, Proc*. of the Chamber of Civil Engineers Turkey (in Turkish).

Jakka, R., Desai, A., Foti, S., (2022) "Guidelines for Minimization of Uncertainties and Estimation of a Reliable Shear Wave Velocity Profile Using MASW Testing: A State-of-the-Art Review", *Advances in Earthquake Geotechnics*, pp 211–253, http://dx.doi.org/10.1007/978-981-19-3330-1_12.

Kalantary, F., MolaAbasi, H., Salahi, M., and Veiskarami, M. (2013). "Prediction of liquefaction induced lateral displacements using robust optimization model", *Scientia Iranica*, 20(2), 242-250.

Kruiver, P. P., de Lange, G., Kloosterman, F., Korff, M., van Elk, J., & Doornhof, D. (2021). "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, 106471, http://dx.doi.org/10.1016/j.soildyn.2020.106471.

Madiai, C., Simoni, G. (2004). "Shear wave velocity-penetration resistance correlation for Holocene and Pleistocene soils of an area in central Italy", *Proceedings ISC-2 on Geotechnical and Geophysical Site Characterization*, Viana da Fonseca & Mayne (eds.), Millpress, Rotterdam, ISBN 90 5966 009 9, 1687-1694.

Mayne, P.W., Rix, G.J. (1995). "Correlations between shear wave velocity and cone tip Resistance in natural clays", *Soils and Foundations*, 35(2), 107-110.

Mayne, P.W. (2006). "Undisturbed sand strength from seismic cone tests", *Geomechanics and GeoEngineering Taylor & Francis Group*, London, 1(4), 239-257.

Mayne, P.W. (2007). “Cone Penetration Testing, National Cooperative Highway Research Programme”, Synthesis 368, *Transportation Research Board*, USA.

McGann, C. R., Bradley, B. A., Taylor, M. L., Wotherspoon, L. M., & Cubrinovski, M. (2015b). “Development of an empirical correlation for predicting shear wave velocity of Christchurch soils from cone penetration test data”, *Soil Dynamics and Earthquake Engineering*, 75, 66- 75, http://dx.doi.org/10.1016/j.soildyn.2015.03.023.

Meng. F. and Pei, H. (2023). “Quasi-site-specific prediction of shear wave velocity from CPTu”, *Soil Dynamics and Earthquake Engineering*, 172, http://dx.doi.org/10.1016/j.soildyn.2023.108005.

Mohammadikish, S., Ashayeri, I., Biglari, M. et al. (2023). “Soil Liquefaction Assessment by CPT and VS Data and Incomplete-Fuzzy C-Means Clustering”, *Geotechnical and Geological Engineering*, https://doi.org/10.1007/s10706-023-02669-1.

Mola-Abasi, H., Dikmen, U., and Shooshpasha, I. (2015). “Prediction of shear-wave velocity from CPT data at Eskisehir (Turkey), using a polynomial model”, *Near Surface Geophysics*, 13(2), 155-167, http://dx.doi.org/10.3997/1873-0604.2015010.

Paoletti, L., Hegazy, Y. Monaco, S., Piva, R. (2010). “Prediction of shear wave velocity for offshore sands using CPT data – Adriatic Sea”, *2nd International Symposium on Cone Penetration Testing, Huntington Beach*, CA, USA, 1-8.

Robertson, P.K. (2009). “Interpretation of cone penetration tests – a unified approach”, *Canadian Geotechnical Journal*, 46(11), 1337–1355.

Stolte, A. and Cox, B. (2020).“Towards consideration of epistemic uncertainty in shear-wave velocity measurements obtained via seismic cone penetration testing (SCPT)”, *Canadian Geotechnical Journal*, 57, http://dx.doi.org/10.1139/cgj-2018-0689.

Sturm, J.F. (1999). “Using SeDuMi 1.02. a MATLAB toolbox for optimization over symmetric cones”, *Optimization methods and software*, Taylor and Francis, 11(1-4), pp.625-653.

Sturm, J. (2002). “Implementation of Interior Point Methods for Mixed Semidefinite and Second Order Cone Optimization Problems”, *Optimization Methods and Software*, 17 (6), 1105-1154.

Sykora, D.W., Stokoe, K.H. (1983). “Correlations of in situ measurements in sands of shear wave velocity, soil characteristics and site conditions”, *The University of Texas, Austin*, *Geotechnical Engineering Report*, GR83-33

Tun, M., (2003). “Investigation of the characteristics of Eskisehir soils due to shear wave velocity and determination of their fundamental vibration periods”, *Master Thesis, Anadolu University Institute of Science and Technology Department of Physics* (in Turkish).

Tun, M., Ayday, C. (2018). “Investigation of correlations between shear wave velocities and CPT data: a case study at Eskisehir in Turkey”, *Bulletin of Engineering Geology and the Environment*, 77, 225–236. https://doi.org/10.1007/s10064-016-0987-y.

Wang, J., Hwang, J., Lu, C. (2022). “Measurement uncertainty of shear wave velocity: A case study of thirteen alluvium test sites in Taipei Basin”, *Soil Dynamics and Earthquake Engineering*, 155, http://dx.doi.org/10.1016/j.soildyn.2022.107195.

Wang, JS., Hwang, JH., Lu, CC. et al. (2022). “Empirical formulas for shear wave velocity prediction and their uncertainties: a case study of thirteen alluvium test sites in the Taipei Basin”, *Bulletin of Engineering Geology and the Environment*, 81, 450. https://doi.org/10.1007/s10064-022-02949-9.

Wang, T., Xiao, S., Zhang, J. and Zuo, B. (2022). “Depth-consistent models for probabilistic liquefaction potential assessment based on shear wave velocity” , *Bulletin of Engineering Geology and the Environment*, 81, 255, http://dx.doi.org/10.1007/s10064-022-02754-4.

Yang, H., Liu, Z., Xie, Y., Li, S. (2023). “A probabilistic liquefaction reliability evaluation system based on CatBoost-Bayesian considering uncertainty using CPT and Vs measurements”, *Soil Dynamics and Earthquake Engineering*, 173, http://dx.doi.org/10.1016/j.soildyn.2023.108101.

Zhai, S., Du, G. & He, H. (2024). “Bayesian probabilistic characterization of the shear-wave velocity combining the cone penetration test and standard penetration test”, *Stochastic Environmental Research and Risk Assessment*, 38(1), 69-84, https://doi.org/10.1007/s00477-023-02566-2.

Zhang, Y., Xie, Y., Zhang, Y., Qiu, J. and Wu, S. (2021). “The adoption of deep neural network (DNN) to the prediction of soil liquefaction based on shear wave velocity”, Bulletin of *Engineering Geology and the Environment*, 80(6), 5053–5060, http://dx.doi.org/10.1007/s10064-021-02250-1

Zhao, Z., Duan, W., Cai, G. (2021). “A novel PSO-KELM based soil liquefaction potential evaluation system using CPT and Vs measurements”, *Soil Dynamics and Earthquake Engineering,* 150, http://dx.doi.org/10.1016/j.soildyn.2021.106930.

Zhao, Z., Duan, W., Cai, G., Wu, M., Liu., S. (2022). “CPT-based fully probabilistic seismic liquefaction potential assessment to reduce uncertainty: Integrating XGBoost algorithm with Bayesian theorem”, *Computers and Geotechnics*, 149, http://dx.doi.org/10.1016/j.compgeo.2022.104868.

Zhou, J., Huang, S., Wang, M. and Qiu, Y. (2022). “Performance evaluation of hybrid GA–SVM and GWO–SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation”, *Engineering with Computers*, 38, 4197–4215, http://dx.doi.org/10.1007/s00366-021-01418-3.