Random-effects regression model for shear wave velocity as a function of standard penetration test resistance, vertical effective stress, fines content, and plasticity index

Mohammad Motalleb Nejad[1], Kalehiwot Nega Manahiloh[2], Mohammad Sadegh Momeni [3]

Abstract

A random effect regression model is used to formulate a unified equation and estimate shear wave velocity (V_s). Standard penetration test (*SPT*) number, effective overburden pressure, plasticity index, and the fines content (F_c) are used as input parameters. First, a fixed model regression is used to obtain the regression parameters. *SPT* number and shear wave velocity are measured at 2 m intervals up to a depth of 10 m, for 71 boreholes, distributed evenly in Urmia city. Plasticity index and fines content are evaluated from laboratory tests that were performed on 355 samples obtained from the 71 boreholes (i.e., 5 samples from each borehole). Statistical analysis performed on the fixed effect model showed the need for examining the random effects arising from variable *SPT* test conditions in each borehole. A mixed effect regression model is employed to investigate such effects. The distribution of residuals is found to satisfy the normality criteria for the mixed effect model. A strong fit for the model is obtained, and through statistical evidence, it is implied that the proposed model is practical. The model's most prominent feature is the capability of unifying different soil types via the incorporation of plasticity index and fines content as inputs.

Key words: Shear wave velocity; standard penetration test; mixed effect regression; statistical analysis; random effects; unified equation.

Citation Information: Please cite this work as follows

Motalleb Nejad, M., K.N. Manahiloh, and M.S. Momeni, *Random-effects regression model for shear wave velocity as a function of standard penetration test resistance, vertical effective stress, fines content, and plasticity index.* Soil Dynamics and Earthquake Engineering, 2017. **103**: p. 95-104.

[1] University of Delaware, Department of Civil and Env. Engineering, 301 DuPont Hall, Newark, DE 19716.
[2] Corresponding Author: University of Delaware, Department of Civil and Env. Engineering, 301 DuPont Hall, Newark, DE 19716, email: knega@udel.edu.
[3] ZTI Consulting Engineers, 1st street, Dorostkar Street, Urmia, Iran.

Introduction

Casualties and massive infrastructure damages indicate the urgent need for dynamic site characterization and robust models for seismic evaluation of a site's sustainability during natural hazards. Seismic characterization of a site is necessary to minimize damage caused by earthquakes. One of the oldest, yet efficient approaches of seismic characterization is performed by systematically obtaining statistical models that estimate the response of soil layers to earthquake excitations. The incorporation of a set of appropriate geotechnical properties plays a pivotal role in the efficiency of such models. In addition, these geotechnical variables should be correlated to a unique practical seismic parameter. Shear wave velocity is widely used for this purpose. The strong correlation of shear wave velocity with maximum shear modulus (G_{max}) of a soil is a great indicator of its importance in earthquake analysis. G_{max} can be correlated to the deformation potential of a given site during a seismic action.

Figure 1 represents a typical modulus reduction curve that shows the rate of decrease in shear modulus with an increase in strain level. Several curves representing this relationship were created for different types of soils (e.g., [1-4]). For very small strain levels (i.e., approximately equal or less than 10^{-3}), the shear modulus of the soil is very close to the value of G_{max}. Therefore, employing an appropriate method to obtain the shear wave velocity of the soil for very small strains is necessary in seismic analysis. Once the shear wave velocity is obtained for very small strains, the small strain shear modulus can be computed as $G_{max}=\rho V_s^2$. In addition, V_s is directly used for ground motion prediction using next generation attenuation relations [5-9]. These relations employ V_{s30} as a required variable which is defined by Choi and Stewart [10] as the average V_s in the upper 30m of the ground. Boore [11] proposed four methods to estimate V_{s30} for situations where data is not available for up to 30 meters below the ground level. In general, when it comes to seismic

analysis, V_s and G_{max} are the most important parameters employed in soil classification, liquefaction potential, and soil-structure interaction analysis [10].

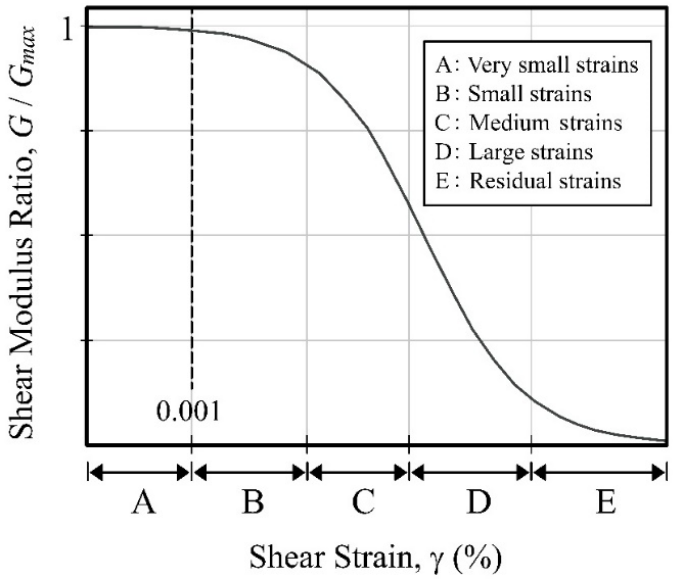

Figure 1 - Typical modulus reduction curve.

There are three different approaches utilized for obtaining the shear wave velocity of soils. The first two approaches make use of laboratory and geophysical field measurements. The third approach, which is adopted in this article, aims at obtaining a robust correlation between shear wave velocity and simple geotechnical parameters (i.e., index properties) of a given soil.

Laboratory measurements of shear wave velocity require devices that are precise enough to measure the shear wave velocity at very small strain levels. For example, resonant column test is used to obtain shear wave velocity from the resonant frequency and the weight and dimensions of specimens [12-15]. Bender elements and shear plates are two types of piezoelectric transducers that are used to obtain the shear wave velocity from the distance, between the two transducers located at two ends of a specimen, and the wave travelling time [16]. Piezoelectric transducers are accommodated in a cyclic triaxial apparatus combined with precise axial strain measurement devices to obtain the shear wave velocity. The accuracy of the results for laboratory measurements

is highly sensitive to sample disturbance. During sampling, the weak boundaries between soil particles are broken and some level of disturbance occurs. Since the effect of sampling disturbance on the stiffness of the soil is remarkable for low strain laboratory tests, accurate results for shear wave velocity measurements are not possible unless expensive freezing techniques are used [17].

Seismic geophysical field tests are the most reliable methods to obtain the shear wave velocity for a soil at various depths. Crosshole test (CHT)[18], downhole test (DHT) [19], seismic cone penetration tests (SCPTs) [20, 21], multichannel analysis of surface waves (MASW) [22], and spectral analysis of surface waves (SASW) [23] are well established methods for V_s measurement at very small strain levels. Although these tests are performed in low disturbance conditions, various restrictions such as space, cost, and noise limit their universal utilization. In addition, field tests have proven to be rather expensive and time consuming. The aforementioned drawbacks of the laboratory and field tests led to the development of a new approach in which statistical methods are used to correlate shear wave velocity and simple geotechnical parameters, such as the index properties of a soil.

Statistical approaches are powerful tools often used to find correlations between shear wave velocity and geotechnical parameters of soil. Most studies in this area attempted to correlate V_s with *SPT* blow counts (N), directly [24-33]. Table 1 summarizes some of the empirical relationships suggested to estimate V_s from SPT number (N) and depth (D) of the soil.

Hara [34] found that dynamic Poisson's ratios were insignificantly influenced by the change in Young's moduli when axial strains were in the order of 10^{-3}. The very first attempt to present a relationship between shear moduli and *SPT* N-value of the soil was made by Kanai *et al.* [35], where they introduced two linear boundaries for the relationship between shear modulus and *SPT* values for clay and sand. Since their pioneering work, other researchers have attempted to

obtain similar correlations for different types of soils. Imai and Yoshimura [26, 27] correlated the mechanical properties of soils to the primary and secondary body wave velocities. They directly measured S- and P-wave velocities using a PS logging system. They acquired the standard penetration test resistance, shear wave velocity, and unconfined compressive strength of soil samples from 242 boreholes that were distributed all over Japan. They proposed three separate equations for three types of soils and correlated the *SPT* blow count, *N*, to the shear wave velocity of clay, sand, and silt. It was the first time that the shear wave velocity of soil was correlated to *N* using an exponential form (i.e., $V_s=AN^B$).

Table 1- Empirical relationships between shear wave velocity, SPT resistance, and depth.

References	**Soil type**			
	All	**Sand**	**Clay**	**Silt**
Imai and Yoshimura [26]	$V_s=76N^{0.39}$	-	-	-
Ohba and Toriumi [31]	$V_s=84N^{0.31}$	-	-	-
Imai and Yoshimura [27]	$V_s=91N^{0.337}$	$V_s=80.6N^{0.331}$	$V_s=102N^{0.292}$	-
Seed and Idriss [32]	$V_s=61.4N^{0.5}$	-	-	-
Seed et al. [33]	-	$V_s=56.4N^{0.5}$	-	-
Jinan [29]	$V_s=116(N+.318)^{0.202}$ $V_s=90.9(D+.62)^{0.212}$	-	-	-
Lee [36]	-	$V_s=57.4N^{0.49}$ $V_s=57.4D^{0.46}$	$V_s=114.4D^{0.31}$	$V_s=105.6D^{0.32}$
Iyisan [28]	$V_s=51.5N^{0.516}$	-	-	-
Hasançebi and Ulusay [25]	$V_s=90N^{0.309}$	$V_s=90.82N^{0.319}$	$V_s=97.89N^{0.269}$	-
Anbazhagan and Sitharam [37]	$V_s=78N_{1\text{-}60}^{0.4}$	-	-	-
Dikmen [24]	$V_s=58N^{0.39}$	$V_s=73N^{0.33}$	$V_s=44N^{0.48}$	-
Brandenberg et al. [38]	$\mathrm{Ln}(Vs)\mathrm{ij}=\beta_0+\beta_1\,\mathrm{Ln}(N_{60})_{ij}+\beta_2\,\mathrm{Ln}(\sigma')_{ij}$			
Kuo et al [30]	$V_s=114N^{0.56}D^{0.168}$			
Ghorbani et al. [17]	$V_s=3.02+a.8839Y_2+0.9307Y_3+0.33683Y_2^2+0.3532Y_3^2+0.6899Y_2Y_3$ $Y_3=f(\sigma',Y_1)$, $Y_2=f(N_{1\text{-}60},Y_1)$, $Y_1=f(\sigma', N_{1\text{-}60})$			

The units for *D*, V_s and σ' are foot, meter per second and kPa, respectively.
βs in equation proposed by Brandenberg et al.(2010) are presented for different types of soils.

Ohba and Toriumi [31] and Ohta et al. [39] obtained same-type equations for alluvial soil deposits including sandy, clayey, and their alternate layers. Since then, the work of these

researchers have been followed by others in an attempt to obtain similar correlations for different types of soils. Seed and Idriss [1, 32] proposed a simple equation that correlates shear wave velocity of a soil to *SPT* values. Using statistical analysis, Lee [36] presented numerous regression models that estimate shear wave velocity from *SPT* resistance, depth, effective overburden pressure, and soil type. The effect of *SPT* number (*N*) and soil type on shear wave velocity of the soil has been studied by Iyisan [28]. The study showed that the same *N* values result in the same V_s for different type of soils, with an exception of gravel. Hasançebi and Ulusay [25] presented several equations for V_s versus *SPT N* number by using 97 sets of data gathered from the Northwest area of Turkey.

Different equations were obtained for clayey and sandy soils. Using datasets gathered from seismic micro zonation studies in India, Anbazhagan and Sitharam [37] presented an equation to determine the shear wave velocity of the soil based on the modified standard penetration test ($N_{1\text{-}60}$). Brandenburg et al. [38] used statistical regression analysis and presented an equation to estimate V_s for soils under Caltrans bridges. Gathering datasets from 79 logs in 21 bridges, they correlated the natural logarithm of V_s (i.e., $Ln(V_s)$) with *SPT N* number and effective overburden pressure for sandy, silty, and clayey soils. Using datasets gathered from Taiwan, Kuo et al. [30] presented an equation to determine V_s based on SPT *N* number and depth. Ghorbani et al. [17] presented an equation to estimate the shear wave velocity from the modified *SPT N* number, $N_{1\text{-}60}$, and effective overburden pressure. They employed polynomial neural networks for their model and used datasets from different zones of the world.

The effect of parameters like *SPT*, effective overburden pressure, the percentage of fine grains, depth, and tip resistance in the cone penetration test for the shear wave velocity of soils have been studied by multiple researchers [e.g., 24, 25, 28, 38]. The outcome from different studies

performed in the past has been diverse. Several studies have revealed that effective overburden pressure, porosity, and geological age influence the value of G_{max} for different soils. Others have reported that pre-consolidation stress has a negligible effect on G_{max} [4, 40-43]. The effect of plasticity index (*PI*), on shear wave velocity of soils, remained controversial. Some studies showed a direct relationship between G_{max} and *PI* (e.g., [4, 41, 42]) while reverse relationships were reported by others (e.g.,[4, 41, 42, 44]). Hardin and Drnevich [41] showed that the most influential parameters in the evaluation of G_{max} and V_s of soils are unit weight, porosity, and effective pressure. Age and cementation have been found to have little or no effect on G_{max} and V_s, depending on the type of soil. Direct relationships between V_s and vertical effective pressure, age, cementation, and the pre-consolidation stress have been obtained by Dobry and Vucetic [45]. They also showed that a reverse relationship governs the correlation between V_s and porosity.

By reviewing the existing literature, it can be said that a one-parameter linear equation is not capable of correlating shear wave velocity and the index properties of a soil. In addition, the scattered datasets and the associated very-weak trend lines suggest the importance of including the effect of parameters other than the *SPT N* number. On the other hand, the wide variety of soil types makes it difficult to define variables that have strong correlation with shear wave velocity. Nonetheless, efforts should continue to be made to obtain correlations that adequately estimate the seismic behavior of soils, while limiting the utilization of existent correlations to the estimation of the need for seismic consideration.

Some researchers have considered depth, as an input parameter for shear wave velocity estimation (e.g., [29, 36]), while the unit weight of the soil is neglected. However, effective overburden pressure can be argued to be a better input-variable as it nicely captures the combined effect of the unit weight and depth of the soil. Statistics of the models proposed by all studies show

that N_{60} is the most significant parameter in shear wave velocity estimation and cannot be ignored. N_{60} unifies the characteristics such as stiffness, and young and shear moduli for soils in a single field test parameter. In addition, the range of N_{60} is wide enough to allow the consideration of the most influential parameters that play a role in shear wave propagation through soil. The simplest way to incorporate the effect of soil type is to include plasticity index and fines content as input variables. Plasticity index contains information about pre-consolidation stress, mineralogy, and energy absorption during seismic stimulations. A combination of fine grain percentage and plasticity index increases the robustness of a model and allows the realization of a unified model that captures behavior across multiple soil types. In this study, a laboratory and field data gathered from boreholes and samples distributed throughout the city of Urmia, Iran, has been used to find a unified correlation between shear wave velocity (V_s), *SPT N* number (N_{60}), effective overburden pressure (σ_v'), fine grains percentage (F_c), and plasticity index (PI).

In addition to proposing the correlation, fixed and random effect models have been utilized to perform regression analysis. In soil studies, random effect models are mostly used to perform regression analysis for earthquake ground motions. The pioneering work of Abrahamson and Youngs [46] proposed a random effect model to evaluate ground motions partitioned into intra-event and inter-event regression terms. Yaghmaei-Sabegh et al. [47] proposed an empirical random effect regression model to predict earthquake ground motion duration in Iran. Jayaram and Baker [48] evaluated the effect of spatial correlation consideration on ground-motion models by using a mixed effect regression model.

Random effect model is also used to obtain attenuation relationships for ground motion. Takahashi et al. [49] presented a spectral acceleration attenuation model for response spectra using data-set derived from Japanese strong-motion records. They showed that the influence of source

depth, tectonic source type, and faulting mechanism on ground motion attenuation are significant. Özbey et al. [50] employed random effect model and used 195 recordings from 17 recent events for Northwestern Turkey ground motions to develop empirical attenuation relationships for the geometric mean of horizontal peak ground acceleration. Another example of application of random effect regression model in geotechnical engineering is the study performed by Shahi and Baker [51] where the probability of near-fault earthquake ground motion pulses, and their period is modeled.

Similar to the current study, Brandenburg et al. [38] proposed a random effect regression model to obtain shear wave velocity as a function of vertical stress and SPT number.

Experimental work and data acquisition

Location

As the 10th most populated city in Iran and the second largest city in the Iranian Azerbaijan, the city of Urmia benefits greatly from data-driven geotechnical and seismic studies that contribute to the minimization of damage as a result of earthquakes and other natural hazards. The world's sixth-largest endorheic saltwater lake, "Lake Urmia," lies to the east of the city. In addition, the rate of precipitation has led the city towards becoming a major trading center for fruit produce. All of these features combined with the picturesque mountainous terrain makes the city a mesmerizing destination to visitors. The city of Urmia lies on Cenozoic stress fields and faults in its very eastern border, and has been shaken by several high intensity earthquakes. Urmia has been listed among the earthquake prone cities and any urban development in the city is mandated to perform comprehensive seismic analysis.

Experimental work

In this study, two field tests, namely the standard penetration test (*SPT*), and downhole test and two laboratory tests, namely the Atterberg limit test, and sieve analysis are conducted to gather index parameters that are important for a regression model. The field data was gathered from 71 boreholes distributed throughout the city of Urmia. Samples were collected in each borehole at two meter intervals up to a depth of 10 m.

Standard Penetration Tests (*SPT*) are performed on all boreholes at every sampling point and the blow counts (i.e. SPT *N* values) are measured. The retrieved samples are transferred to the geotechnical laboratory where the Atterberg limit tests and sieve analyses are performed for each sample. A downhole test is performed for each borehole that was dug up to the engineering bedrock level. In this study, the engineering bedrock was considered as the depth at which the shear wave velocity is equal or greater than 700 m/sec. The depth of bedrock (i.e. depth of borehole) varied between 10 to 50 m. A mobile sensor was embedded inside each borehole to receive the shear waves generated by a simple hammer. The travelling time of the seismic waves was analyzed and the shear wave velocity was obtained for the depths of interest. PVC pipes of three to six inches in diameter were used to stabilize the walls of the boreholes. Although the data-measuring device benefited from noise minimizing software, the downhole test operations were performed at night to minimize initial noise.

Data acquisition

Overall, 355 datasets aree collected from 71 boreholes. At each location shear wave velocity, SPT number (*N*), effective overburden pressure (σ_v'), fine grain percentage (F_c), and the plasticity index (*PI*) of the soils are measured. The geographic coordinates and satellite view of the city of Urmia and borehole locations are shown in Figure 2. Table 2 shows the statistics of the data obtained

from filed and laboratory tests. Table 3 presents the number of specimens and/or data collection points for each soil type used in this study.

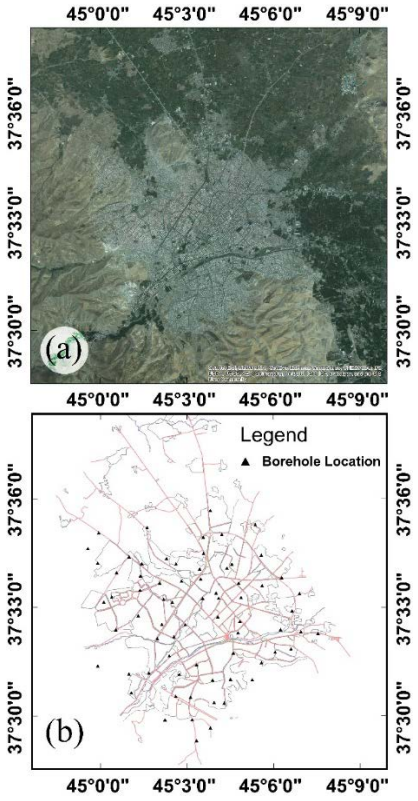

Figure 2 - (a) Satellite view of Urmia city; (b) location of boreholes.

Table 2 - Statistical indices for data obtained from field tests.

Variables	Statistics			
	Max	**Min**	**Mean**	**Standard deviation**
N_{60}	130	4.67	73.82	38.25
σ_v'(kPa)	176.4	17.3	87.85	49.13
F_c	98	6	71.51	26.6
PI	55.8	0	5.87	9.86
V_s (m/s)	652	90	383.3	123.91

Table 3 - Number of data points for each soil type used in this study.

Soil	Clay	Silt	Sand	Gravel	Total
Number	98	180	51	26	355

Statistical regression models

The typical form of the regression equation for V_s and N_{60} is as follows:

$$V_s = A(N_{60})^B \quad (1)$$

The basic form of the equation that has been utilized for the proposed statistical model is identical with the general form introduced in the introduction. The value of *SPT* blow count, *N*, is not corrected for overburden pressure but is modified for energy (i.e., 60% energy transfer from the safety hammer to the drill rod), length of the probe, and internal diameter of the sampler. This is to allow for the inherent contribution of overburden pressure in the regression model. For simplicity, the modified *SPT N* value, that is not corrected for overburden pressure is represented by $N_{m\text{-}60}$.

Input Parameters

The input parameters have been normalized first to ease the interpretation of the regression coefficients. The value of $N_{m\text{-}60}$ is normalized with respect to a fixed number of one hundred.

$$X_1 = X_N = (\frac{N_{m-60}}{100}) \times 100 = N_{m-60} \quad (2)$$

The final form of the regression model will be written as a logarithmic equation. Therefore, the plasticity index of the soil is increased by one in order to prevent undefined values from the model.

$$X_2 = X_{PI} = (PI + 1) \quad (3)$$

Since the value of F_c is reported as the percentage of particles passing sieve number 200, there is no need to normalize this parameter. However, with the same logic described for the plasticity index, F_c is increased by one percent.

$$X_3 = X_{F_c} = \left(F_c + 1\right) \quad (4)$$

It should be noted that Equations (3) and (4) satisfy the possible boundary values for *PI* and F_c within the general regression equation obtained in the following section of the paper.

The value of overburden pressure is normalized with respect to atmospheric pressure as follows.

$$X_4 = X_S = \left(\frac{\sigma'_v}{P_a = 101\ Kpa}\right) \times 100 \tag{5}$$

Derivation of Regression equation

The statistical equation chosen for the regression has been inspired by the typical form of regressions used to correlate V_s to *SPT* number. The shear wave velocity is usually corrected for overburden pressure using exponential functions [e.g., 38, 52, 53]. Branderberg *et al*. [38] combined the general form of the shear wave velocity estimation equation with the correction factor function for overburden pressure. In this study, it has been assumed that the correction equation for plasticity index (*PI*) and fines content (F_c) follows an exponential function and the preliminary form of the regression model is written as follows:

$$V_s = A\left(\prod_{i=1}^{4} X_i^{m_i}\right)^B \tag{6}$$

Where *A* and *B* are the regression parameters and each *X* indicates the index parameter that contributes to shear wave velocity estimation. One can take the natural logarithm of both sides of the equation to obtain a linear form of the equation that can be used for linear regression as follows:

$$\ln(V_s) = \ln A + B\sum_{i=1}^{4} m_i \ln X_i \tag{7}$$

Expanding the Equation 7 using the four input parameters considered in this study and accounting for residuals from the regression model (i.e., fixed effect errors) and any errors from *k* number of other sources (i.e., random errors), the regression model take the following form.

$$\ln(V_s) = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \beta_4 \ln X_4 + \varepsilon + \sum_{source=1}^{k} (Error_k) \quad (8)$$

Fixed Effect Regression Model

A simplest estimate of the model can be made by assuming other sources of error to be negligible. Such an assumption leads to the following equation:

$$\ln(V_s) = \beta_0 + \beta_N \ln(X_N) + \beta_{PI} \ln(X_{PI}) + \beta_{F_c} \ln(X_{F_c}) + \beta_S \ln(X_S) + \varepsilon \quad (9)$$

Where ε represents the residuals due to the difference between the predicted and real values of the natural logarithm of the shear wave velocity for a given dataset. The subscripts *N*, *PI*, F_c and *S* refer to the model parameters *SPT* resistance, plasticity index, fines contents and vertical overburden pressure, respectively. The *Lm* function in the R software has been utilized to obtain the fixed model for the equation. The results of the fixed model regression are presented in Table 4. The values of the coefficient of correlation (CoC) and root mean square error (RMSE) indicate that the model has successfully fitted to the data.

Table 4 - regression parameters for fixed effect regression model.

β_0	β_N	β_{PI}	β_{F_c}	β_S	**CoC**	**RMSE**
3.79363	0.44715	0.02596	0.02964	0.02827	0.874	52.14

The residual for each set of input parameters is defined as ε = ln (V_S) - $[\beta_0 + \beta_N \ln X_N + \beta_{PI} \ln X_{PI} + \beta_{F_c} \ln X_{F_c} + \beta_S \ln X_S]$. The Quantile-Quantile probability plot has been used to show the standardized vs. normal theoretical residuals (Figure 3). The set of residuals are standardized using the mean and standard deviation values. This allowed for a comparison with respect to the 1:1 linear line. The normality of residuals is clear from Figure 3 where the normalized residual quantiles in Q-Q plot fall along a straight line.

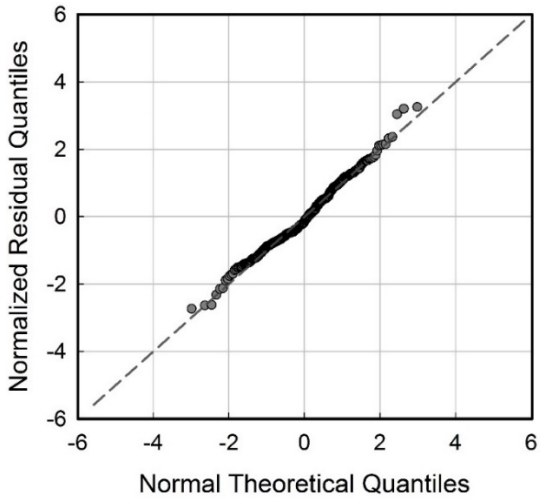

Figure 3 - Q-Q plot for fixed effect regression model.

Although CoC, RMSE, and the Q-Q plot show that the model is successful in estimating the shear wave velocity, the maximum and minimum values for residuals are significant. Figure 4 shows the residuals versus the fitted values of the model. In addition, the standardized residuals versus the leverage along with cook's distance are presented in Figure 4. Cook's distance contours show the presence of some outliers that should be eliminated from the regression model. However, for the data sets with residuals that fall in Cook's distances equal or less than 0.5, their participation in regression would not distort the accuracy of the results. The mean of the residuals is equal to zero, but the model shows a high standard deviation from the mean. This is an indicator of heavy and long tail for residuals, while the number of outliers is not significant. In addition, such a high standard deviation shows that the model might not be very successful in the estimation of V_s and the possible presence of random effects needs to be eliminated in order to obtain a better estimation. To examine this hypothesis, it is worth to study residuals of the model in terms of the difference between measured and model-calculated shear wave velocities.

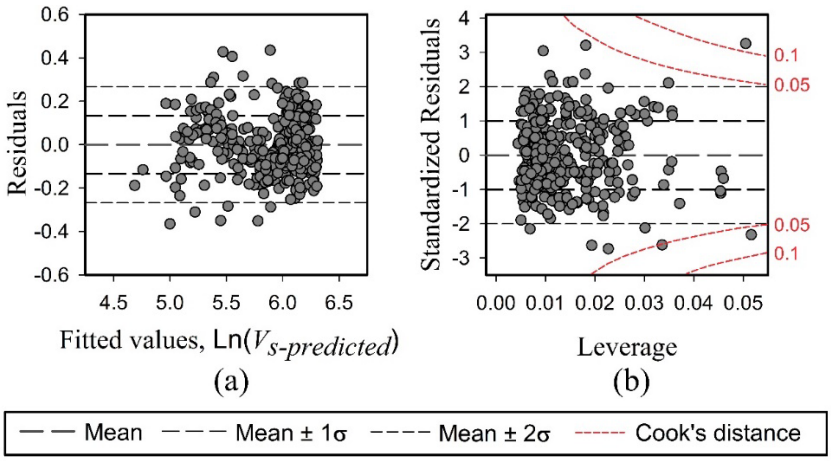

Figure 4 - (a) Residuals versus Fitted values, (b) Standardized Residuals versus Leverage along with cook's distance, for fixed effect regression model.

Figure 5 shows the histogram of the residuals for V_s (e.g., measured minus predicted). The standard deviation of the residuals is relatively high and the frequency of the residuals in the area of values higher than 50 and lower than -50 is significant. The maximum and minimum values for residuals is -101 and 196, respectively.

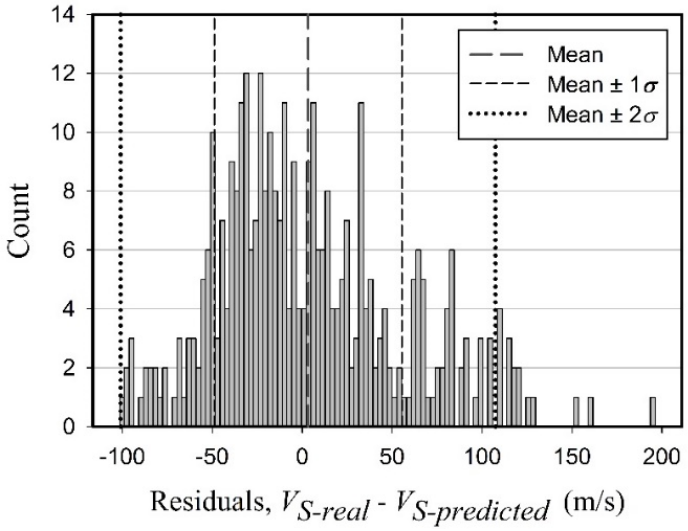

Figure 5 - Residuals frequencies for fixed effect regression model.

Although the statistical equations used for V_s estimation are known to be very rough, they are very useful in evaluating the worthwhileness of taking dynamic considerations in engineering design.

Mixed Effect Regression Model

The source of error observed in term of high residuals in the fixed effect model can be attributed to various sources. The errors can be introduced during data acquisition while performing laboratory and field tests. Plasticity indices were obtained in the laboratory and the same approach has been used to obtain *PI* values for each sample. F_c values are also less sensitive to such error source as all samples were weighted carefully, and washed with the most care to obtain the percentage of fines. The overburden pressure is not sensitive to different boreholes. The most probable source of error is the variation in test conditions during the *SPT* test. Each borehole has been drilled in different situations and weather conditions, which implies that there might be a mixed error from one borehole to another associated with to the measured *SPT N* values.

Figure 6 shows the relationship between $N_{m\text{-}60}$ and shear wave velocity for 10 boreholes chosen from the existing 71 boreholes. The trendlines are sorted from the gentlest to steepest slope. The intercept and the slope of the 10 samples were different, which indicates that there is a random effect for each borehole.

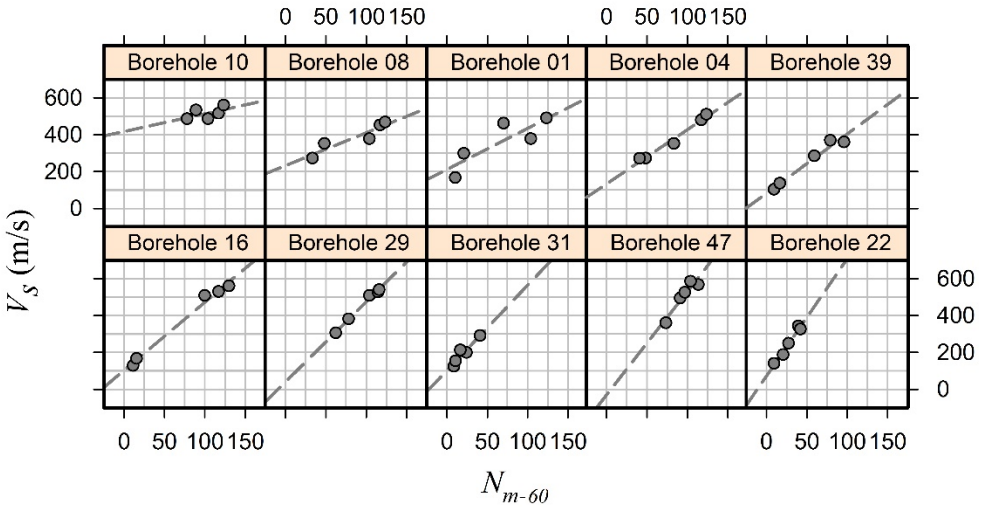

Figure 6 - Linear relationship between $N_{m\text{-}60}$ and shear wave velocity for 10 random boreholes.

Although Figure 6 shows the random effect for boreholes, the normality of the distribution for slope and intercept of the linear relationships for all boreholes should still be checked. The regression line for the j[th] borehole can be written as follows:

$$(V_s)_j = (\alpha_0)_j + (\alpha_1)_j \, N_{m-60} \tag{9}$$

Figure 7 shows the distribution of intercepts and slopes for the dataset. Although there are some outliers that are far from the mean value, a relatively normal distribution is observed here.

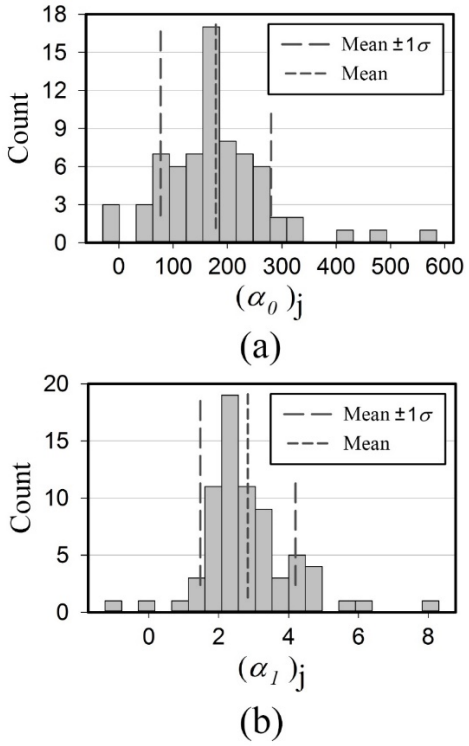

Figure 7 - (a) Distribution of the slopes and (b) Distribution of the intercepts, of the linear relationships for all boreholes.

As discussed above, the elimination of random effect due to different test conditions in each borehole can result in a better model. In order to examine the validity of this hypothesis, a mixed effect regression model is designed to consider the errors for each borehole. In this model, the *SPT N* values from each borehole are gathered into a single group. The correlation of coefficients

between different groups allows the model to consider the effect of varying *SPT* test conditions for each borehole. The statistical equation for the mixed effect regression model is as follows:

$$\ln\left(V_s\right)_{ij} = \beta_0 + \beta_N \ln\left(X_N\right)_{ij} + \beta_{PI} \ln\left(X_{PI}\right)_{ij} + \beta_{F_c} \ln\left(X_{F_c}\right)_{ij} + \beta_S \ln\left(X_S\right)_{ij} + \varepsilon_{ij} + \theta_j \quad (10)$$

In equation 10, β s are the fixed effect parameters in the regression model. ε_{ij} is the residual due to the difference between the predicted value and real value of natural logarithm of the shear wave velocity for the ith *SPT N* value in jth borehole. θ_j is the error due to the different *SPT* test condition in jth borehole, which is known as random effect and it can be presented with random effect parameters as $\theta_j = \left(\beta_{0-random}\right)_j + \left(\beta_{N-random}\right)_j \ln\left(X_N\right)$ where j is the index of the borehole.

For the sake of simplicity, the random effects are shown as the standard deviation of the error variates. Standard deviation of the random effect for intra-boring effects and inter-boring effects are assumed to be σ and τ, respectively. The *Lmer* function in the R software has been utilized to obtain the mixed regression parameters. Table 5 presents the fixed effect regression parameters and the standard deviation of the random effects obtained from the model. The standard deviation of the random intercept parameters is also included and shown as $\sigma_{\text{intercept}}$ in Table 5.

Table 5 - regression parameters for mixed effect regression model.

Fixed effects					**Random effects**		
β_0	β_N	β_{PI}	β_{F_c}	β_S	$\sigma_{\text{intercept}}$	σ_{X_n}	τ
3.83985	0.41035	0.01711	0.02852	0.05444	0.38800	0.09058	0.08058

The histogram of the residuals is shown in Figure 8. Comparing Figure 8 with the histogram of the fixed effect model residuals (i.e., Figure 5), the frequency of the residuals in the area of values higher than 50 and lower than -50 are significantly reduced. This finding keeps in mind that there may be a slight effect of boring dependency in the conducted *SPT* tests.

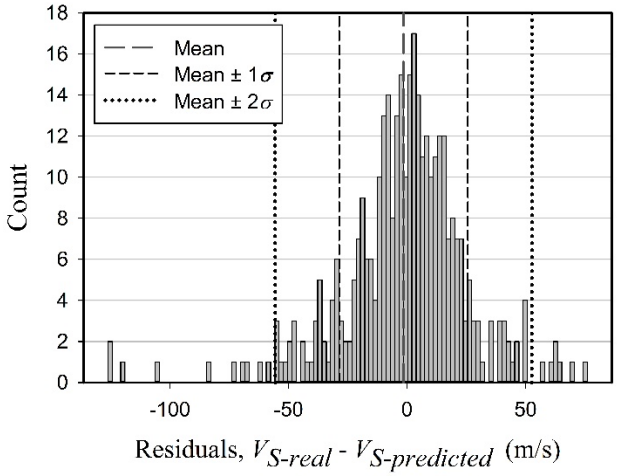

Figure 8 - Residuals frequencies for mixed effect regression model.

Figure 9 shows the Pearson residuals which are defined as ε_{ij} = ln (V_S) - [$\beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \beta_4 \ln X_4 + \theta$] divided by the square root of the variance function. The mean of the residuals is very close to zero, which indicates a reliable model fit. The standard deviation of the residuals is significantly reduced compared to the fixed effect model.

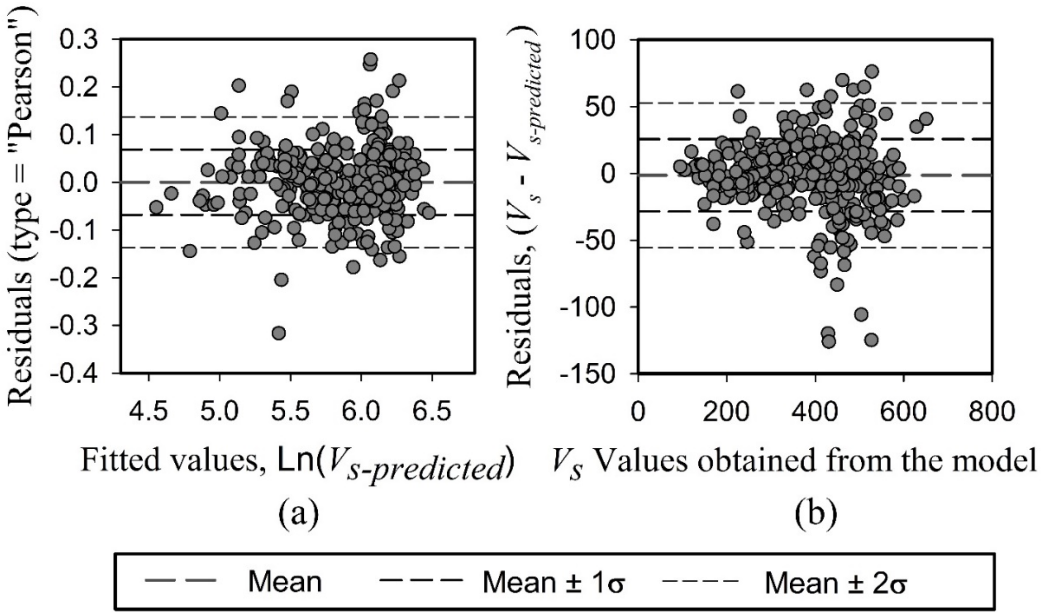

Figure 9 - Residuals versus Fitted values (a) in natural logarithmic scale and (b) normal numeric scale for mixed effect regression model.

Figure 10 shows the Q-Q plot for the mixed effect regression. Although the tails of the data deviate from the 1:1 line, the histogram of the residuals shows that only very few outliers exist in the first and forth quartile of the residuals. This indicates that the model is successful in estimating the value of V_s.

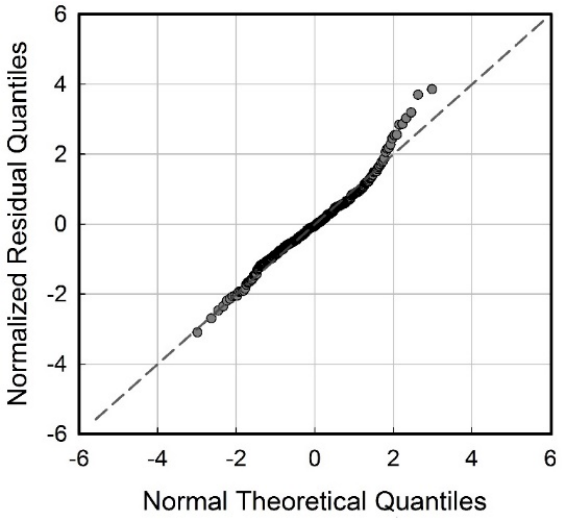

Figure 10 - Q-Q plot for mixed effect regression model.

The values of the coefficient of correlation (CoC) and root mean square error (RMSE) are shown in Table 6.

Table 6 - CoC and RMSE for mixed effect regression model.

Coefficient of correlation	**Root Mean Square Error**
0.976	27.08

Figure 11 shows the residuals versus the $N_{m\text{-}60}$, PI, F_c, and overburden pressure. It is clear from the plots that the trendline for each set of data is fairly close to zero, which means there is no bias in the data with respect to any of the input parameters in the model.

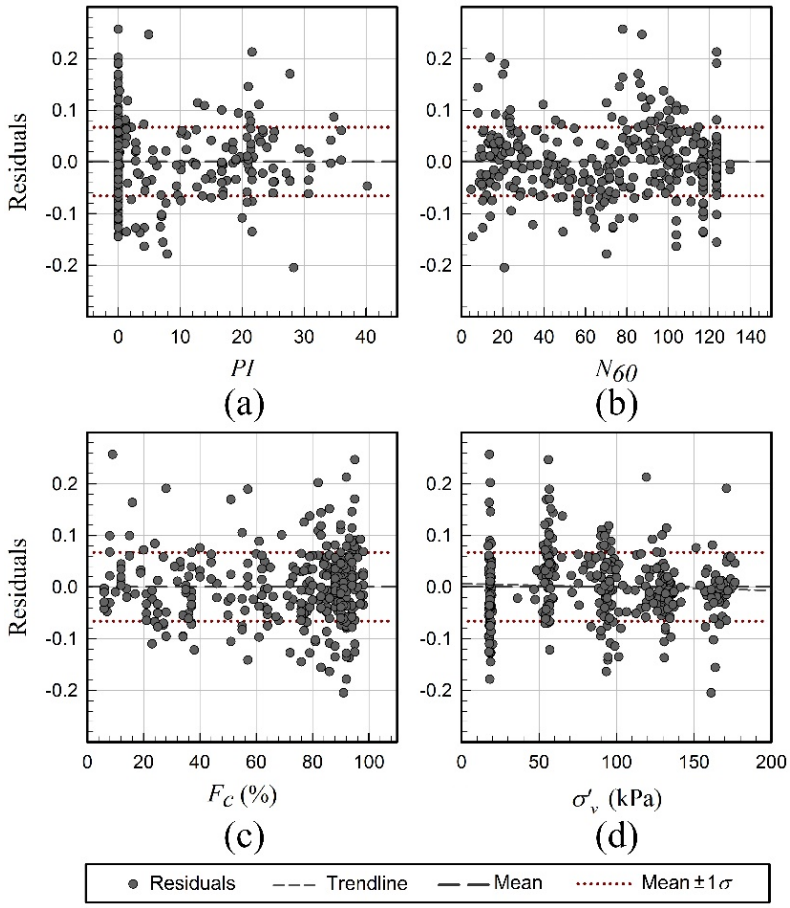

Figure 11 - Residuals versus regression variables (i.e. (a) Plasticity Index, (b) N_{60}, (c) F_c and (d) effective overburden stress) for mixed effect regression model.

In order to show the performance of the model, the results of the regression equation have been presented for three values of F_c and *PI* in Figure 12. For the first set of data, the values of F_c and *PI* are assumed to be zero. Such a soil can be categorized as sandy soil. For the second set, values of 100 and 0 percent are assigned for F_c and *PI*, respectively, which resembles a silty soil. Finally, a clayey soil is resembled by assigning a value of 100 and 15 percent for F_c and *PI* to the third set. The results of the regression equation are provided in Figure 12 for these three sets along with the data set values for $N_{m\text{-}60}$ and overburden pressure, σ_v'.

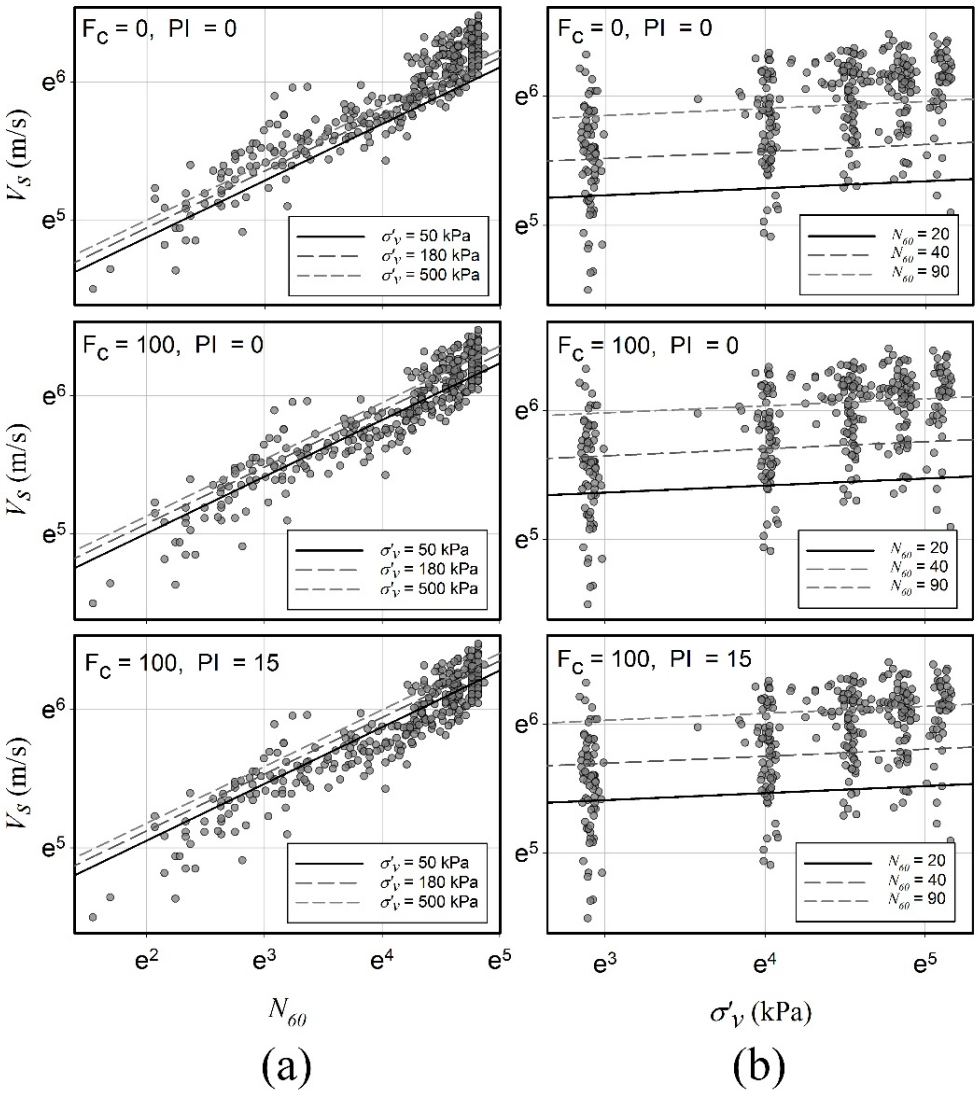

Figure 12- Shear wave velocity obtained from regression equation versus (a) N_{60} and (b) effective overburden stress, for three different sets of F_c and *PI* values.

The value of the shear wave velocity has been shown with respect to both $N_{m\text{-}60}$ and *PI*. For V_s versus $N_{m\text{-}60}$ plots, regression equations have been shown for three values of effective overburden pressure, being 50, 180, and 500 kPa. Also for V_s versus σ_v' plots, regression equations have been shown for three values of *SPT* numbers, being 20, 40, and 90. The marginal change for *N* value is significantly higher than that of effective overburden pressure, which means that the *N* value is still the most important governing variable in the estimation of the shear wave velocity.

The results show that as the plasticity index and fine content of the soil increases, the shear wave velocity increases. Although the marginal changes for both F_c and PI parameters are observed to be insignificant, the shear wave velocity undergoes a significant change when F_c and PI varies in the range of engineering practice.

In order to analyze the effect of each variable in the model, the Analysis of Variance (ANOVA) table is presented in Table 7.

Table 7 - Analysis of variance (ANOVA) for mixed effect regression model.

Variable	Chisq	Df	P-value	Significance Level
$\ln(N_{m-60})$	663.779	1	< 2.2e-16	***
$\ln(PI+1)$	8.7485	1	0.003099	**
$\ln(F_c+1)$	6.661	1	0.009855	**
$\ln(\sigma_v')$	32.8884	1	9.76E-09	***
***, **, * indicate significance levels of 0.001, 0.01, and 0.05, respectively.				

From p-values observed in Table 7, the coefficient of N_{m-60} and σ_v', are highly significant (with $p<0.0001$), suggesting there is considerable effect for both SPT values and overburden pressure in the estimation of shear wave velocity for a given soil. The coefficient of PI and F_c is significant and positive with p-values both less than 0.01, which indicates that these values cannot be omitted from the model. Generally, the effect of N_{m-60} values are observed to be superior to the effect of other variables regardless of the type of soil. Effective overburden pressure is the second most influential parameter in the estimation of the shear wave velocity. Plasticity index and fine content of the soil are the third and fourth significant parameters, which affect the shear moduli and shear wave velocity of the soil.

Table 8 compares the result of the proposed model with three models for estimation of shear wave velocity. The Coefficient of Correlation and Root Mean Square Error are significantly improved for the proposed model. It should be noted that since Brandenberg et al. [38]obtained

three different relationships for three different soils, for each soil type the associated relationship is used.

Table **8** - Comparison of the results of the model obtained in this study with the model.

	Coefficient of correlation	Root Mean Square Error
Seed and Idriss[32]	0.87	129.84
Hasançebi and Ulusay[25]	0.87	98.48
Dikmen[24]	0.87	116.65
Brandenberg et al[38]	0.75	153.82
This study	0.98	27.08

Conclusions

Statistical regression models were used to correlate the shear wave velocity of soils to common geotechnical parameters, namely the standard penetration test resistance, vertical effective overburden pressure, fines content, and plasticity index. Through correlating these parameters, the ultimate goal of obtaining a single equation that estimates the shear wave velocity for different types of soil was achieved. Downhole, standard penetration, Atterberg limit tests, and sieve analysis were performed to obtain the geotechnical parameters required, as input, for regression models. First, a fixed effect model was used, the results of which showed a high standard deviation for the residuals of the model. Then, a random effect model was employed to remove inter-borehole random errors. The distribution of residuals for the random effect model showed a very low standard deviation. The effect of each parameter was examined using statistical tools (i.e., analysis of variance). It was found out that the *SPT* number, $N_{m\text{-}60}$ is the most influential parameter in the estimation of shear wave velocity. Overburden pressure (S) was found to be the second influential parameter in V_s estimation. For the specimens tested in this study, the effects of plasticity index (PI) and fines content (F_c) on the shear wave velocity obtained were found to be low. This was because the variation of these parameters was marginal for the studied region. In

the larger context of geotechnical practice, however, these parameters vary highly across different soil types and that renders them influential in the V_s estimation. It should be noted that the statistical equation obtained in this study should not be used for sites where the exact shear wave velocity of the soil was obtained using geophysical tests, such as downhole and Crosshole tests. It is recommended to limit the utilization of the proposed equation to preliminary V_s estimation and in the evaluation of the need for seismic consideration in design.

References

1. Seed, H.B. and Idriss, I.M., *Soil Moduli and Damping Factors for Dynamic Response Analyses*. 1970, Earthquake Engineering Research Center. University of California,: Berkeley, CA.
2. Stokoe II, K.H. and Lodde, P.F. *Dynamic response of SanFrancisco Bay mud*. in *American Society of Civil Engineers Conf. on Earthquake Engineering and Soil Dynamics*. 1978. Reston,VA: ASCE.
3. Sun, J.I., Golersorki, R., and Seed, H.B., *Dynamic moduli and damping ratios for cohesive soils*, in *Ref. No. UCB/EERC-88/15.* . 1988: College of Engineering, University of California at Berkeley, USA.
4. Vucetic, M. and Dobry, R., *Effect of soil plasticity on cyclic response.* Journal of Geotechnical Engineering, 1991. **117**(1): p. 89-107.
5. Abrahamson, N. and Silva, W., *Summary of the Abrahamson & Silva NGA Ground-Motion Relations.* Earthquake Spectra, 2008. **24**(1): p. 67-97.
6. Boore, D.M. and Atkinson, G.M., *Ground-Motion Prediction Equations for the Average Horizontal Component of PGA, PGV, and 5%-Damped PSA at Spectral Periods between 0.01 s and 10.0 s.* Earthquake Spectra, 2008. **24**(1): p. 99-138.

7. Campbell, K.W. and Bozorgnia, Y., *NGA Ground Motion Model for the Geometric Mean Horizontal Component of PGA, PGV, PGD and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s.* Earthquake Spectra, 2008. **24**(1): p. 139-171.
8. Chiou, B.-J. and Youngs, R.R., *An NGA Model for the Average Horizontal Component of Peak Ground Motion and Response Spectra.* Earthquake Spectra, 2008. **24**(1): p. 173-215.
9. Idriss, I.M., *An NGA Empirical Model for Estimating the Horizontal Spectral Values Generated By Shallow Crustal Earthquakes.* Earthquake Spectra, 2008. **24**(1): p. 217-242.
10. Choi, Y. and Stewart, J.P., *Nonlinear site amplification as function of 30 m shear wave velocity.* Earthq spectra, 2005. **21**(1): p. 1-30.
11. Boore, D.M., *Estimating Vs(30) (or NEHRP Site Classes) from Shallow Velocity Models (Depths < 30 m)* Bulletin of the Seismological Society of America, 2004. **94**(2): p. 591-597.
12. Isenhower, W.M., *Torsional Simple Shear/Resonant Column Properties of San Francisco Bay Mud* Geotechnical Engineering Thesis, 1979. **GT 80-1, University of Texas, Austin, TX.**
13. D'Elia, B. and Lanzo, G. *'Laboratory and field determinations of small-strain shear modulus of natural soil deposits*. in *Proc. of the 11th World Conference on Earthquake Engineering*. 1996.
14. Khosravi, A. and McCartney, J., *Resonant Column Test for Unsaturated Soils With Suction–Saturation Control.* 2011.
15. Walton-Macaulay, C., Bryson, L.S., Hippley, B.T., and Hardin, B.O., *Uniqueness of a Constitutive Shear Modulus Surface for Unsaturated Soils.* International Journal of Geomechanics, 2015. **15**(6).

16. Brignoli, E. and Gotti, M., *Misure della velocità di onde elastiche di taglio in laboratorio con l'impiego di trasduttori piezoelettrici.* Rivista Italiana di Geotecnica, 1992. **1**: p. 5-16.
17. Ghorbani, A., Jafarian, Y., and Maghsoudi, M.S., *Estimating shear wave velocity of soil deposits using polynomial neural networks: Application to liquefaction.* Comput Geosci, 2012. **44**: p. 86-94.
18. ASTM D4428/D4428M-00, *Standard Test Methods for Crosshole Seismic Testing.* ASTM International, 2014.
19. ASTM Standard D7400, *Standard Test Methods for Downhole Seismic Testing*. 2014, ASTM International: West Conshohocken.
20. Campanella, R.G., Robertson, P.K., and Gillespie, D., *Cone Penetration Testing in Deltaic Soils.* Canadian Geotechnical Journal, 1983. **20**: p. 23.
21. Robertson, P.K., Campanella, R.G., Gillespie, D., and Rice, A., *Seismic Cpt to Measure in Situ Shear Wave Velocity.* Journal of Geotechnical Engineering, 1986. **112**(8): p. 791-803.
22. Park, C.B., Miller, R.D., and Xia, J., *Multichannel analysis of surface waves.* GEOPHYSICS, 1999. **64**(3): p. 800-808.
23. ASTM Standard D6758-02, *Standard Test Method for Measuring Stiffness and Apparent Modulus of Soil and Soil-Aggregate In-Place by an Electro-Mechanical Method*, in *ASTM International*. 2002.
24. Dikmen, Ü., *Statistical correlations of shear wave velocity and penetration resistance for soils.* J Geophys Eng, 2009. **6**(1): p. 61-72.
25. Hasancebi, N. and Ulusay, R., *Empirical correlations between shear wave velocity and penetration resistance for ground shaking assessments.* Bull Eng Geol Environ, 2007. **66**(2): p. 203-13.

26. Imai, T. and Yoshimura, M., *Elastic wave velocities and characteristics of soft soil deposits.* Soil Mechanics and Foundation Engineering, The Japanese Society of Soil Mechanics and Foundation engineering, 1970. **18**(1).
27. Imai, T. and Yoshimura, M., *The relation of mechanical properties of soils to P and S wave velocities for soil ground in Japan*. 1976, Urana Research Institute, OYO Corporation: Japan.
28. Iyisan, R., *Correlations between shear wave velocity and in-situ penetration test results.* Technical journal of Turkish Chamber of Civil Engineers, 1996. **7**: p. 371-374.
29. Jinan, Z., *Correlation between seismic wave velocity and the number of blow of SPT and Depths.* Chin J Geotech Eng (ASCE), 1987: p. 92-100.
30. Kuo, C., Wen, K., Hsieh, H.H., Chang, T.M., Lin, C.M., and Chen, C.T., *Evaluating empirical regression equations for Vs and estimating Vs30 in northeastern Taiwan.* Soil Dynamics and Earthquake Engineering, 2011. **31**(3): p. 431-439.
31. Ohba. S and Toriumi, I. *Dynamic response characteristics of Osaka Plain*. in *In: Proceedings of the annual meeting, AIJ (in Japanese)*. 1970.
32. Seed, H.B. and Idriss, I.M., *Evaluation of liquefaction potential sand deposits based on observation of performance in previous earthquakes*, in *ASCE National Convention*. 1981: Missouri.
33. Seed, H.B., Idriss, I.M., and Arango, I., *Evaluation of liquefaction potential using field performance data.* J Geotech Eng, 1983. **109**(3): p. 458-82.
34. Hara, A., *Research on dynamic characteristics by dynamic triaxial tests, part 2.* Technical Meeting of Architectural Institute of Japan (In Japanese), 1970.

35. Kanai, K., Tanaka, T., Morishita, T., and Osada, K., *Observation of microtremors, XI. Matsushiro earthquake swarm areas.* Bulletin of Earthquake Research Institute, XLIV, Part 3, University of Tokyo, 1966.
36. Lee, S.H., *Regression models of shear wave velocities in Taipei basin.* Journal of the Chinese Institute of Engineers, 1990. **13**(5): p. 519-532.
37. Anbazhagan, P. and Sitharam, T., *Site characterization and site response studies using shear wave velocity.* J. Seismol. Earthquake Eng. , 2008. **10**(2): p. 53-67.
38. Brandenberg, S., Bellana, N., and Shantz, T., *Shear wave velocity as a statistical function of standard penetration test resistance and vertical effective stress at Caltrans bridge sites.* Soil Dynamics and Earthquake Engineering, 2010. **30**(10): p. 1026-1035.
39. Ohta, Y., Goto, N., Kagami, H., and Shiono, K., *Shear wave velocity measurement during a standard penetration test.* Earthquake Engineering & Structural Dynamics, 1978. **6**(1): p. 43-50.
40. Darendeli, M.B., *Development of a new family of normalized modulus reduction and material damping curves*. 2001, University of Texas at Austin: Austin, Texas.
41. Hardin, B.O. and Drnevich, V.P., *Shear modulus and damping in soils: measurement and parameter effects (terzaghi leture).* J Soil Mech Found Div, 1972. **98**(6): p. 603-24.
42. Kim, T. and Novak, M., *Dynamic properties of some cohesive soils of Ontario.* Canadian Geotechnical Journal, 1981. **18**: p. 371-389.
43. Rampello, S. and Viggiani, G. *Panel discussion: The dependence of G_o on stress state and history in cohesive soils*. in *Prefailure Deformation of Geomaterials*. 1995. Balkema, Rotterdam.

44. Carlton, B.D. and Pestana, J.M. *Small strain shear modulus of high and low plasticity clays and silts*. in *Proc. 15th World Conference on Earthquake Engineering*. 2012.
45. Dobry, R. and Vucetic, M., *Dynamic properties and seismic response of soft clay deposits*. 1988, Department of Civil Engineering, Rensselaer Polytechnic Institute.
46. Abrahamson, N.A. and Youngs, R.R., *A stable algorithm for regression analyses using the random effects model.* Bulletin of the Seismological Society of America, 1992. **82(1)**: p. 505-510.
47. Yaghmaei-Sabegh, S., Shoghian, Z., and Neaz Sheikh, M., *A new model for the prediction of earthquake ground-motion duration in Iran.* Natural Hazards, 2014. **70**(1): p. 69-92.
48. Jayaram, N. and Baker, J.W., *Considering spatial correlation in mixed-effects regression and the impact on ground-motion models.* Bulletin of the Seismological Society of America, 2010. **100(6)**: p. 3295-3303.
49. Takahashi, T., Asano, A., Saiki, T., Okada, H., Irikura, K., Zhao, J.X., Zhang, J., Thio, H.K., Somerville, P.G., Fukushima, Y., and Fukushima, Y. *Attenuation models for response spectra derived from japanese strong-motion records accounting for tectonic source types*. in *13th World Conference on Earthquake Engineering*. 2004. Vancouver, Canada.
50. Özbey, C., Sari, A., Manuel, L., Erdik, M., and Fahjan, Y., *An empirical attenuation relationship for Northwestern Turkey ground motion using a random effects approach.* Soil Dynamics and Earthquake Engineering, 2004. **24**(2): p. 115-125.
51. S.K., S. and Baker, J.W., *Regression Models for Predicting the Probability of Near-Fault Earthquake Ground Motion Pulses, and their Period*, in *11th International Conference on Applications of Statistics and Probability in Civil Engineering*. 2011: Zurich, Switzerland.

52. Skempton, A., *Standard penetration test procedures.* Geotechnique, 1986. **36**(3): p. 425-557.
53. Sykora, D., *Creation of a data base of seismic shear wave velocities for correlation analysis: ,Vicksburg, Mississippi, U.S. Army Engineer Waterways Experiment Station.* Geotechnical Laboratory Miscellaneous Paper, 1987: p. GL-87-26, 114.