Soil Velocity Models Informed by Remote Sensing and Artificial Intelligence

Qinlin Yu

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science in Civil Engineering

University of Washington

2021

Committee:

Brett Maurer

Michael Gomez

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Soil Velocity Models Informed by Remote Sensing and Artificial Intelligence

Qinlin Yu

Chair of the Supervisory Committee:
Brett Maurer
Civil and Environmental Engineering

Near-surface soil conditions can significantly alter the amplitude, duration, and frequency content of incoming ground motions – often with profound consequences for the built environment – and are thus important inputs to any ground-motion prediction. In current practice, the shear-wave velocity (Vs) time-averaged over the upper 30 m (Vs30) is widely used as a proxy for site effects, forming the basis of seismic site class and underpinning site-amplification factors in empirical ground-motion models (GMMs). Similarly, wave-propagation based GMMs rely on depth-continuous models of Vs, at least to where hard rock is found. Consequently, earthquake simulations rely on knowledge of either Vs30 or Vs-versus-depth (referred to herein as a soil velocity model, SVM), depending on which type of GMM is adopted. In either case, the need for these inputs at regional scale presents a challenge, given the infeasibility of subsurface testing over vast areas. At present, a patchwork of Vs30 models exists in the U.S., with the USGS National "baseline" model being a regression equation based on one parameter – topographic slope. Several regional models attempt to improve upon this baseline, generally by incorporating mapped geology into similar regression equations. Likewise, a sparse collection of SVMs, commonly called "Community Velocity Models," exists in the U.S. These generally: (i) are available only in select urban regions; (ii) provide predictions with low spatial resolution; and (iii) focus more on deep geologic structure and less on near-surface conditions, even though the latter could alter motions most significantly. Given the growth of community geophysical and geotechnical datasets, satellite remote sensing, and artificial intelligence, unified and more accurate solutions, commensurate with the sophistication of emergent ground-motion models, are conceivable. Accordingly, this research develops U.S. National Vs30 and SVM solutions using machine- and deep-learning (ML/DL) techniques, wherein geospatial variables are used to predict subsurface wave velocities. These velocities lack theoretical links to above-ground parameters, but correlate in complex, interconnected ways – a prime problem for ML/DL. The resulting, trained models are tested against existing models using unbiased tests and shown to provide efficient predictions that match or best existing models. Ultimately, the proposed approach could be expanded using additional training data and new predictor variables, which were each relatively modest in the present study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

[The acknowledgement section goes here. The acknowledge section does not go into the table of contents.]

DEDICATION

[This page is optional and does not go into the Table of Contents]

Chapter 1. INTRODUCTION

Subsurface seismic-wave velocities (e.g., shear-wave velocity, Vs), have a significant effect on ground-motion characteristics. Measurements or estimates of these velocities are thus needed to predict ground motions and, in-turn, coseismic phenomena such as liquefaction, landslides, infrastructure damage, economic loss, etc. Ideally, these velocities would be obtainable: (i) rapidly (i.e., in a relatively time- and cost-efficient manner); (ii) at high resolution (e.g., commensurate with the scale at which subsurface velocities change); and (iii) over the regional extents that experience strong ground-motion during earthquakes (e.g., those of cities and transportation networks). Problematically, state-of-practice methods for measuring Vs rely on either downhole, direct measurements (which are more accurate, but also more expensive) or non-invasive, indirect measurements via surface waves (which are less expensive, but also less accurate). Regardless of which is used, both collect 1D profiles and require considerable time and cost. As a result, it is simply infeasible to perform Vs measurements over vast areas, as would be required for regional earthquake simulations. Even in cases where Vs is needed at discrete sites for important purposes (e.g., at seismic-recording stations, to develop empirical ground motion models, or GMMs), it is often the case that Vs is estimated, rather than measured (e.g., Ahdi et al. 2017), due to the cost.

Accordingly, efforts have been made to predict Vs profiles in the absence of direct measurement (e.g., Boore and Joyner, 1997; Holzer et al., 2005; Allen and Wald, 2007; Castellaro et al., 2008; Boore et al., 2011; Boore, 2016; Parker et al., 2017). These efforts have focused somewhat more on predicting the time-averaged VS in the upper 30 m (VS30), which: (i) is widely used as a proxy for site effects; (ii) forms the current basis of seismic site class; (iii) underpins site-amplification functions (e.g., Stewart et al., 2017); and (iv) is a required input to empirical GMMs.

By comparison, efforts have focused somewhat less on explicitly predicting the variation of VS versus depth, which: (i) is a required input to ground-motion prediction methods based on wave propagation (e.g., Schnabel et al., 1972); (ii) is increasingly needed, given the growth and utility of physics-based simulations (e.g., among many, Ramirez-Guzman et al., 2015; Bradley et al., 2017; Frankel et al., 2018; Wirth et al., 2018); and (iii) relative to VS30, is more difficult to predict. Collectively, Vs30 and Vs-versus-depth play important roles in: post-earthquake data products, such as those from the United States Geological Survey (USGS) (Worden et al., 2010); amplification functions; scenario-earthquake studies; seismic hazard analyses (e.g., Crowley et al., 2019); and ultimately, the National Seismic Hazard Model (Petersen et al., 2019).

1.1 PREDICTION OF VS30

With respect to Vs_{30}, a patchwork of models presently exists in the U.S., with the USGS National "baseline" model being a regression equation based on one parameter – topographic slope (Wald and Allen, 2007; Allen and Wald, 2009). The underlying, seminal concept – that flat ground tends to be soft and steep ground tends to be hard – is quite effective, but also unabashedly simple. Several regional models attempt to improve upon this baseline, generally by using higher-resolution digital elevation models, more advanced statistical/interpolation schemes, and/or by binning the data and regression equations on mapped geologic units (e.g., Wills and Clahan, 2006; Thompson et al., 2014; Ahdi et al., 2017). At present, four U.S. regional Vs_{30} models are adopted by the USGS, covering all or part of five Western states, as shown in Figure 1.1. One or more other regional solutions exist but were not adopted because they were not digital and otherwise not easy to implement. While there is no mechanistic necessity for Vs_{30} models to be region-specific, the historical pace at which measurements are made, in combination with the tedium of regression

modelling, has led to the current patchwork of regional models overlaying a baseline. In some states, for example, there is, or at least formerly was, a complete paucity of Vs_{30} measurements.

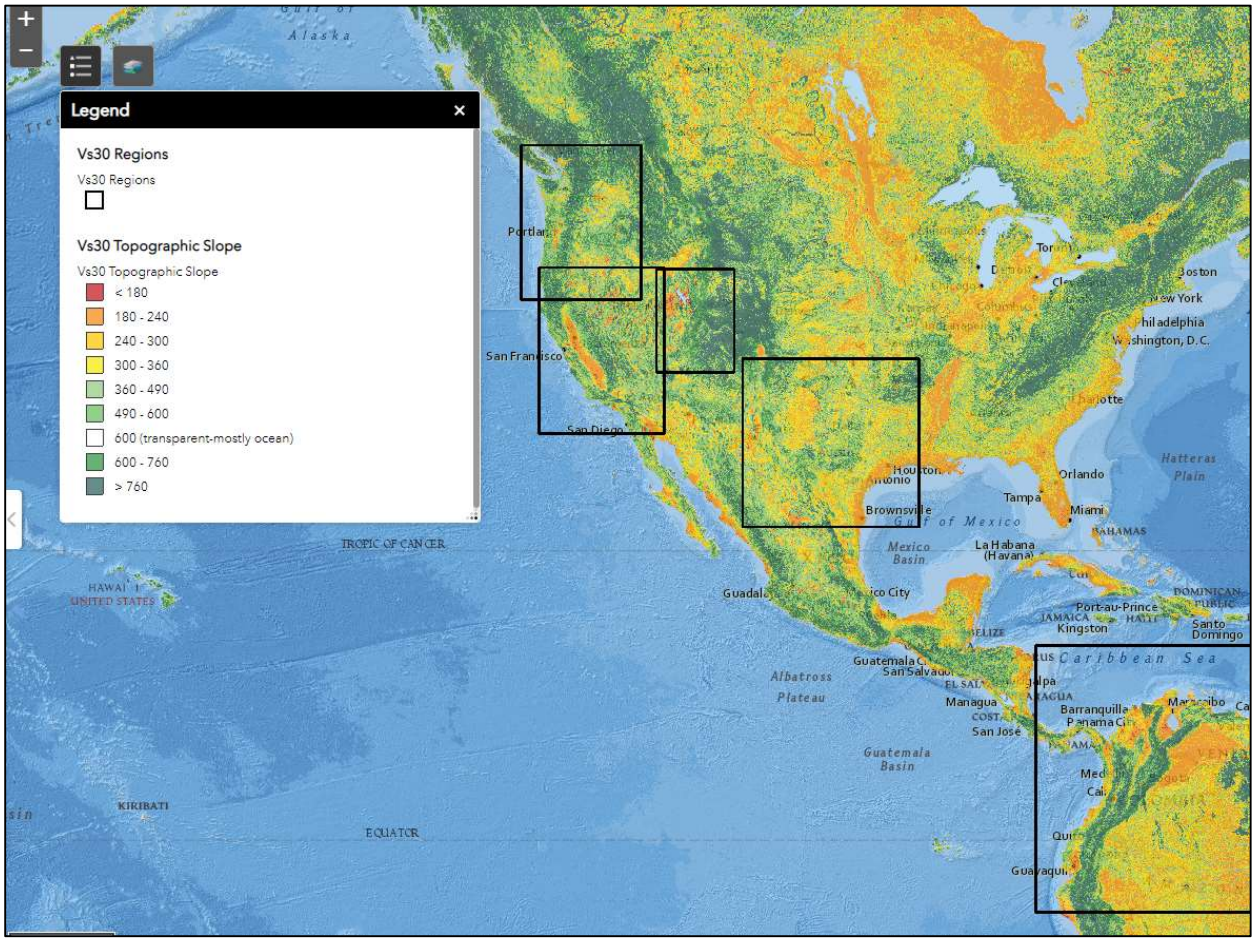

Figure 1.1 Mosaic Vs30 map of the continental United States. Boxes outline regional maps inserted over the topographic slope-based background map (as described in Heath et al., 2020).

1.2 Prediction of Vs versus Depth

Similar to Vs_{30}, a patchwork of *Vs*-versus-depth models, referred to herein as a "soil velocity model" (SVM), presently exists in the U.S. These are also commonly called "Community Velocity Models," owing to their community scale (e.g., that of an urban region) and/or to pay homage to the community of individuals who collected the data used in the model. SVMs are necessary for physics-based simulations, which are especially useful in regions lacking historical records (e.g.,

where the earthquake return-period exceeds the observation interval), examples of which include the Alpine Fault Zone of New Zealand (Bradley et al., 2017), the Central-Eastern U.S. (Ramirez-Guzman et al., 2015), and the Cascadia Subduction Zone of North America (Frankel et al., 2018; Wirth et al., 2018). In addition to populating gaps in ground-motion datasets, physics-based simulations help to elucidate and quantify complex ground-motion phenomena (e.g., the effects of directivity, basins, and topography) via explicit modeling of kinematic fault rupture, wave propagation, and the subsurface velocity structure.

With respect to the latter, U.S. SVMs include Magistrale et al. (2009) (Wasatch Valley, Utah), Small et al. (2017) (Southern California; among others in the state); Stephenson et al. (2017) (Oregon and Washington, west of the Cascade Range); and Cramer et al. (2016) (St. Louis), among a limited number of others. In general, these models are: (i) available only in select urban areas of the U.S.; (ii) have coarse spatial resolution; and (iii) are concerned more with predicting deep geologic structure, such as basin geometry, and less with near-surface conditions that can also significantly alter motions. The Stephenson et al. (2017) model of the Cascadia Subduction Zone, for example, has a minimum Vs of 600 m/s. Thus, while this model was utilized in recent simulations (Frankel et al. 2018, Wirth et al. 2018), it does not consider the presence and influence of surface conditions. A unified National SVM would serve as a "Baseline" à la the USGS Vs_{30} product and allow ground-motion simulations to be performed nationally.

Efforts to develop solutions that are more generalized and broadly applicable include Boore and Joyner (1997) (revisited by Boore, 2016), who proposed a Vs-versus-depth model for rock sites, applicable to the Western U.S. This model, shown in Figure 1.2, adopts a simple power-law form between Vs and depth, constrained such that Vs_{30} is that of a rock site. While this model has

been used in many applications, it does not tackle the more difficult problem of predicting the depth of rock or the *Vs*-profile of material overlaying rock.

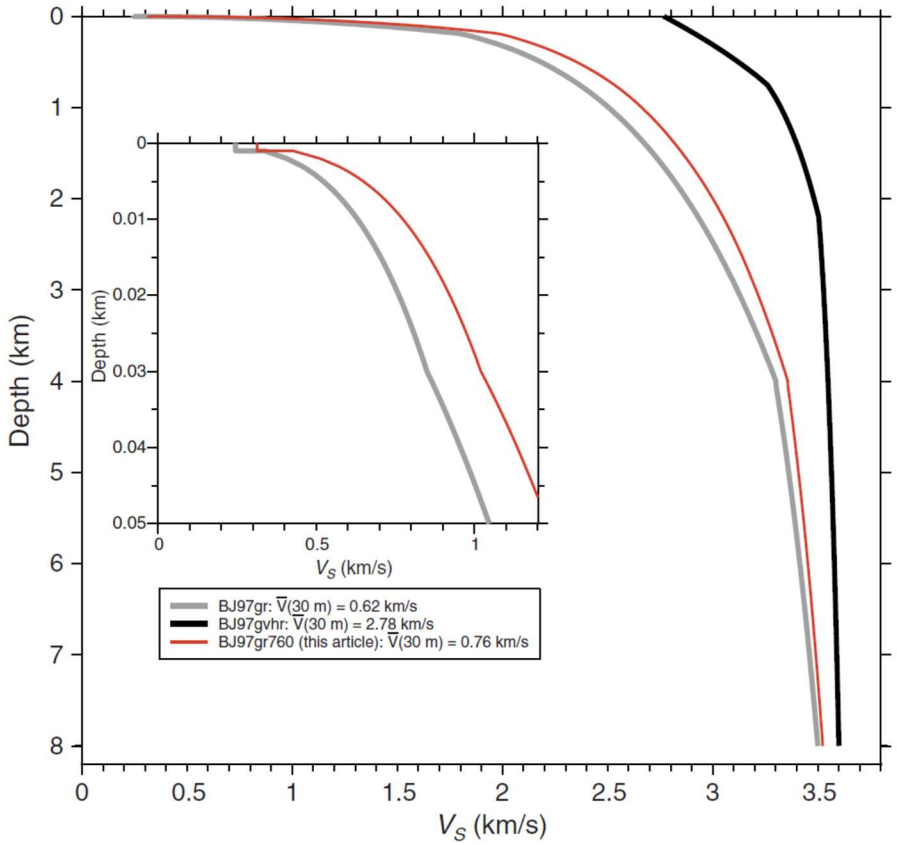

Figure 1.2 Vs as a function of depth as modeled by Boore and Joyner (1997) and revisited by Boore (2016). Figure after Boore (2016).

More recently, Shi and Asimaki (2018) proposed a California SVM conditioned on V_{S30} and based on the functional form of Vrettos (1996). Provided a prediction of V_{S30}, it predicts V_S as a function of depth (z):

$$V_s(z) = \begin{cases} V_{S0} & , \ z < 2.5 \text{ m} \\ V_{S0}\left(1 + k(z - 2.5)\right)^{\frac{1}{n}}, & z \geq 2.5 \text{ m} \end{cases} \quad (1)$$

where V_{S0}, k, and n are fitting parameters that are predicted solely as a function of V_{S30}, as defined in Shi and Asimaki (2018). This model: (i) provides depth-continuous predictions of near-surface

V_S; (ii) can be implemented at relatively fine spatial resolution (i.e., the resolution of the input V_{S30}); and (iii) was fit to a set of measured profiles from California. A depiction of the model's fit to the training data compiled by Shi and Asimaki (2018) is shown in Figure 1.3.

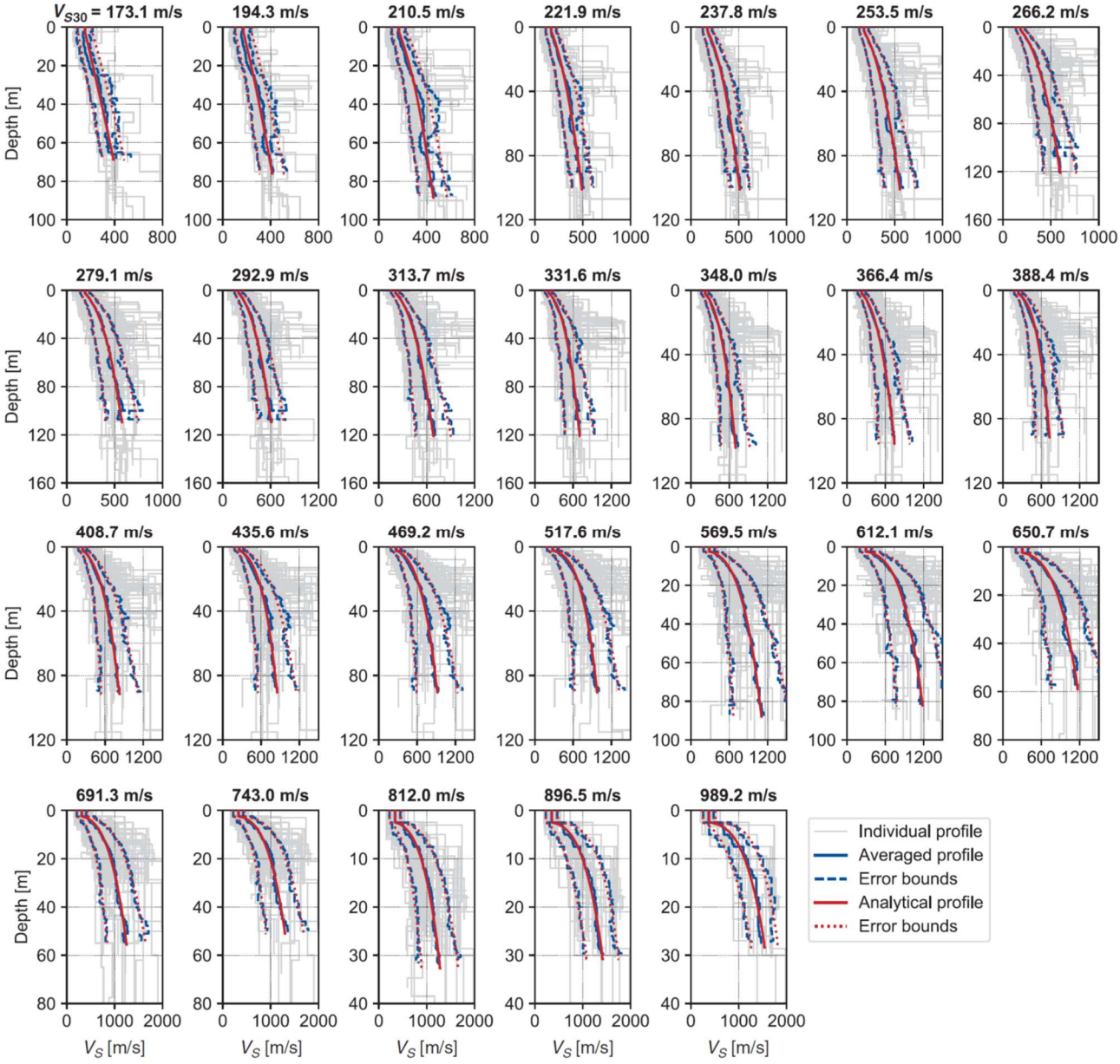

Figure 1.3 Vs as a function of depth as modeled by Shi and Asimaki (2018) for different bins of Vs30. Individual training profiles are in gray; the model prediction appears in red. Figure from Shi and Asimaki (2018).

While the Shi and Asimaki (2018) SVM represents a significant conceptual advance, it still suffers from several limitations. *First*, it was trained only on data from California, so it is not nationally applicable. *Second*, it relies on existing methods to predict V_{S30} (the only model input), which are themselves quite uncertain. Third, it assumes that all profiles of a given V_{S30} have the same velocity profile, which is to assume that V_{S30} is sufficient for capturing the V_S-depth relationship. *Fourth*, and closely related to the preceding, it relies on a separate method to predict the depth and velocity of rock, given that its predictions imply the presence of soil to very large depths. That is, it must be supplanted by some other information, such as the predicted depth to a V_S indicative of rock, such as 760 m/s, 1 km/s, or some other value, x ($Z_{0.76}$, $Z_{1.0}$ or Z_x), which could be used to override the SVM at depth. This could be obtained from an existing community velocity model, although such models are sparse, or from an empirical V_{S30}-Z_x correlation (e.g., Abrahamson and Silva, 2008; Chiou and Youngs, 2008; Campbell and Bozorgnia, 2007), although such correlations are notoriously uncertain. This is because a deposit of soft soil can be thin or thick, independent of the underlaying rock, so there is little reason for V_{S30} and Z_x to be closely related when V_{S30} is low.

Considering these limitations, Marafi et al. (2021) studied the suitability of the Shi and Asimaki (2018) SVM in the Cascadia Region of Washington, Oregon, and California using 909 V_S profiles compiled by Ahdi et al. (2017). Figure 1.4 quantifies the accuracy of predictions in terms of the difference between the natural-log of the predicted V_S ($\ln V_{S,SA18}$) and the natural-log of the measured V_S ($\ln V_S$), termed the V_S residual.

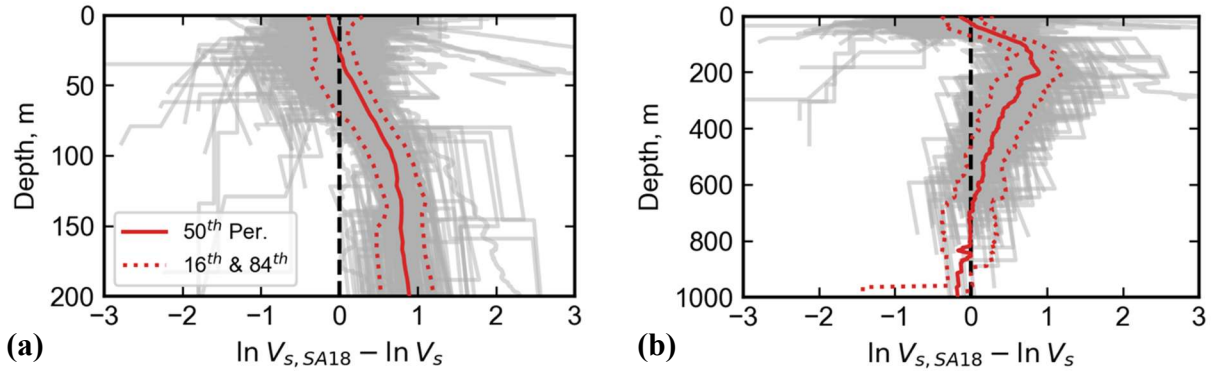

Figure 1.4 Vs residuals (predictions from Shi and Asimaki, 2018) versus depth for 909 profiles in Cascadia, for depths up to: (a) 200 m; and (b) 1000 m. After Marafi et al. (2018).

As shown in Figure 1.4, the medians of V_S residuals exhibit strong bias between depths of 50 m and 600 m, reaching a peak value of +0.84 (~132% overprediction of V_S). In particular, the Shi and Asimaki (2018) model tends to severely overpredict V_S in deeper sedimentary basins, which are either not found in California or were poorly represented in the California training set. In addition, the variance of residuals indicates that even at depths where the median residual is zero, 32% of V_S profiles are either under- or overpredicted by at least 35%. It should be emphasized that this performance assumes V_{S30}, *the only model input,* is *perfectly predicted* (i.e., the actual profile V_{S30} was used as the input). Hence, Figure 1.4 does not convey the true error that would occur in a forward prediction, wherein a V_{S30}-model would be used to predict V_{S30}. Consequently, if the Shi and Asimaki (2018) SVM was employed within a ground-motion simulation in Cascadia, the resulting predictions would, in many locations, be erroneous for a range of spectral periods.

In summary, a unified National SVM to predict near-surface V_S-versus-depth, which is needed for physics-based ground motion simulations, does not exist. A unified National V_{S30} model, which is used to develop and employ empirical GMMs, does exist but is a one-parameter equation.

1.3 A New Approach to Model Development

Given the growth of community geophysical/geotechnical datasets, satellite remote sensing, and artificial intelligence, more advanced and accurate solutions could be achieved, commensurate with the sophistication of existing and emergent ground-motion models. Accordingly, this research develops U.S. National Vs30 and SVM solutions using machine- and deep-learning (ML/DL), wherein geospatial variables will be used to predict subsurface Vs, as trained on a new National dataset compiled herein. Examples of geospatial predictor variables, which are obtained from satellite remote-sensing and existing, mapped information, include surface slope, mineralogy, roughness, wetness, and reflectance; distance to and elevation above rivers, streams, and other water bodies; and various values describing geology, geomorphology, bedrock and water depth, hydrology, lithology, etc. While the basic concept of a geospatial Vs model is not new – the current National Vs30 model, with one parameter, might be described as such – neither algorithmic learning nor a larger number of predictor variables have previously been used. While accurate prediction of subsurface traits surely requires more than one variable, regression approaches require hypotheses of what is believed to matter and how, limiting the number of variables easily modelled. Because such beliefs are unnecessary with ML/DL, it can provide insights that are simply unlikely, if not altogether infeasible, with regression. The proposed approach thus allows for a very large body of geospatial information to be used, with greater potential for these variables/data to be exploited fully. This thesis results in: (i) a new dataset of 2737 Vs profiles in the United States compiled from numerous existing publications, with many profiles requiring digitization; (ii) a new National "baseline" VS30 model driven by ML/DL; and (iii) building off the VS30 solution, a new National "baseline" SVM to predict VS-versus-depth, wherein a

mechanistic backbone is used to ensure physically sensible behavior. Both models are tested against existing methods using unbiased test data and provided in Matlab format.

1.4 Research Objectives and Overview of Tasks

The objective of this project is to develop unified National "Baseline" models, informed by ML/DL, for predicting Vs_{30} and Vs-versus-depth. The research tasks are: **(1)** compile all available Vs profiles in the U.S.; **(2)** compile an array of geospatial predictor variables at the site of each Vs measurement; **(3)** adopt an SVM functional form and fit V_S profiles; **(4)** Use machine- and deep-learning to develop models that predict V_{S30} and SVM parameters remotely, at any location, using only geospatial predictor variables (as part of this process, models for predicting the depth-of-bedrock will also be developed); **(5)** Test the prediction efficiency of the final models against existing models using an unbiased test set; and **(6)** code implementation. A synopsis of the research plan is shown in Figure 1.5.

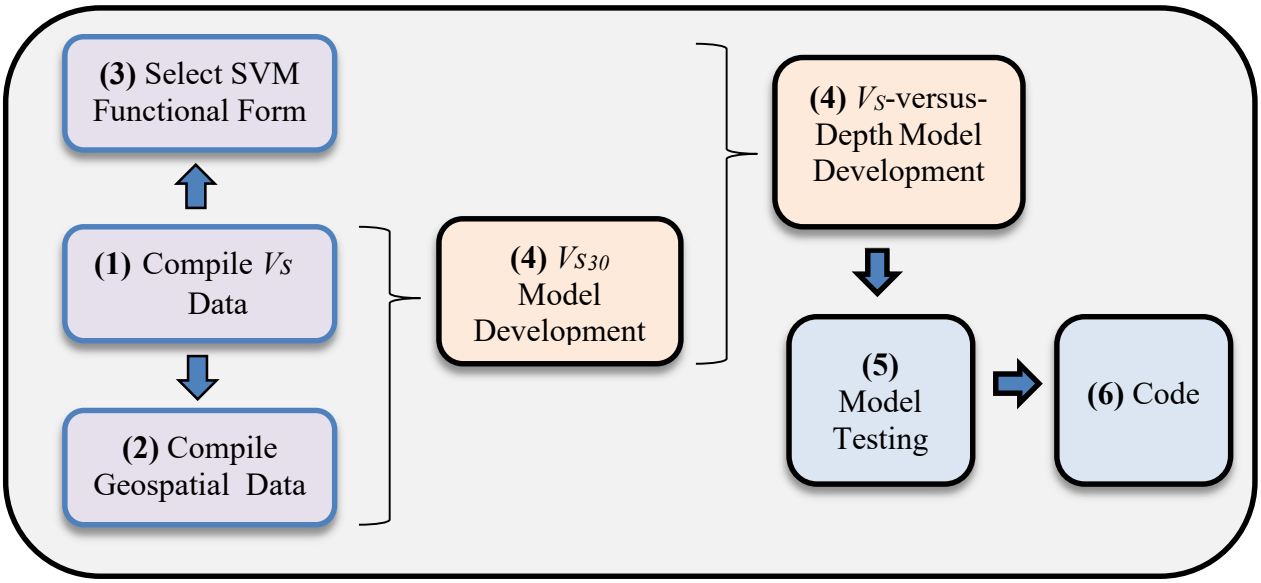

Figure 1.5 Synopsis of the research.

Chapter 2. DATA AND METHODOLOGY

2.1 COMPILATION OF VS PROFILES

A crucial element of this research was the compilation of a new database of measured Vs profiles in the United States, with the goal of achieving the greatest spatial distribution possible. At present, the largest such database known to the author is that of Kwak et al. (2021), which contains 915 profiles, located predominantly in California. Building off the Kwak et al. (2021) database by aggregating profiles from across the literature, a new database of 2737 profiles was developed, as summarized in Table 2.1. and mapped in Figure 2.1. This database includes 909 Vs profiles compiled by Ahdi et al. from Oregon, Washington, and British Columbia (Canada) and 180 profiles transformed from shear-wave time travel data in the USGS national CPT dataset, among other profiles from various sources. Several hundred profiles were digitized from hard copy as part of this process.

Table 2.1 Sources of compiled Vs profile measurements.

Reference	Number of Profiles
Kayen et al. (2011)	23
Ahdi et al.(2017)	909
USGS CPT Dataset	180
Carlos et al. (2018)	3
Kwak et al. (2021)	915
Louie, John N. (2020)	515
Pancha et al. (2011)	35
Rosenblad (2006)	10
Salomone et al. (2013)	34
Stephenson et al. (2009)	6
Stephenson et al. (2007)	17
Thompson et al. (2014)	27
Turner et al. (2012)	52
U. S. Geological Survey. (2017)	11
Total	2737

Figure 2.1 Map of compiled shear-wave velocity profiles.

Given this dataset, histograms of Vs30, maximum Vs, and maximum depth values calculated or directly obtained from the profiles of each dataset are shown in Figures 2.2-2.4. The vast majority of sites have Vs30 between 100 and 500 m/s, maximum Vs between 100 and 700m/s, and maximum measured depth less than 60 m. Measurements are relatively scarce outside these ranges.

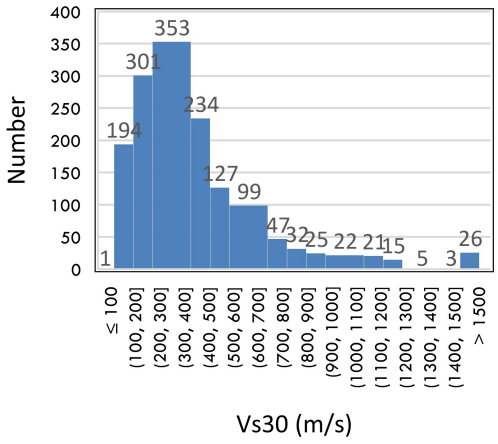

Figure 2.2 Histogram of Vs30, as computed from the compiled dataset.

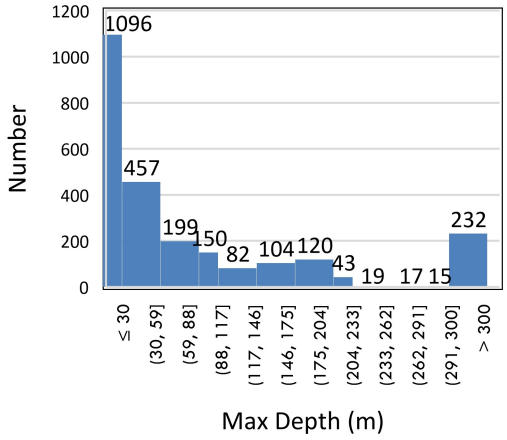

Figure 2.3 Histogram of maximum profile depth, as obtained from the compiled dataset.

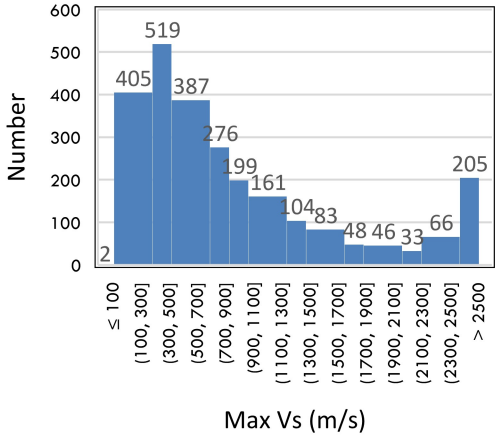

Figure 2.4 Histogram of maximum Vs, as obtained from the compiled dataset.

While the compiled dataset represents a valuable product, not all data meet the requirements for further analysis. Accordingly, profiles were removed for the following reasons: (i) quality control (non-monotonic behavior); (ii) insufficient depth (many measurements are shallow); (iii) and insufficient predictor variables (to be discussed). This process resulted in 735 high-quality profiles having spatial distribution similar to the original, complete database. This filtered dataset was further divided by randomly selecting 20% for unbiased testing. The other 80% form the

training set. The training set will never be used for testing and the test set will never be used for training. This will provide an unbiased assessment of model performance.

Figure 2.5 Map of filtered shear-wave velocity profiles.

2.2 Compilation of Geospatial Data

Using the proposed modelling approach, geospatial predictor variables readily obtained from satellite remote sensing or existing maps will be used to predict below ground conditions (Vs30, Vs versus depth, and depth to rock). Accordingly, 17 different geospatial predictors were next compiled from the coordinates of each measured Vs profile, as summarized in Table 2.2.

Table 2.2 Summary of geospatial predictor variables compiled for modelling.

Parameter Name	Type	Reference
Topographic Position Index	Numeric	e.g., Amatulli, et al. (2018)
Terrain Ruggedness Index	Numeric	
Roughness	Numeric	
Slope	Numeric	
Profile curvature	Numeric	
Tangential curvature	Numeric	
Vector Ruggedness Measure	Numeric	
Ground Water Table depth	Numeric	Fan and Miguez-Macho (2013)
Distance to River	Numeric	Verdin (2017)
Compound Topographic Index	Numeric	Verdin (2017)
Distance to coast	Numeric	Nasa (2020)
Global Depth to Bedrock Prediction	Numeric	Shangguan et al. (2017)
Annual Precipitation	Numeric	WorldClim (2020)
Unconsolidated Material	Binomial	USUS National Geologic Map Compilation
UnconsolidatedCoarsedetrital_Sand	Binomial	
UnconsolidatedFinedetrital_Clay	Binomial	
UnconsolidatedFinedetrital_Silt	Binomial	

2.3 Machine and Deep Learning Algorithms

Various ML/DL techniques were explored as part of the modelling approach, including Gaussian process models, support vector machines, decision trees, model ensembles with bagging, gradient boosting, or random forests, and neural networks. Gaussian process regression is nonparametric (i.e., not limited by a functional form), so rather than calculating the probability distribution of parameters of a specific function, GPR calculates the probability distribution over all admissible functions that fit the data. Support-vector machines, also known as support-vector networks, are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, a support-vector machines training algorithm builds a model that predicts whether a new example falls into one category or the other. Decision tree learning uses a decision tree as a predictive model to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). Artificial neural networks (ANNs), or connectionist systems, are

computing systems vaguely inspired by the biological neural networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

In general, modelling approaches that are easier to interpret tend to have lower predictive accuracy (e.g., single decision trees, support vector machines), while those with higher accuracy (e.g., neural networks, or ensembles of decision trees) are typically very complex to interpret. Each approach has numerous options and internal parameters (i.e., “hyperparameters”) (e.g., neural net optimization algorithm, activation function, layer quantity and size; regression tree leaf size; Gaussian basis and kernel functions; SVM kernel scale and box constraint). Once promising models were identified, hyperparameter optimization was employed, such that the hyperparameter values that minimized the model error were identified via an automated optimization scheme. 5-fold cross validation was used to control overfitting, as is common in model development. Additionally, training and test performance metrics were compared for signs of overfitting (i.e., significantly better training performance than test performance). In general, models with slightly lower accuracy but little-to-no sign of overfitting were favored over models that achieved the highest training accuracy but with suspicion of overfitting.

2.4 Modelling Approach and Functional Forms

Vs30 and depth-to-rock will be directly predicted via the aforementioned modelling approaches using all available geospatial variables:

Vs30, Depth-to-rock = AI_Trained_Model (Geospatial Variables as predictors)

This will allow more predictor variables to be used, with greater potential for those variables to be exploited. Each trained model will be evaluated by the mean absolute error (MAE), calculated over all test profiles. The MAE is defined as:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |V_{s30.predicted} - V_{s30.Measured}|$$

where *i* stands for number of each profile, *N* stands for number of all the profiles calculated. To predict the Vs-Depth profile, it is unfeasible to use the same approach as Vs30 and depth-to-rock, since the Vs-Depth relationship is too complex to be predicted directly (i.e., to independently predict the Vs at any and all depths of interest. Accordingly, a functional form will be fit to each Vs profile and the model coefficients producing the optimal fit will be recorded. Models will then be trained to predict each of the model coefficients independently. In forward prediction, prediction of the model coefficients will in turn produce a prediction of the full Vs profile.

Toward the selection of a functional form, the Shi and Asimaki (2018) functional form was first considered:

$$V_s(z) = \begin{cases} V_{S0} \quad , z < 2.5\ m \\ V_{S0}(1 + k(z - 2.5))^{\frac{1}{n}},\ z \geq 2.5\ m \end{cases}$$

Because the entire function is multiplied by Vs0, the predictions are strongly tied to Vs0, which indicates the initial velocity. As a result, predicted Vs at small and large depths are strongly correlated, but real Vs profiles don't always follow this trend. For example, soft soil could be shallow or deep, independent of the harder material that lays beneath. In addition, because the 3 parameters in the function are highly correlated, non-unique solutions are readily produced, such

that various values of Vs0, n, and k produce a similar profile. This presents a challenge during the optimization and modelling process. Accordingly, a modified functional from was adopted:

$$V_s = \begin{cases} V_{s0}, \ 0 < z < 2.5m \\ V_{s0} + (k - V_{s0})\left(\dfrac{z - 2.5}{15 - 2.5}\right)^{\frac{1}{n}}, \ z \geq 2.5m \end{cases}$$

In this function, there are also three parameters to optimize (during training) and predict (in forward prediction): Vs0 controls the initial shear wave velocity; k controls the rate-of-change of Vs at small depth; and n controls the overall curvature of the profile and the rate-of-change of Vs at large depth. Each represents one aspect of the profile in a more independent manner, decreasing the chances of non-unique solutions.

To fit each profile with the functional form, the Global-Search algorithm in MATLAB was b used to find optimal parameters Vs0, k and n by minimizing the error between the measured Vs and that predicted by the function, with error computed as:

$$error = \frac{1}{N}\sum_{i=1}^{N} |V_{s.predicted}(z_i) - V_{s.Measured}(z_i)|$$

To carry out this equation, each profile was discretized into 1-m thick layers; z_i is the depth from the surface to the midpoint of layer i, and N is the total number of layers for that profile.

Chapter 3. RESULTS

3.1 Vs30 Model and Testing

Following the aforementioned methodology and using 17 geospatial variables as predictors, a boosted tree ensemble model was ultimately adopted. This type of model combines many relatively weak learners into one relatively strong prediction model. Using the unbiased test set, the computed MAE of the newly proposed model is 61.11 m/s, whereas that of the USGS National baseline model is 183.05 m/s. The error of new model is thus 33% of the USGS model. The predictions made by the two models are contrasted in Figure 3.1(a), where the improved prediction efficiency of the new model is readily apparent. In addition, an index of error reduction, defined as the absolute error of the USGS model minus that of the new model, was computed and is shown in Figure 3.1(b). It can be seen that the new model tends to provide improved predictions at all values of Vs30, but that the improvement is most pronounces as Vs30 increases above 500 m/s.

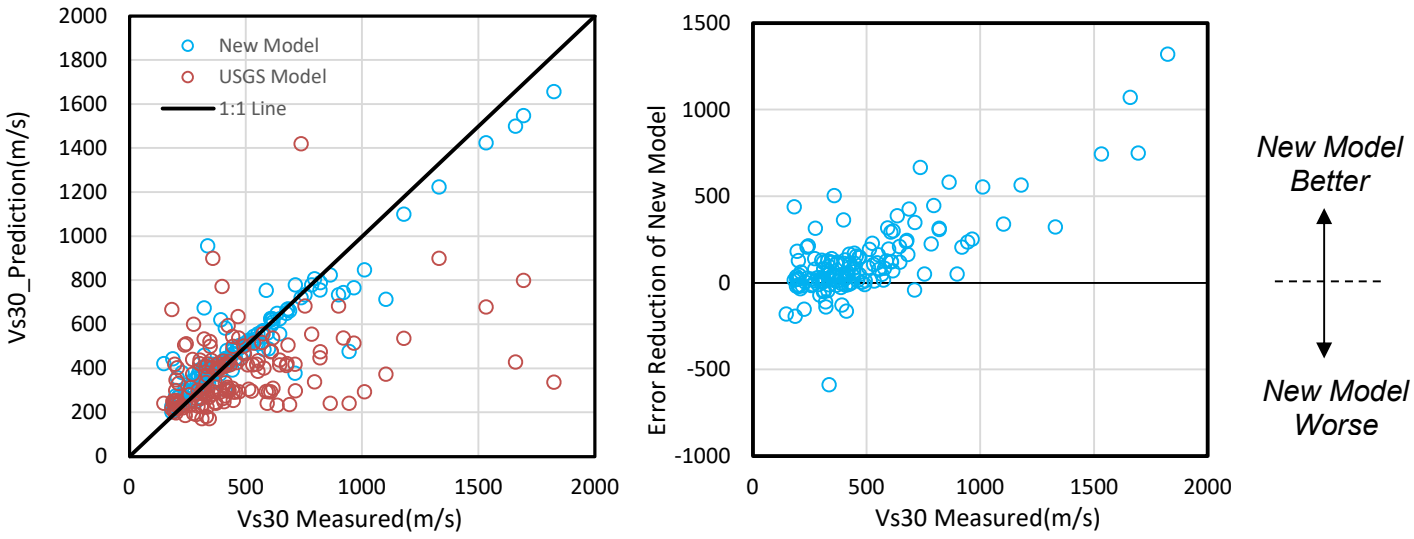

Figure 3.1 Vs30 prediction evaluation: (a) Vs30 as predicted by the newly proposed model vs. Vs30 as predicted by the USGS baselined model; (b) Error reduction (new model vs. USGS model) as a function of measured Vs30.

3.2 Vs versus Depth Model and Testing

Using the newly developed Vs30 prediction as an input, in addition to the aforementioned 17 geospatial variables, the following model types were ultimately adopted for each parameter in the function form: Coarse Gaussian support-vector machines (n prediction); Quadratic support-vector machines (Vs0 prediction); and Robust Linear (k prediction). Using the unbiased test set, prediction residuals (i.e., Vs predicted – Vs measured) were computed using the Shi and Asimaki (2018) model as well as the newly developed model. These residuals are plotted in Figure 3.2. Averaging the prediction error across all depths and all profiles, the proposed model was found to have an MAE of 189.5 m/s, whereas that of Shi and Asimaki (2018) is 286.19. Thus, the newly proposed model provides a significant improvement.

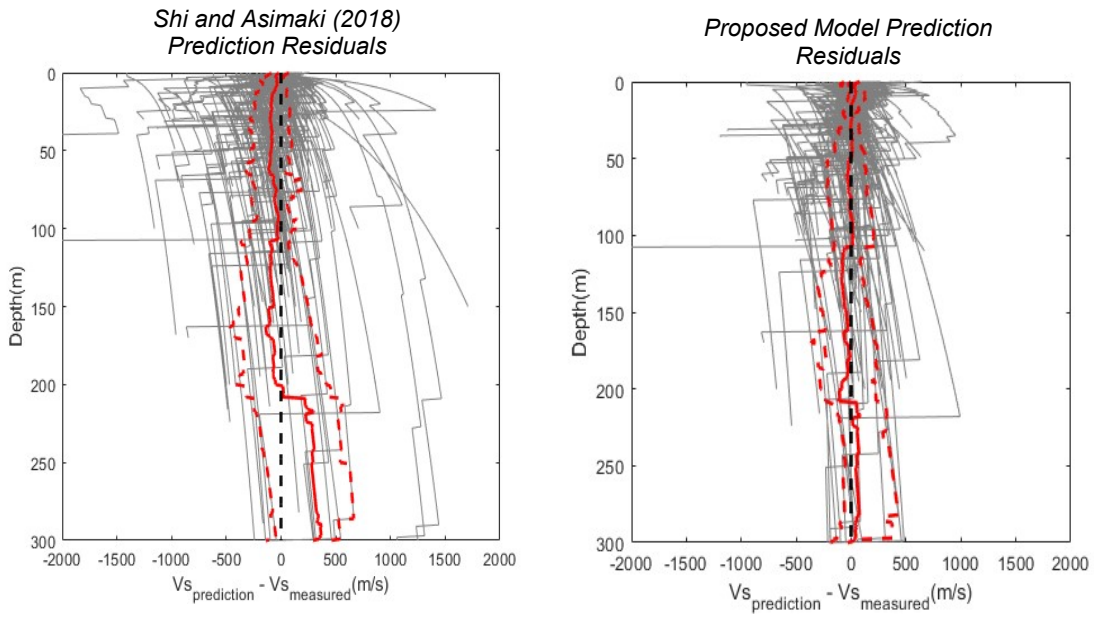

Figure 3.2 Prediction residuals: (a) Shi and Asimaki (2018) Vs-depth model; (b) newly proposed Vs-depth model.

As was done with Vs30, an index of error reduction, defined as the absolute error of the Shi and Asimaki (2018) model minus that of the new model, was computed and is shown in Figure 3.3. It can be seen that the new model tends to provide improved predictions at all values of Vs30, but that the improvement is most pronounces as Vs30 increases above 500 m/s. This is similar to the finding shown in Figure 3.1(b), where it was shown that predictions of Vs30 similarly improved most when Vs30 exceeded 500 m/s. It would appear, then, that in locations where Vs30 is predicted more accurately, the full Vs profile is predicted more accurately as well. To confirm, the two error reductions plotted in Figures 3.1(b) and 3.3 are plotted against one another in Figure 3.4. It can be seen that the reductions in error are strongly correlated, meaning that improved prediction of Vs30 generally leads to an improved prediction of the full Vs profile, as could be expected.

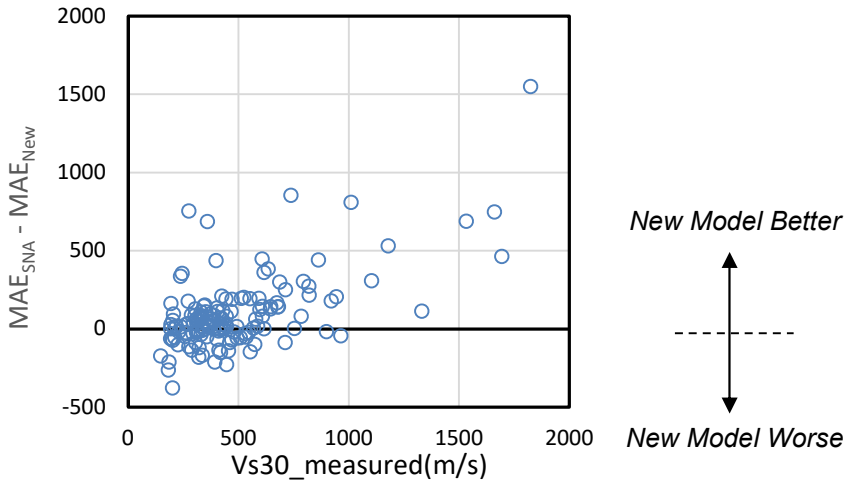

Figure 3.3 Error Reduction of new Vs-depth model compared to Shi and Asimaki (2018).

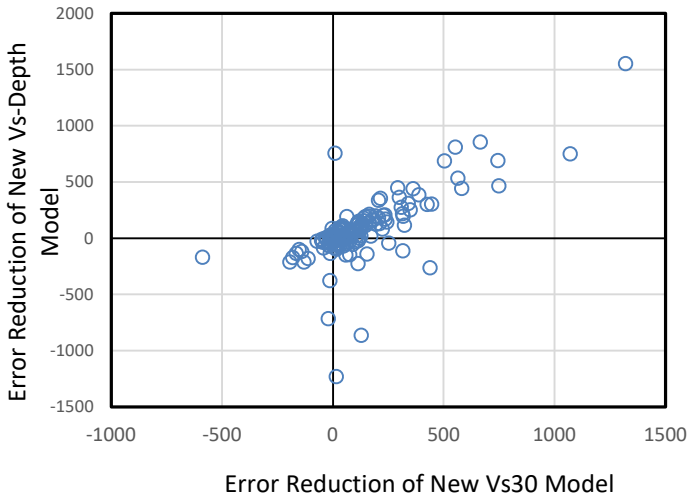

Figure 3.4 Error reduction of new Vs-depth model (as compared to Shi and Asimaki, 2018) versus error reduction of new Vs30 model (as compared to the USGS baseline model).

3.2.1 *Depth-to-Bedrock Model*

Because Vs-depth-models do not explicitly predict depth of rock (i.e., where a strong velocity contrast may be found), predicting depth-to-rock could improve the proposed model by defining the functional range of the soil velocity predictions. Traditionally, "rock" has been defined several ways from a wave-velocity standpoint. A velocity of 1000 km/s, for example, is widely used as a site term in empirical ground-motion prediction models (e.g., as an indication that a site is located in a sedimentary basin). A velocity of 760 m/s is also commonly used to define rock in international and U.S. building codes, given that this corresponds to the boundary between site class "B" (rock) and "C" (firm soil). Accordingly, models were separately developed to predict both the depth to a Vs of 1000 m/s ($Z_{1.0}$) and the depth to a Vs of 760 m/s ($Z_{0.760}$).

Existing models for predicting $Z_{1.0}$ include Abrahamson and Silva (2008) and Chiou and Youngs (2008), with both predicting $Z_{1.0}$ as a function solely of Vs30. Abrahamson and Silva (2008) define $Z_{1.0}$ using three exponential functions:

$$Z_{1.0} = \begin{cases} \exp(6.745) & \text{for } V_{S30} < 180\,\text{m/s} \\ \exp\left[6.745 - 1.35 \cdot \ln\left(\frac{V_{S30}}{180}\right)\right] & \text{for } 180 \le V_{S30} \le 500\,\text{m/s} \\ \exp\left[5.394 - 4.48 \cdot \ln\left(\frac{V_{S30}}{500}\right)\right] & \text{for } V_{S30} > 500\,\text{m/s}. \end{cases}$$

Chiou and Youngs (2008) model $Z_{1.0}$ using a single exponential function:

$$Z_{1.0} = \exp\left[28.5 - \frac{3.82}{8} \cdot \ln\left({V_{S30}}^8 + 378.7^8\right)\right].$$

The author is unaware of any existing model for predicting $Z_{0.760}$. Following the aforementioned modelling methodology, quadratic gaussian process regression models were ultimately adopted to predict s $Z_{1.0}$ and $Z_{0.760}$. Using the unbiased test set, the performance of the new $Z_{1.0}$ model is shown in Figure 3.6 in terms of measured vs. predicted $Z_{1.0}$. Across the test set, the new model has an MAE of 23.99 m. For comparison, the performance of the Abrahamson and Silva (2008) and Chiou and Youngs (2008) models are shown in Figures 3.6 and 3.7 respectively. These models had respective MAEs of 111.63 m and 85.87 m. Thus, the newly proposed model provides predictions that are significantly more accurate.

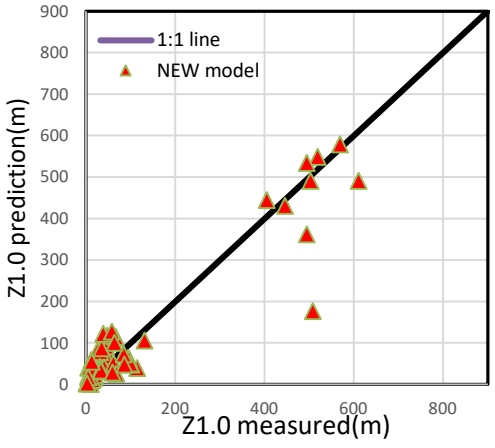

Figure 3.5 Performance of new $Z_{1.0}$ model.

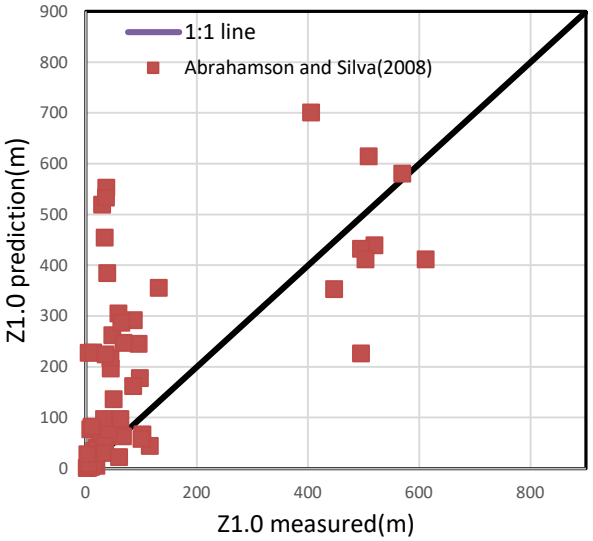

Figure 3.6 Performance of the Abrahamson and Silva (2008) $Z_{1.0}$ model.

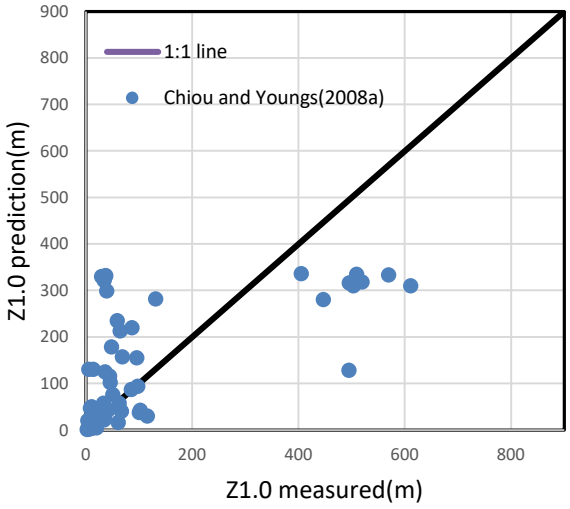

Figure 3.7 Performance of the Chiou and Youngs $Z_{1.0}$ model.

Finally, following the same approach, the performance of the developed For $Z_{0.760}$ model is shown in Figure 3.8. The model had a computed MAE of 24 m, very similar to that of the $Z_{1.0}$ model.

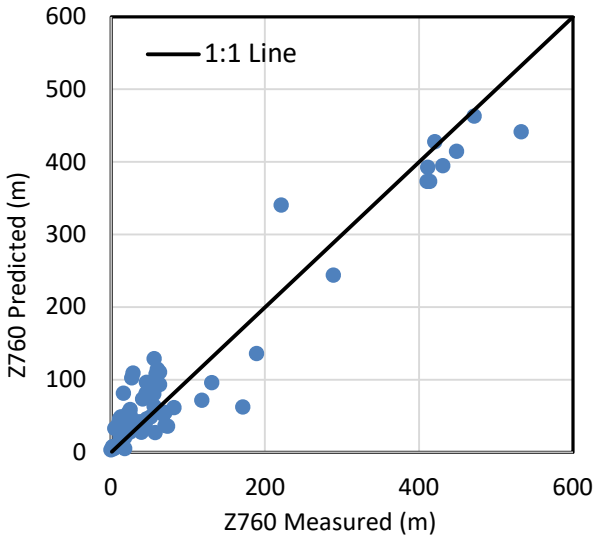

Figure 3.8 Performance of the new $Z_{0.760}$ model.

Chapter 4. CONCLUSIONS

This study presented a new Vs30 model, a new Vs-depth model, and a new depth-to-bedrock model. All three models provided more accurate predictions than existing solutions. Such models are critical to the evaluation of local amplification of ground motions at sites where detailed, site-specific profiles are not available (e.g., regional-scale simulations). The new models improved on previous works by using machine learning algorithms instead of traditional regression, and by increasing the number of predictor variables from 1 to 17. These variables can be easily and freely obtained without a detailed soil investigation. Ultimately, the modelling approach demonstrated herein could be expanded to further improve seismic-wave modelling by using a larger training set, additional predictor variables, and new learning algorithms, all of which are continuously growing and improving.

BIBLIOGRAPHY

Abrahamson, N. A., and Silva, W. J., (2008). Summary of the Abrahamson & Silva NGA ground-motion relations, *Earthquake Spectra* 24, 67–97.

Ahdi, S.K., Stewart, J.P., Ancheta, T.D., Kwak, D.Y., and Mitra, D. (2017). Development of VS Profile Database and Proxy-Based Models for VS 30 Prediction in the Pacific Northwest Region of North America. *Bulletin of the Seismological Society of America*, *107*(4), 1781-1801.

Allen, T. I., & Wald, D. J. (2007). Topographic slope as a proxy for global seismic site conditions (VS30) and amplification around the globe: U.S. Geological Survey Open-File Report 2007-1357.

Allen, T.I. and Wald, D.J. (2009). "On the use of high-resolution topographic data as a proxy for seismic site conditions (Vs30)." *Bulletin of the Seismological Society of America* 99: 935–943.

Boore, D. M. (2016). "Determining Generic Velocity and Density Models for Crustal Amplification Calculations, with an Update of the Boore and Joyner (1997) Generic Site Amplification for $V_S(Z)$ = 760 m/s." *Bulletin of the Seismological Society of America*, 106(1), 316–320.

Boore, D. M., and Joyner, W. B. (1997). Site amplifications for generic rock sites. *Bulletin of the seismological society of America*, *87*(2), 327-341.

Boore, D. M., Thompson, E. M., and Cadet, H. (2011). "Regional Correlations of VS30 and Velocities Averaged Over Depths Less Than and Greater Than 30 Meters." *Bulletin of the Seismological Society of America*, Seismological Society of America, 101(6), 3046–3059.

Bradley, B.A., Bae, S.E., Polak, V., Lee, R.L., Thomson, E.M. and Tarbali, K. (2017). "Ground motion simulations of great earthquakes on the Alpine Fault: effect of hypocenter location and comparison with empirical modelling." *New Zealand Journal of Geology and Geophysics*, 60(3), 188-198.

Campbell, K. W., and Bozorgnia, Y., (2007). Campbell-Bozorgnia NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters, PEER Report No. 2007=02, Pacific Earthquake Engineering Research Center, University of California, Berkeley.

Castellaro, S., Mulargia, F., and Rossi, P. L. (2008). "VS30: Proxy for seismic amplification?" *Seismological Research Letters,* 79(4), 540-543.

Chiou, B. S.-J., and Youngs, R. R., (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* 24, 173–215.

Cramer, C. H., Bauer, R. A., Chung, J. W., David Rogers, J., Pierce, L., Voigt, V., ... & Hempen, G. L. (2016). "St. Louis area earthquake hazards mapping project: Seismic and liquefaction hazard maps." *Seismological Research Letters*, *88*(1), 206-223.

Crowley, H. et al. (2019). The European Seismic Risk Model 2020 (ESRM 2020). Available at: https://www.globalquakemodel.org/gempublications/The-European-Seismic-Risk-Model-2020- %28ESRM-2020%29 (last accessed January 2021).

FGDC (2006). Metadata Quick Guide. *Federal Geographic Data Committee,* Available from https://www.fgdc.gov/metadata/documents/MetadataQuickGuide.pdf, Accessed 5/12/2018/.

Fick, S. E., and Hijmans, R. J., (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37(12), 4302-4315.

Frankel, A., Wirth, E., Marafi, N., Vidale, J., and Stephenson, W. (2018). "Broadband Synthetic Seismograms for Magnitude 9 Earthquakes on the Cascadia Megathrust Based On 3D Simulations and Stochastic Synthetics (Part 1): Methodology and Overall Results." *BSSA* 108 (5A): 2347-2369.

Geyin, M. and Maurer, B.W. (2020). Horizon: CPT-based liquefaction risk assessment and decision software. DesignSafe-CI, doi: 10.17603/ds2-2fky-tm46.

Heath, D. C., Wald, D. J., Worden, C. B., Thompson, E. M., & Smoczyk, G. M. (2020). A global hybrid VS 30 map with a topographic slope–based default and regional map insets. *EQS*, *36*(3), 1570-1584.

Holzer, T. L., Padovani, A. C., Bennett, M. J., Noce, T. E., and Tinsley, J. C. (2005). "Mapping NEHRP VS30 site classes." *Earthquake Spectra*, *21*(2), 353-370.

Horton, J.D., San Juan, C.A., and Stoeser, D.B., (2017). The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (ver. 1.1). USGS Data Series 1052, 46 p.

Jaume S.C. (2006). Shear-wave velocity profiles via seismic cone penetration test and refraction microtremor techniques at ANSS strong motion sites in Charleston, SC, *Seismol. Res. Lett.*, 77: 771-779.

Kayen, R. E., Carkin, B. A., Corbett, S. C., Zangwill, A., Estevez, I., & Lai, L. (2015). Shear wave velocity and site amplification factors for 25 strong-motion instrument stations affected by the M5. 8 Mineral, Virginia, earthquake of August 23, 2011 (No. 2015-1099). US Geological Survey.

Kwak D.Y.; Ahdi S.K.; Wang P.; Zimmaro P.; Brandenberg S.J.; Stewart J.P. (2021). Web portal for shear wave velocity and HVSR databases in support of site response research and applications. UCLA Geotechnical Engineering Group. DOI:10.21222/C27H0V

Lehner, B., Verdin, K, and Jarvis A., (2006). *HydroSHEDS Technical Documentation*. World Wildlife Fund US, Washington, D.C

Magistrale, H., Pechmann, J., & Olsen, K. (2009). "The Wasatch Front, Utah, community seismic velocity model". *Seismol. Res. Lett*, *80*, 368.

Marafi, N., Grant, A., Maurer, B.W., Rateria, G., Eberhard, M., and Berman, J. (2021). A generic soil velocity model that accounts for near-surface conditions and deeper geologic structure. *Soil Dynamics and Earthquake Engineering,* 140: 106461.

Maurer, B.W., Baird, A.J., and Geyin, M. (2019). Rapid prediction of infrastructure damage and loss due to earthquake-induced soil liquefaction. *USGS Report G18AP00006*, 85pp.

McPhillips, D.F., Herrick, J.A., Ahdi, S., Yong, A.K., and Haefner, S., (2020), Updated Compilation of VS30 Data for the United States: USGS data release, https://doi.org/10.5066/P9H5QEAC.

Odum, J. K., Williams, R. A., Stephenson, W. J., Worley, D. M., von Hillebrandt-Andrade, C., Asencio, E., ... & Cameron, A. (2007). Near-surface shear wave velocity versus depth profiles, Vs 30, and NEHRP classifications for 27 sites in Puerto Rico. U. S. Geological Survey.

Parker, G. A., Harmon, J. A., Stewart, J. P., Hashash, Y. M., Kottke, A. R., Rathje, E. M., ... & Campbell, K. W. (2017). Proxy-based VS 30 estimation in central and eastern North America. *Bulletin of the Seismological Society of America*, *107*(1), 117-131.

Petersen M.D. et al. (2019). “The 2018 update of the US National Seismic Hazard Model: Overview of model and implications.” *Earthquake Spectra* 36: 5–41.

Ramirez-Guzman, L., Graves, R. W., Olsen, K. B., Boyd, O. S., Cramer, C., Hartzell, S., ... & Zhong, J. (2015). Ground-motion simulations of 1811–1812 New Madrid earthquakes, central United States. *Bulletin of the Seismological Society of America*, *105*(4), 1961-1988.

Salomone, L. A., Hamel, J. F., & Kassawara, R. P. (2012). EPRI (2004, 2006) ground-motion model (GMM) review project: Shear wave velocity measurements at seismic recording stations. *Draft Report, EP-P43952/C19088*.

Schnabel, P. B., Lysmer, J., & Seed, H. B. (1972). “SHAKE: A computer program for earthquake response analysis of horizontally layered sites.” Report No. *EERC72-12, EERC*, *12*.

Shangguan, W., Hengl, T., de Jesus, J. M., Yuan, H., & Dai, Y. (2017). Mapping the global depth to bedrock for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *9*(1), 65-88.

Shi, J., & Asimaki, D. (2018). A Generic Velocity Profile for Basin Sediments in California Conditioned on VS30. *Seismological Research Letters*, 89(4), 1397–1409. https://doi.org/10.1785/0220170268

Small, P., Gill, D., Maechling, P. J., Taborda, R., Callaghan, S., Jordan, T. H., ... & Goulet, C. (2017). “The SCEC unified community velocity model software framework.” *SRL*, *88*(6), 1539-1552.

Stephenson, W. J., Reitman, N. G., & Angster, S. J. (2017). P- and S-wave velocity models incorporating the Cascadia subduction zone for 3D earthquake ground motion simulations—

Update for Open-File Report 2007–1348. In Open-File Report (Version 1.). https://doi.org/10.3133/ofr20171152

Stewart JP, Parker GA, Harmon JA, Atkinson GM, Boore DM, Darragh RB, Silva WJ and Hashash YM (2017) Expert panel recommendations for ergodic site amplification in central and eastern North America. PEER Report 2017/04. Berkeley, CA: Pacific Earthquake Engineering Research Center.

Thompson, E. M., L. G. Baise, R. E. Kayen, Y. Tanaka, and H. Tanaka (2010). A geostatistical approach to mapping site response spectral amplifications, *Eng. Geol*. 114: 330–342.

Thompson, E.M., Wald, D.J., and Worden, C. (2014). A Vs30 map for California with geologic and topographic constraints. *Bulletin of the Seismological Society of America* 104: 2313–2321.

Vrettos, C. (1996). "Simple inversion procedure for shallow seismic refraction in continuously nonhomogeneous soils." *Soil Dynamics and Earthquake Engineering*, *15*(6), 381-386.

Wald, D. J., and Allen, T. I., (2007). Topographic slope as a proxy for seismic site conditions and amplification. *BSSA* 97, 1379–1395.

Wald, D. J., P. S. Earle, T. I. Allen, K. Jaiswal, K. Porter, and M. Hearne (2008). Development of the U.S. Geological Survey's PAGER system (Prompt Assessment of Global Earthquakes for Response). *Proc. 14th World Conf. Earthq. Eng*., Beijing, China, 8 pp.

Williams R.A., Stephenson W.J., Odum J.K., Worley D.M. (2003). Seismic velocities from high-resolution surface seismic imaging at six ANSS sites near Memphis, Tennessee, U.S. Geological Survey, *USGS Open-File Report 03-218*, Reston, VA.

Wills, C. and Clahan, K. (2006). Developing a map of geologically defined site-condition categories for California. *Bulletin of the Seismological Society of America* 96: 1483–1501.

Wirth, E. A., Frankel, A. D., Marafi, N., Vidale, J. E., and Stephenson, W. J. (2018). "Broadband Synthetic Seismograms for Magnitude 9 Earthquakes on the Cascadia Megathrust based on 3-D Simulations and Stochastic Synthetics (Part 2): Rupture Parameters and Variability." *Bulletin of the Seismological Society of America,* 108 (5A): 2370-2388.

Worden, C., Wald, D., Allen, T., Lin, K., Garcia, D. and Cua, G. (2010). A revised ground-motion and intensity interpolation scheme for ShakeMap. *BSSA* 100: 3083–3096.