Final Technical Report (2021-2022)

Towards a Three-Dimensional Geotechnical Layer Model for Northern California

Collaborative Research with the University of Nevada Reno and California Institute of Technology

USGS Award Numbers G21AP10518 and G21AP10448

Hesam Tehrani[1], Grigorios (Greg) Lavrentiadis[2], Elnaz Seylabi[3], David McCallen[4], and Domniki Asimaki[5]

[1]Graduate Student Researcher, University of Nevada Reno
[2]Postdoctoral Scholar Research Associate, California Institute of Technology
[3]Principal Investigator, University of Nevada Reno (elnaze@unr.edu)
[4]Co-Principal Investigator, University of Nevada Reno
[5]Co-Principal Investigator, California Institute of Technology

June 17, 2023

Acknowledgements

This material is based upon work supported by the U.S. Geological Survey under Grant No. G21AP10518 and G21AP10448.

Disclaimer

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the U.S. Geological Survey. Mention of trade names or commercial products does not constitute their endorsement by the U.S. Geological Survey.

Abstract

This project gathers measured shear wave velocity (Vs) profiles in the Bay region to develop a parameterized, region-specific near-surface velocity model. Currently, the finest discretization step size in the detailed velocity model of Northern California is 100 m in the horizontal direction and 25 m in the vertical direction (Aagaard et al., 2020a), which inevitably constrains the capabilities of physics-bases earthquake simulations to capture site effects in concert with path and source effects over a broad range of frequencies concerning earthquake engineering applications. The proposed model's formulation is the same as the one proposed by Shi and Asimaki (2018). However, new scaling relationships are considered to enhance the predictive capability of the proposed model and its extrapolation capacity; additionally, a robust global regression analysis based on Monte Carlo Markov chain Bayesian inference is used to constrain the model's hyperparameters such that almost zero bias in computed shear wave velocity residuals is guaranteed irrespective of depth. The results suggest that the proposed model introduces a finer scale structure in near-surface layers, potentially improving modeling site effects in three-dimensional ground motion simulations. It is recommended to use the V_{S30} map by Wills et al. (2015) in computing Vs profiles using the proposed model because they show a better correlation to the V_{S30} of measured profiles in the Bay region. The maximum depths of most measured profiles were less than 100 m, and only a few had $Z_{1.0}$ information. Therefore, the proposed model should be used cautiously in deeper layers.

Contents

1 Introduction **4**

2 USGS San Francisco velocity model (SFVM) **6**

3 Overview of recently developed near-surface velocity models **7**
3.1 Sedimentary velocity model for California 7
3.2 Generic velocity model for Cascadia region 10

4 Measured Vs profiles in the Bay region **12**

5 Near-surface velocity model proposals for the Bay region **13**
5.1 General functional form 13
5.2 Model regression: First approach 14
5.3 Model regression: Second approach 15
5.4 Modeling of Velocity Variability 21

6 Model evaluation in the Bay region **21**

7 Concluding remarks **27**

A Bayesian Regression Overview **32**

1 Introduction

Three-dimensional (3D) earthquake ground motion simulations have been constrained, until recently, to low frequencies (Olsen et al., 2006; Graves and Pitarka, 2010; Bielak et al., 2010). However, over the last decade, the continuous growth of high-performance computing systems has made higher frequency simulations an increasingly realistic target (Taborda and Bielak, 2013, 2014; Bielak et al., 2016; Rodgers et al., 2018, 2019). This fact has motivated researchers to focus on refining existing source and geology models to capture source, path and site effects more accurately in broadband deterministic ground motion simulations (e.g., Shi and Day, 2013; Taborda et al., 2012; Roten et al., 2014; Seylabi et al., 2019; Savran and Olsen, 2016). A higher resolution shallow velocity structure, specifically, can significantly affect the amplitude and frequency content of the simulated ground motions, especially at frequencies concerning earthquake engineering applications, and can result in localized nonlinear behavior that significantly impacts observed surface motions. Thus, It is critical to understand better the coupled source, path, and site effects in broadband 3D earthquake simulations.

Weathered rocks and sedimentary deposits in the shallow crust, and their transition to the stiffer bedrock, are frequently represented by the optional geotechnical layer (GTL) in community velocity models accessible through the Unified Community Velocity Model (UCVM) framework (Small et al., 2017). The model, proposed by Ely et al. (2010), is a geometric function that, given V_{S30} measurements, yields the smooth variation of the shear wave velocity profile for the top 350 m of the crust. Various functions–e.g., linear spline and polynomial interpolation–were evaluated to represent the smooth shear wave velocity (V_S) profiles with depth based on criteria such as (a) the capability to represent a wide range of soil and rock velocity profile types; (b) facilitating a smooth transition to the crustal velocity model; (c) an ability to reasonably handle poor spatial correlation of V_{S30} and crustal velocity data; (d) simplicity and minimal parameterization (Ely et al., 2010). These considerations resulted in the model with cubic and square-root depth dependence extended to a transition depth of z = 350 m. The shear-wave velocity at the surface is derived from V_{S30} by a uniform scaling. The UCVM framework of the Southern California Earthquake Center (Small et al., 2017) includes a geology-based V_{S30} map of California (Wills and Clahan, 2006) to support this model. It should be noted that the method can also be applied to regions without direct measurement of V_{S30} by using V_{S30} estimates from topographic slope (Wald and Allen, 2007).

More recently, Shi and Asimaki (2018) used nearly a thousand measured velocity profiles with V_{S30} ranging from 150 m/s to 1000 m/s to develop a sedimentary velocity model (SVM) that translates V_{S30} and depth to shear wave velocity of 1 km/sec ($Z_{1.0}$) into a generic one dimensional (1D) velocity profile suitable for use in wave propagation based ground motions. They considered four data sets of V_S measurements for the development of the model: (1) 178 profiles measured by Yong et al. (2013) (AY), (2) 277 profiles documented in Boore (2003) (DB), (3) 137 profiles collected by Chris Wills (CW) from the California Geological Survey (personal correspondence, 2016), and (4) 322 profiles measured by LeRoy Crandall (LC) and Associates (personal correspondence, 2016). The V_S profiles were measured using two families of site characterization techniques: the DB, CW, and LC profiles were measured using invasive methods (e.g., suspension logging, cross-hole, and down-hole tests), and the AY profiles were measured using noninvasive methods based on the inversion of surface-wave dispersion curves. As shown in Figure 1, the V_S profiles in the first three data sets are concentrated mainly in Los Angeles and San Francisco. It is noted that the last dataset was proprietary and therefore is not shown in the Figure.

In another study, Marafi et al. (2021) proposed another generic velocity model for the Cascadia

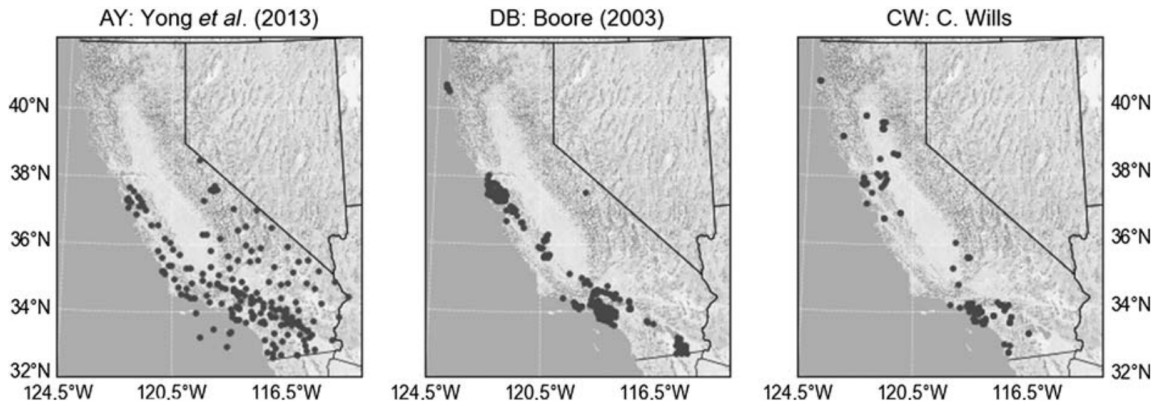


Figure 1: Locations of Vs profile measurements of three of the four datasets: AY (Yong et al., 2013), DB (Boore, 2003), and CW (Chris Wills). Each dot on the map denotes the location of a Vs profile measurement (Shi and Asimaki, 2018).

region, modifying the proposed model by Shi and Asimaki (2018). In developing this model, both V_{S30} and $Z_{1.0}$, as a means to consider deeper velocity structure, were used for model parameterization. A set of 218 shear wave velocity profiles from Ahdi et al. (2017) with $Z_{1.0}$ value was employed to develop the generic velocity model (Figure 2).

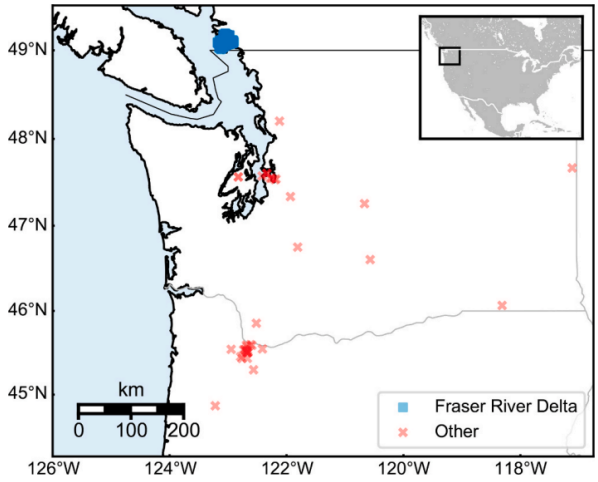


Figure 2: Locations of Vs profile measurements with $Z_{1.0}$ information for the Cascadia region (Marafi et al., 2021).

This project proposes using available V_S profile measurements in the Bay region to develop a parameterized, region-specific near-surface velocity model. The current USGS crustal velocity model comprises a high-resolution model of 290 km by 140 km by 45 km volume for the greater San Francisco Bay Area, surrounded by a coarser model spanning 650 km by 330 km by 45 km volume. The finest discretization size in the detailed seismic velocity model is 100 m in the horizontal direction and 25 m in the vertical direction (Aagaard et al., 2020a), which inevitably limits the capabilities of physics-based simulations to capture site effects. The proposed model aims to refine the near-surface velocity structure in the USGS velocity model in the Bay Area. The proposed

model and the future integration in the USGS velocity model will allow researchers to study the coupling of 3D source, path, and site effects in the Bay area; and to perform validation studies of deterministic earthquake simulations over a wide range of frequencies.

With this introduction, the remaining sections are organized as follows: Section 2 provides an overview of the USGS San Francisco velocity model (SFVM); Section 3 summarizes two previous studies on developing generic velocity models for California and Cascadia regions. Details of the measured V_S profiles in the Bay region are provided in Section 4, and the considered methods to develop the near-surface velocity model are discussed in Section 5. Section 6 evaluates the proposed model compared to measured profiles and SFVM, and concluding remarks are

2 USGS San Francisco velocity model (SFVM)

The USGS SFVM was developed to support 3D earthquake simulations in the Bay region (Aagaard et al., 2020b; Brocher et al., 2006). The USGS SFVM differs from tomographic models as it takes a forward approach by developing a 3D structural model of the local geologic units using geologic mapping, potential field geophysics, and seismicity (Waldhauser and Ellsworth, 2000; Ellsworth et al., 2000). This method, developed primarily by Brocher (2008), assigns elastic properties to the units based on rock-specific velocity-depth relationships, which is beneficial in areas like the California coast where strike-slip faults cause lateral juxtaposition of rocks with different properties. The SFVM covers a "detailed domain" that spans 290 km parallel to the San Andreas fault, 140 km perpendicular to the fault, and 45 km depth, surrounded by a "regional domain" that represents the geologic structure and elastic properties in a coarser manner (Aagaard et al., 2020b).

The initial version of the SFVM (v.05.0.0) was designed to simulate wave propagation for 1906 (Mw=7.9) San Francisco and 1989 (Mw=6.9) Loma Prieta earthquakes, as well as hypothetical (Mw=7.9) earthquakes on the San Andreas fault (Aagaard et al., 2008a,b). The model assigns elastic properties, i.e., compressional wave velocity V_P, shear wave velocity Vs, and density, to every point in the geologic model based on geologic unit and depth. The Vp-depth relations are based on measurements from various sources, while the Vs-depth relations were derived from V_P-V_S relations. The density is derived from the V_P and density relations, while the attenuation parameters were taken from Olsen et al. (2003).

Subsequent evaluations revealed systematic biases in the synthetic waveforms produced by the model, which were addressed through updates to the SFVM v.08.3.0 to enhance accuracy and consistency. Since then, the updated version has been used in various studies to analyze the interplay between earthquake ruptures and the 3D geologic structure on spatial variability of ground motions (Hirakawa and Aagaard, 2022). This version updates the Brocher (2008) rules based on waveform evaluations and comparisons with a travel-time tomography model. The changes include a decrease in V_P and V_S in some geologic units by a few percent and an increase in V_P and V_S by ten percent in the La Honda basin. Additionally, Qp and Qs are updated to vary as a function of V_P and V_S using the relations given in Brocher (2008).

The last version of SFVM is version v.21.1, released recently based on the final modifications on v.08.03. In this final version, further adjustments have been made to improve the accuracy of ground-motion predictions in the SF Bay region. These adjustments were made by analyzing geologic structures and experimenting with synthetic motions to match observed ones (Hirakawa and Aagaard, 2022).

As discussed before, the SFVM lacks resolution in near-surface layers (Figure 4), and the main

objective of this study is to develop a new near-surface velocity model for the Bay Area using measured V_S profiles within the region to help increase the accuracy of V_S values in near-surface layers. In Figure 4, the comparison of SFVM V_S values with depth obtained from two recent versions and the measured profiles at selected locations (shown in Figure 3) is pictured. The SFVM results do not necessarily comply with measured profiles. It is desired to develop a generic model to modify the SFVM model in near-surface layers to ensure better agreement with measured profiles in the Bay region.

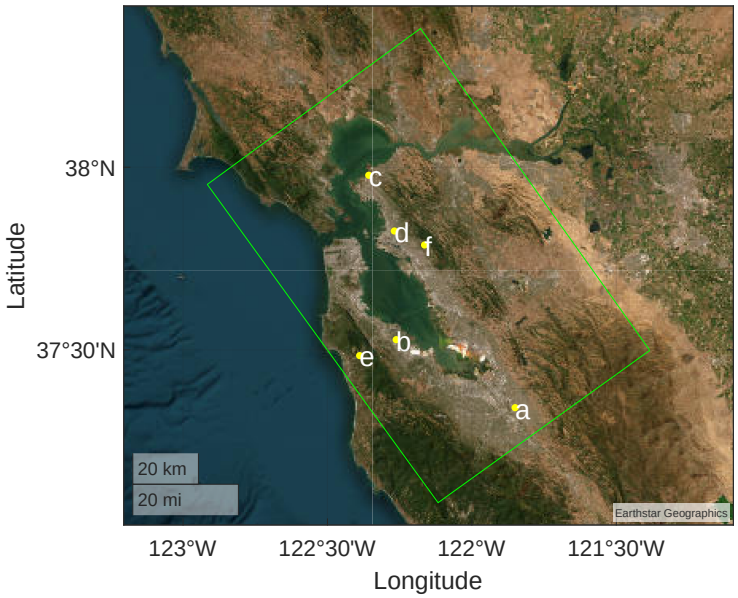


Figure 3: Selected locations to compare measured Vs profiles to USGS SFVM predictions.

3 Overview of recently developed near-surface velocity models

This section summarizes two recent studies on developing near-surface velocity models for California and Cascadia region.

3.1 Sedimentary velocity model for California (Shi and Asimaki, 2018)

Shi and Asimaki (2018) used the following analytic expression to fit the model over an ensemble of average Vs profiles.

$$V_S(z) = \begin{cases} V_{S0} & 0 \le z \le z^* \\ V_{S0}[1 + k(z - z^*)]^{1/n} & z > z^* \end{cases} \tag{1}$$

In Equation (1), z is depth, $z^* = 2.5$ m, and (V_{S0}, k, n) pairs are functions of V_{S30}. In this forward model, V_{S0} controls the initial shear wave velocity, k shows the rate-of-change of shear wave velocity at smaller depths, and n controls the curvature of the Vs profile and the rate-of-change of shear wave velocity values in higher depths.

To constrain the model parameters, the Vs database discussed in Section 1 is divided into several bins based on their V_{S30} values (Figure 5), and the mean profile in each bin is used to optimize (V_{S0}, k, n) for each V_{S30}. Figure 6 shows the trends of the optimized values as a function of V_{S30},

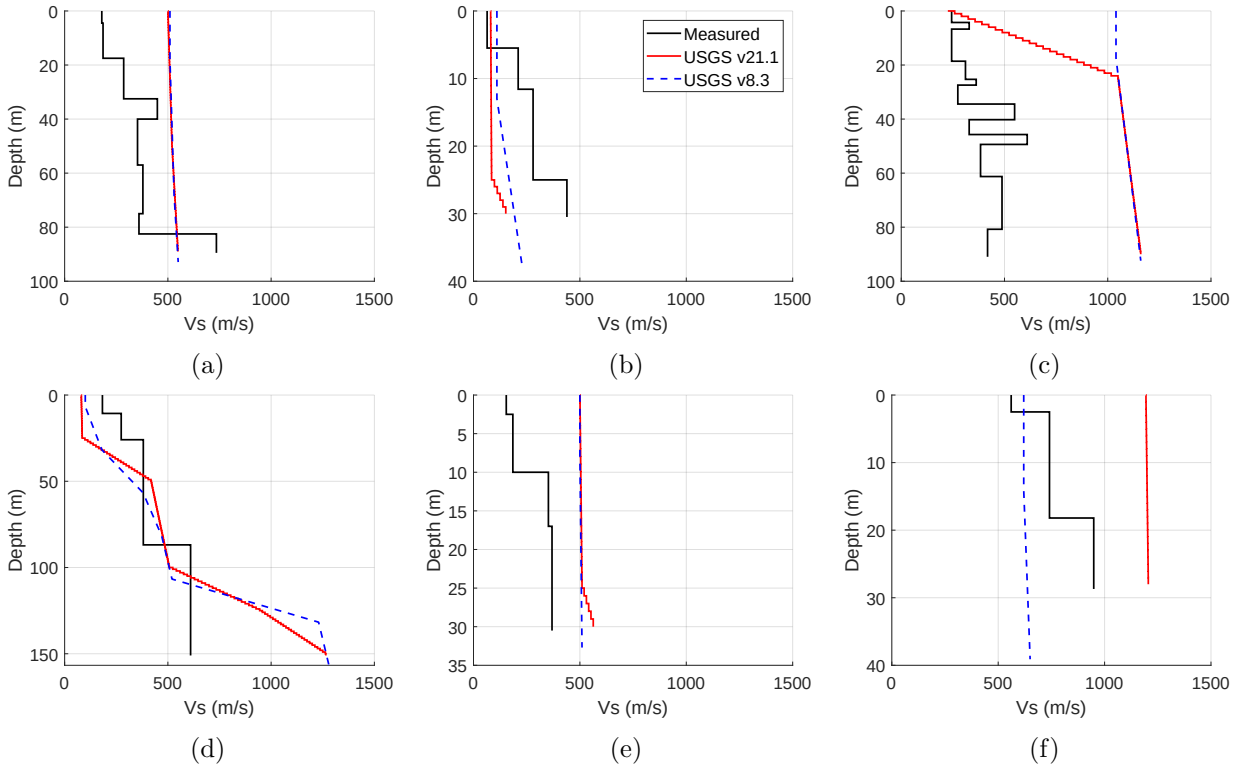


Figure 4: Comparison of Vs profiles obtained using two last versions of SFVM (v.08.03.0 and v.21.1) with measured ones at locations shown in Figure 3.

which are used to find and fit appropriate functions (Equation (2) to Equation (4)). To avoid extrapolation-associated errors, this model is recommended for V_{S30} less than 1 km/s that can be queried from the California V_{S30} maps (e.g., Thompson et al., 2014; Yong et al., 2013; Wills et al., 2015).

$$V_{S0} = p_1(V_{S30}^2) + p_2(V_{S30}) + p_3 \tag{2}$$

$$k = \exp(r_1[V_{S30})^{r_2} + r_3) \tag{3}$$

$$n = s_1 \exp(s_2 V_{S30}) + s_3 \exp(s_4 V_{S30}) \tag{4}$$

where $p_1 = -2.1688 \cdot 10^{-4}$, $p_2 = 0.5182$, and $p_3 = 69.452$ for V_{S0}, $r_1 = -59.67$, $r_2 = -0.2722$, and $r_3 = 11.132$ for k, and $s_1 = 4.11$, $s_2 = -1.0521 \cdot 10^{-4}$, $s_3 = -10.827$, and $s_4 = -7.6187$ for n.

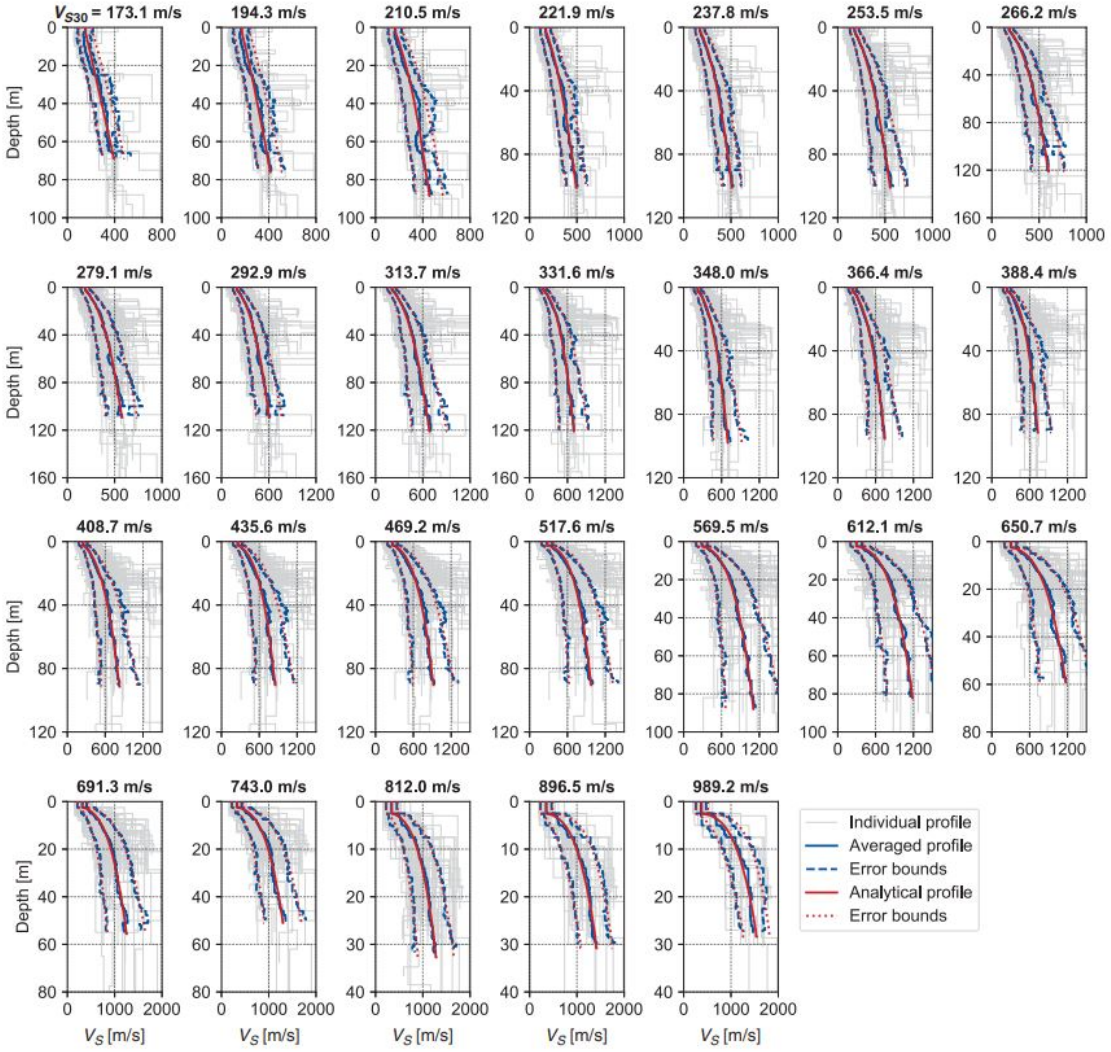


Figure 5: V_{S30} bins defined to parameterize (V_{S0}, k, n) pairs in Equation (1) as a function of V_{S30} in developing the SVM model (the figure is from Shi and Asimaki (2018)).

The goodness of fit of the predicted shear wave velocity profiles (GoF_{V_S}) is used to assess the performance of the proposed model compared to other existing velocity models for Southern California,

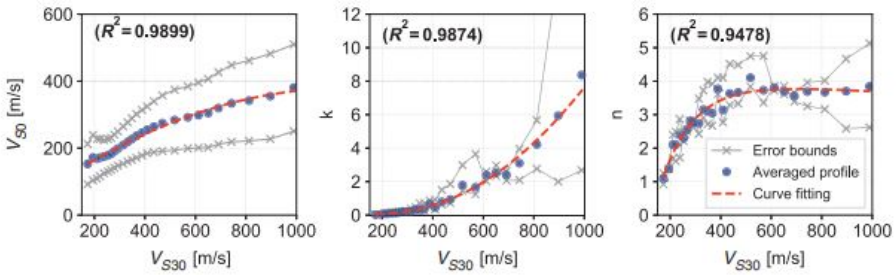


Figure 6: Variation of V_{S0}, k, and n as a function of V_{S30}. Each dot illustrates the result of the curve-fitting for each subplot of Figure 5. The fitted curve over the general trend is shown in red, and the error bounds (come from the red dashed line in Figure 5) are represented in gray (the figure is from Shi and Asimaki (2018)).

including Community Velocity Models (CVMs)-CVM-S4.26.M01 and CVM-H.15.1.0-and the Ely GTL (Ely et al., 2010). These results are shown in Figure 7, confirming its better performance.

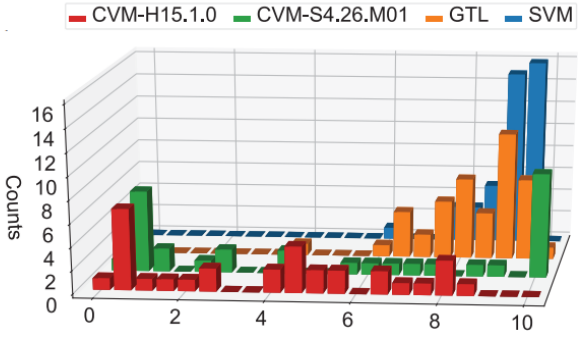


Figure 7: Goodness of fit comparison of different velocity models in Southern California (the figure is from Shi and Asimaki (2018)).

3.2 Generic velocity model for Cascadia region (Marafi et al., 2021)

Marafi et al. (2021) consider a similar generic model as follows:

$$V_S(z) = \begin{cases} V_{s0}, & z \leq 2.5 \\ V_{s0} + 1000.(k\frac{z-2.5}{Z_{1.0}-2.5})^{1/n}, & z \geq 2.5 \end{cases} \tag{5}$$

where $k = (\frac{1000-V_{s0}}{1000})^n$ is defined such that $V_S(Z_{1.0}) = 1000$ m/s. It is expected that by adding $Z_{1.0}$ as another parameter, the model represents better-constrained estimates of shear wave velocity with depth. n and V_{S0} are optimized to minimize the error between the predicted and measured profiles using the Nelder-Mead algorithm (Gao and Han, 2012). After minimizing the error for each profile, V_{S0} and n are derived and plotted versus V_{S30} and $Z_{1.0}$ (Figure 8) to define the trends in Equation (7) and fit their parameters.

$$V_{S0} = a_0 + a_1(V_{S30})^{a_2} \tag{6}$$

$$n = b_0(V_{S30})^{b_1}(Z_{1.0})^{b_2}(V_{S30}Z_{1.0})^{b_3} \tag{7}$$

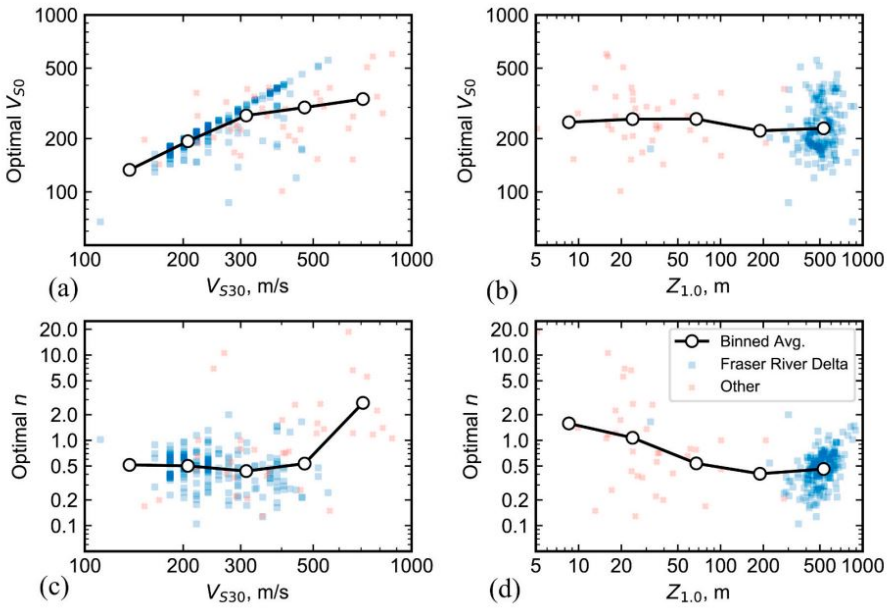


Figure 8: Variations of (a) optimal V_{S0} versus V_{S30}, (b) optimal V_{S0} versus $Z_{1.0}$, (c) optimal n versus V_{S30}, and (d) optimal n versus $Z_{1.0}$ (the figure is from Marafi et al. (2021)).

where $a_0 = -629$, $a_1 = 434$, and $a_2 = 0.122$ for V_{S0} and $b_0 = 0.00912$, $b_1 = 0.646$, $b_2 = -0.201$, and $b_3 = 0.136$ for n.

Figure 9 shows the computed residuals as a function of depth, using the proposed models by Shi and Asimaki (2018) and Marafi et al. (2021) to predict the measured V_S profiles in the Cascadia region. As shown, the Marafi et al. (2021) model results in less residual and bias, irrespective of depth, and therefore has better performance to be used within the intended region.

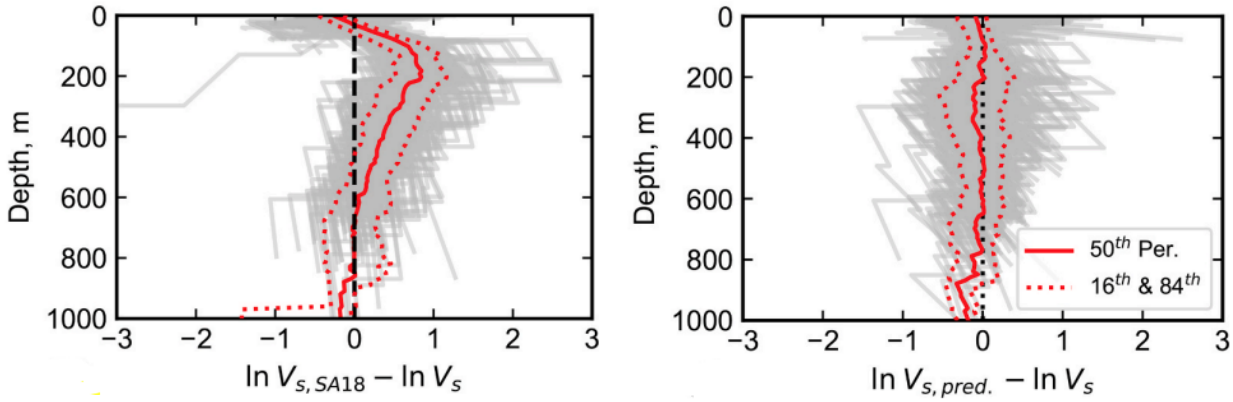


Figure 9: Vs residual comparisons obtained using the proposed model by (a) Shi and Asimaki (2018) and (b) Marafi et al. (2021) compared to measured Vs profiles in the Cascadia region (the figure is from Marafi et al. (2021)).

The main focus of this project is developing a generic near-surface velocity model for the Bay region

in Northern California. Considering this region's abundant active faults and complex geologic structure, developing a velocity model to represent better the near-surface velocity structure is of great importance. To formulate this velocity model, 211 Vs profiles are used. In Section 4, details of the gathered measured Vs profiles are elaborated upon.

4 Measured Vs profiles in the Bay region

For this study, five datasets of Vs profile measurements are collected within the Bay Area, including 13 profiles measured by Yong et al. (2013), 92 profiles documented in Boore (2003), 29 profiles collected by Chris Wills from the California Geological Survey, 3 profiles measured by LeRoy Crandall and Associates, and 70 profiles derived from VSPDB Vs profile database (Kwak et al., 2021). This resulted in 211 total profiles at locations shown in Figure 10.

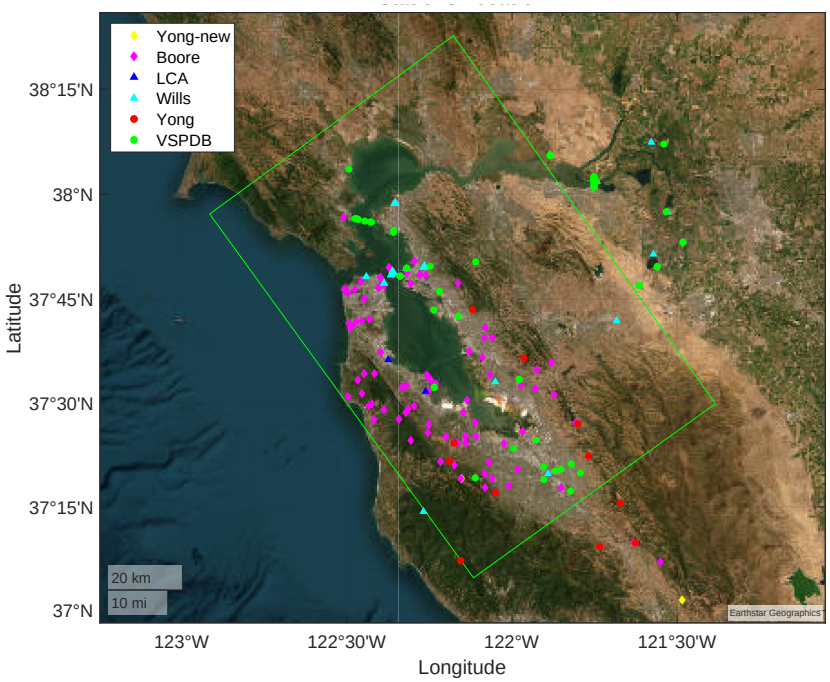
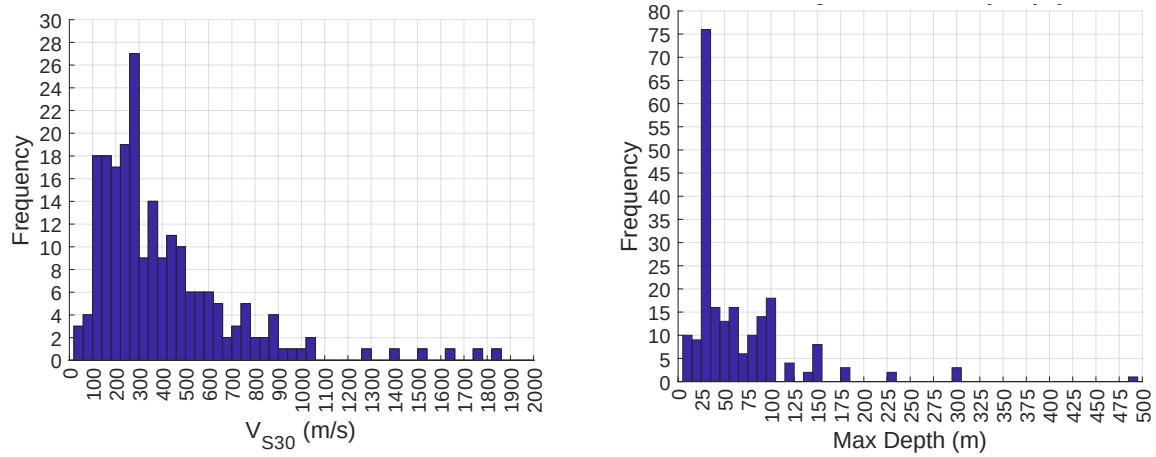


Figure 10: Locations of the 211 gathered Vs profiles in this study, using five different datasets discussed in Section 4.

As mentioned, the Boore, Wills, and Crandall profiles are obtained using invasive techniques, such as suspension logging, cross-hole, and down-hole tests. On the other hand, Yong profiles are derived using noninvasive methods that involve inverting surface-wave dispersion curves, such as spectral analysis of surface waves, multichannel analysis of surface waves, and/or refraction microtremor (ReMi) (Shi and Asimaki, 2018). VSPDB dataset includes a combination of both methods. Previous studies have indicated that noninvasive shear-wave velocity profiling techniques can produce similar results to invasive techniques (Boore and Brown, 1998; Boore and Asten, 2008; Brown et al., 2002; Rix et al., 2002; Stephenson et al., 2005; Bas et al., 2022). Therefore, all five datasets are combined for further analysis in this study.

Figure 11 shows the distribution of the V_{S30} and maximum depth for these profiles. As shown, about 35% of profiles only reach the depth of 25-30m, and the maximum depth of most profiles is

less than 100 m.



(a) (b)

Figure 11: Distributions of (a) V_{S30} and (b) maximum depth for the gathered measured Vs profiles in the Bay region.

5 Near-surface velocity model proposals for the Bay region

5.1 General functional form

The functional form of median shear-wave velocity versus depth follows the model of Equation (8) proposed by Shi and Asimaki (2018); however, using different scaling relationships.

$$V_S(z) = \begin{cases} V_{S0} & z \leq z^* \\ V_{S0}(1 + k(z - z^*))^{1/n} & z \geq z^* \end{cases} . \tag{8}$$

This study proposes a new scaling relationship for V_{S0} that ensures V_{S30} of the profiles are honored in the statistical fit (i.e., the calculated V_{S30} of the velocity model will exactly match the input V_{S30} value). Additional advantages of this decision are a reduction in the input parameters and a better extrapolation behavior. To this end, one can derive the formula for V_{S0} as a function of k, n, and V_{S30} by substituting Equation (8) in Equation (9) for V_{S30}. That is:

$$V_{S30} = \frac{30}{\int_0^{30} \frac{1}{V_S(z)} \mathrm{d}z} , \tag{9}$$

which results in

$$V_{S0} = V_{S30} \frac{(1 + 27.5k)^{1/n} + 2.5k(1 - \frac{1}{n}) - 1}{30k(1 - \frac{1}{n})} . \tag{10}$$

In developing the new model, two approaches with the same forward model are considered. The first approach derived each profile independently, and the scaling relationships of the model parameters were estimated as a second step. This approach is similar to the method used by Marafi et al. (2021). The second approach used global Bayesian inference to parameterize the velocity model and directly capture the hyper-parameters of the scaling relationships. Next, each approach is discussed in detail.

5.2 Model regression: First approach

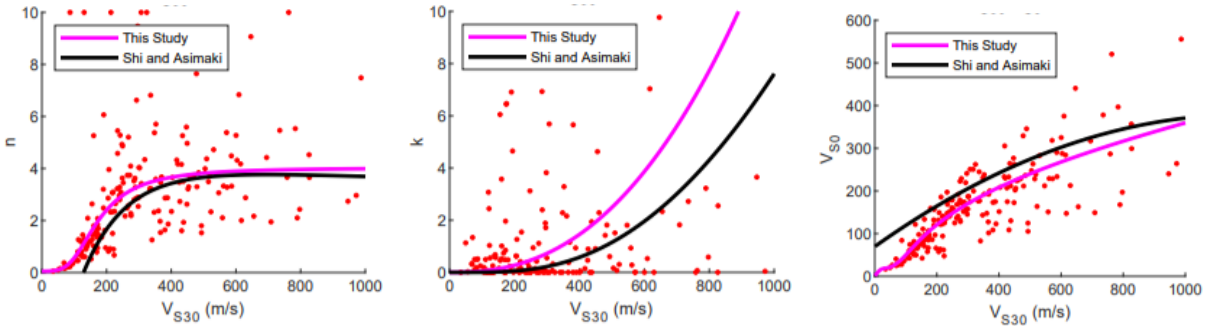
The main difference between the current approach with the procedure done by Shi and Asimaki (2018) is that in their study, profiles with different values of V_{S30} are grouped in different bins, the average profile is calculated in each bin, and the scaling parameters are derived based on fitting the average profile. However, in the first approach, each profile is used to minimize the error using the nonlinear least-square method (Equation (11)). Trust-region algorithm (Moré and Sorensen, 1983) is used in the optimization process. This algorithm helps us minimize errors between each profile's predicted and measured shear wave velocity values. That is,

$$k^*, n^* = \underset{k,n}{\operatorname{argmin}} \|V_{S,predicted} - V_{S,measured}\|^2 \tag{11}$$

where $V_{S,measured}$ is the measured V_S array at a different depth, $V_{S,predicted}$ is the computed V_S array at the same depths as in the measured profile using Equation (8), $\|\ \|$ denotes the L2-norm of the subtended variable, and k^* and n^* are the optimized k and n parameters for each V_S profile.

It should be noted that profiles with V_{S30} less than 100 or greater than 1000 m/s are not used in model development due to the scarcity of data in those ranges. Additionally, profiles with a maximum depth of less than 15 are not considered since they are not deep enough for model development. Overall, among all 211 profiles, 188 profiles are selected for the optimization process.

After regression analysis of each profile, the fitted model parameter values are plotted versus V_{S30} (Figures 12a to 12c). Then, taking the trends of V_{S30}-k, and V_{S30}-n into account, two functional formulations are fitted over them, and the values of hyperparameters are calculated using the trust-region algorithm (Moré and Sorensen, 1983). The fitted function over each parameter is shown in Equation (12).



(a) (b) (c)

Figure 12: Variations of optimized (a) k and (b) n values as a function of V_{S30}; (c) Variation of computed V_{S0} using optimized k and n values versus V_{S30}. Results are compared to those proposed by Shi and Asimaki (2018).

$$\begin{aligned} k &= r_1(V_{S30} - r_2)^{r_3}\,, \\ n &= \frac{s_1}{1 + s_2 V_{S30}^{-s_3}} + s_4\,, \end{aligned} \tag{12}$$

where $r_1 = 15.11$, $r_2 = 0.1885$, $r_3 = 2.015$, $s_1 = 4.014$, $s_2 = 6.069 \times 10^{-3}$, $s_3 = 2.951$, and $s_4 = 9.436 \times 10^{-4}$, and V_{S30} is in km/s unit. From Figure 12, it is clear that parameter k does not show a strong correlation to V_{S30}, and it is attempted to use a relatively simple model for fitting.

To determine the goodness of fit, shear wave velocity residuals are computed as follows, and compared to those from using the SVM model (Shi and Asimaki, 2018).

$$\text{Residual} = \ln V_{S,measured} - \ln V_{S,predicted} \tag{13}$$

Figure 13 shows the resulting residual for all measured profiles as a function of depth. Although the proposed model results in less residual in shallow depths, both models tend to underpredict the velocity at deeper layers, losing accuracy, and introducing bias. To solve this bias and have a better-fitted model for parameter k, it is decided to use a global Bayesian regression model, which is presented in the next section.

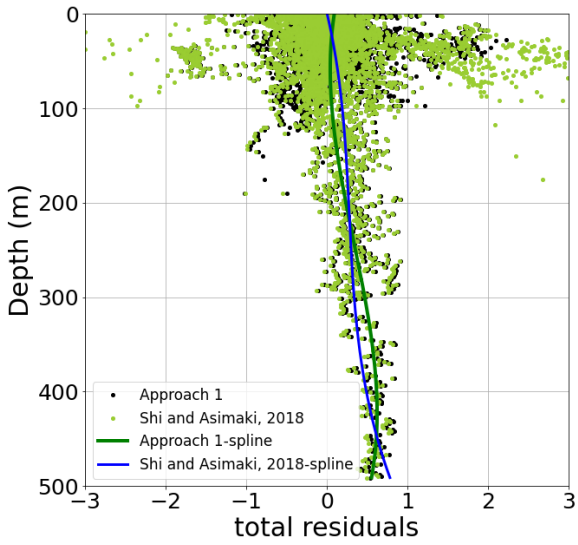


Figure 13: Residuals in computed Vs values using the first approach and Shi and Asimaki (2018) study. Equation (13) is used to compute residuals.

5.3 Model regression: Second approach

The second approach used Bayesian global inference to calibrate the parameters. Section A provides a general overview of Bayesian regression. Before performing Global Bayesian Regression, a shear wave velocity flatfile, including all profiles' data, should be gathered suitable for Bayesian regression analysis in STAN (Carpenter et al., 2017), which is a platform for statistical modeling and high-performance statistical computation. To prepare this flatfile, the columns below are considered:

- `DSID`: As we have different datasets, `DSID` is a number representing the dataset number.
- `DSName`: It contains the name of the dataset related to each row of data. All profiles in a specific dataset have the same `DSID` and `DSName`.
- `VelID`: It is a unique number assigned to each profile. Each profile contains several rows of data; thus, `VelID` is constant for all rows of that specific profile.
- `VelName`: Unique name for each profile.
- V_{S30}: V_{S30} shows the time-averaged shear wave velocity in the upper 30 m of each profile.

- `Lat` and `Lon` indicate the profile location in latitude/longitude coordinates.
- `Depth`, `Depth-MPt`, and `Thk`: They describe the depth to the base, depth to the mid-point, and thickness of each layer.
- `Vs`: Shear-wave velocity of each layer.

$$\begin{cases} r_1 \sim \text{Lognormal}(2.71, 0.10) \\ r_2 \sim \text{Lognormal}(-1.75, 0.40) \\ r_3 \sim \text{Lognormal}(0.65, 0.45) \\ s_1 \sim \text{Lognormal}(1.40, 0.25) \\ s_2 \sim \text{Lognormal}(-1.80, 0.50) \\ s_3 \sim \text{Lognormal}(1.05, 0.30) \\ s_4 \sim \text{Lognormal}(-7.00, 0.35) \\ \phi_{V_S} \sim \text{Lognormal}(-0.30, 0.60) \end{cases} \tag{14}$$

The variability of the measured velocity profiles around the median is modeled as follows:

$$\ln V_{S,measured}(z) = \ln V_S(z) + \delta V_S \tag{15}$$

where δV_S is the misfit between the measured velocity profiles and global median model and follows a normal distribution:

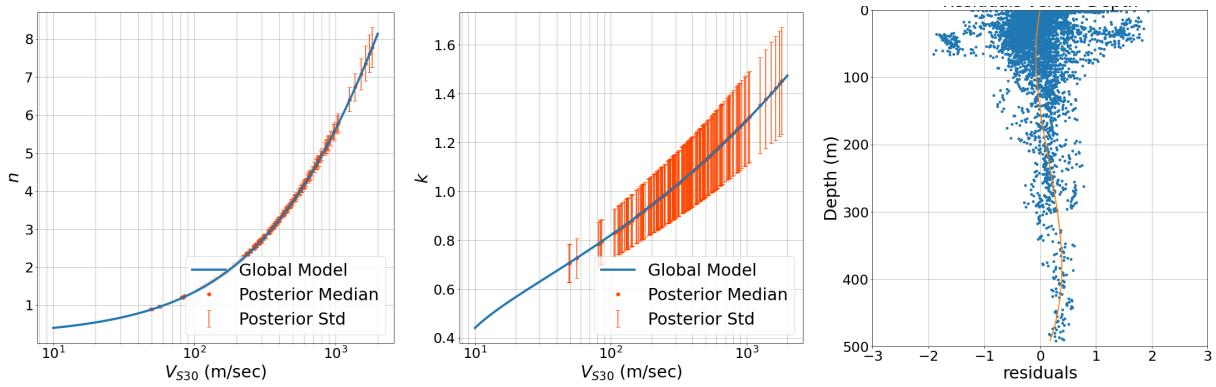
$$\delta V_S \sim \text{Normal}(0, \phi_{V_S}) \tag{16}$$

with ϕ_{V_S} being the standard deviation of the variability.

Figures 14a and 14b show the regression results for n and k as a function of V_{S30}, respectively, and Figure 14c shows the resulting residual versus depth. The global models for n and k are computed using the median of posterior distributions. These results suggest that the posterior distributions of estimated parameters still result in significantly more uncertainty in k than n, irrespective of V_{S30} values, and the residuals still suffer from a bias increasing with depth.

To determine the possibility of fitting a more informative functional form on k, the second regression analysis was performed using a random term δB_r to define $k = \exp(\delta B_r)$ to capture within-profile variability and determine its effects on decreasing bias in residuals. Additionally, it is decided to update the functional form for n to enforce $n \geq 1$. This will ensure fitting a convex function to the measured profiles and therefore avoid unrealistically rapid increases in Vs at deeper layers, which can become problematic in the future integration of this model with the SFVM. To summarize:

$$\begin{aligned} k &= \exp(\delta B_r)\,, \\ n &= 1 + \frac{s_1}{1 + s_2 {V_{S30}}^{-s_3}}\,. \end{aligned} \tag{17}$$



(a) (b) (c)

Figure 14: Relationship between model parameters (a) n and (b) k with V_{S30} using first approach's scaling relationships in global Bayesian regression analysis; (c) total residual values. In this figure, dots indicate the residual values of each profile at each depth, and the line depicts the spline, which is interpolated on residual values.

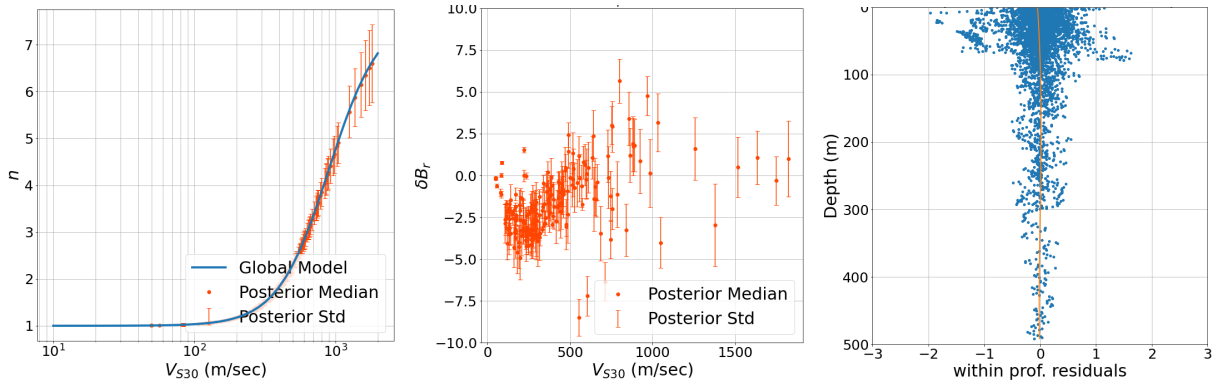
For this case, the prior distributions are defined as follows.

$$\begin{cases} \tau \sim \text{Lognormal}(-0.3, 0.6) \\ \delta B_r \sim \text{Normal}(0, \tau) \\ s_1 \sim \text{Lognormal}(1.25, 0.5) \\ s_2 \sim \text{Lognormal}(1.25, 0.5) \\ s_3 \sim \text{Lognormal}(0.65, 0.4) \end{cases} . \tag{18}$$

Figure 15 shows the variation of n and δB_r using the regression results. In these figures, the global model is obtained using the median of the converged posterior distributions for parameters s_1, s_2, s_3, and $k = 1$. Additionally, within-profile residuals are computed using the median of the estimated parameters. It is clear that the bias can be reduced significantly if one allows k to change spatially as a random term.

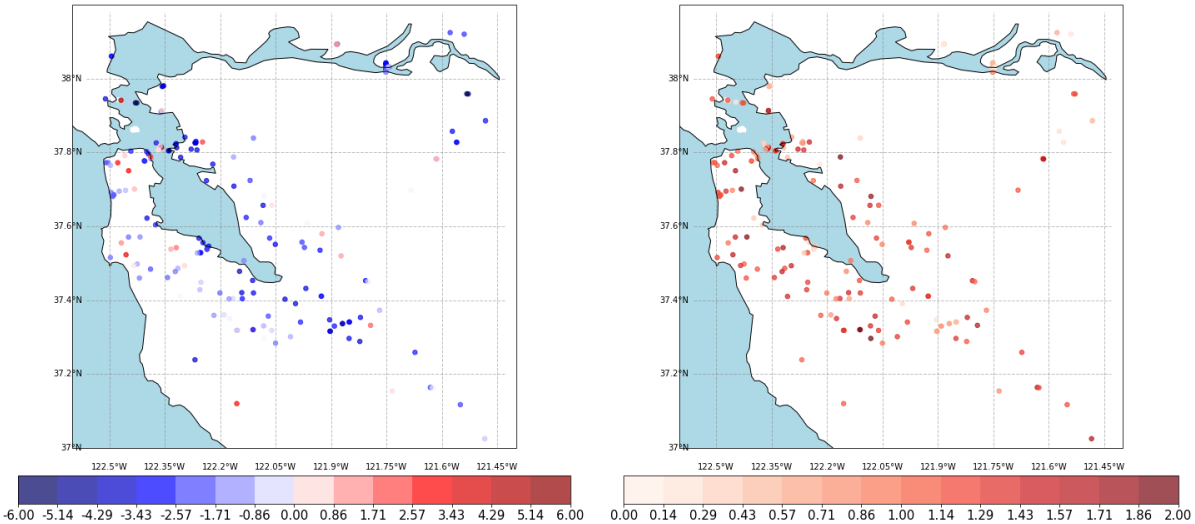
Figure 16 shows the spatial variability of the median and standard deviation for δB_r at the locations of measured profiles. At this stage, no strong spatial correlation is found, and therefore, for the remaining analysis, it is decided to use V_{S30} solely to define a new functional form for k based on the trend observed in Figure 15. To this end, using the trust-region algorithm, a sigmoid function is fitted (Equation (19)) to determine plausible prior distributions for the hyperparameters. Figure 17 shows the fitted model and the 95% confidence interval of each fitted k hyperparameter is used to define the prior distributions. For parameter n, the prior distributions with approximately high standard deviation are initially defined in Equation (18). Nonetheless, after defining k as the new function, the prior distributions of parameter n are modified to have less standard deviation based on the posteriors of the regression analysis performed on the case with the random term. The prior distributions of the hyperparameters for the final Bayesian inference are summarized as follows (Equation (20)).

$$k = \exp\left(\frac{r_1}{1 + r_2 {V_{S30}}^{-r_3}} + r_4\right) \tag{19}$$



(a) (b) (c)

Figure 15: Relationship between model parameters (a) n and (b) k with V_{S30}; (c) Whitin-profile residual, resulting from global Bayesian regression analysis using the second approach's scaling relationship for n and the spatially variable random term for k (Equation (17)). Dots in (c) indicate the residual values of each profile at each depth, and the line depicts the spline, which is interpolated on residual values.



(a) (b)

Figure 16: Spatial variability of δB_r in terms of its (a) median and (b) standard deviation values derived for each profile.

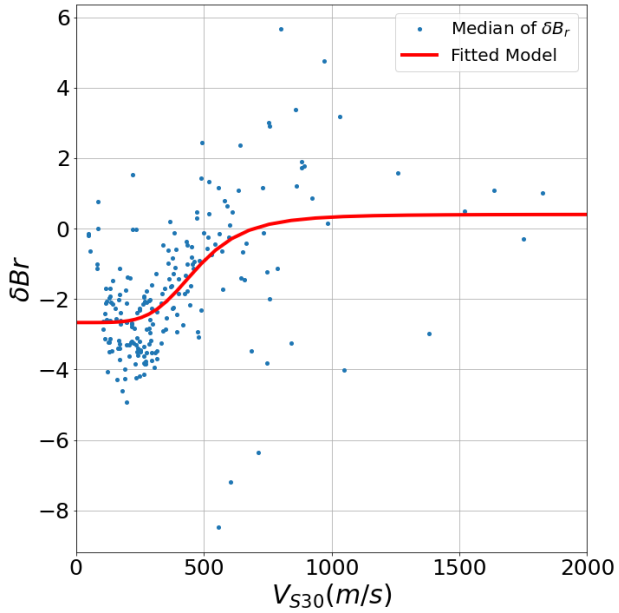


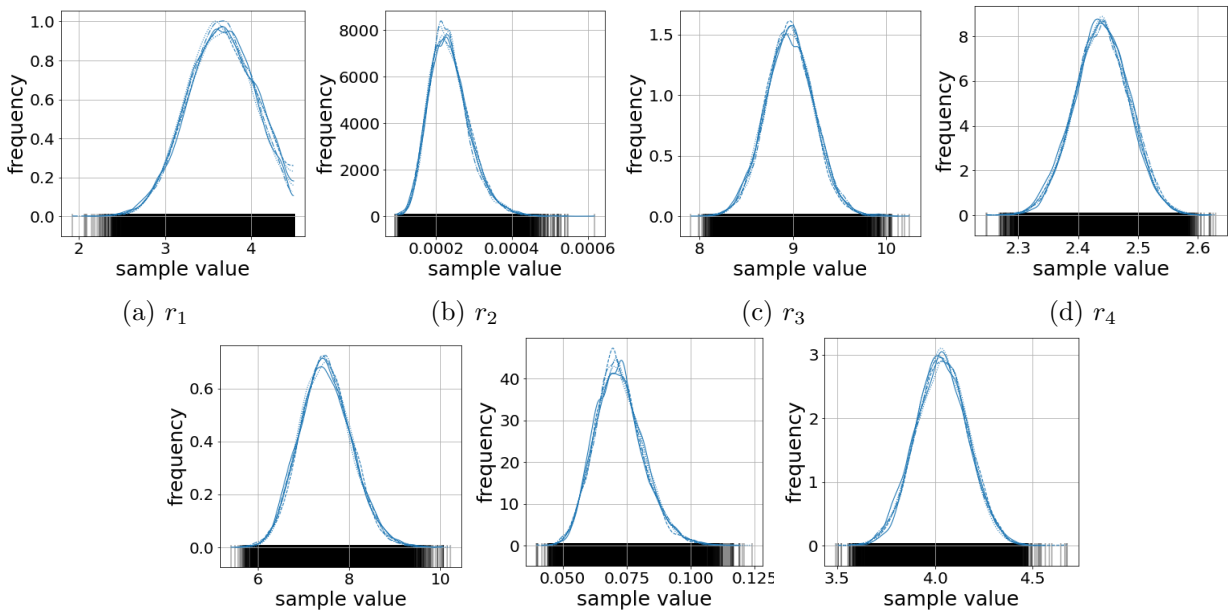
Figure 17: Fitted function over δB_r-V_{S30} using the trust-region algorithm.

$$\begin{cases} r_1 \sim \text{Lognormal}(1.1, 0.2) \\ r_2 \sim \text{Lognormal}(-8.3, 0.5) \\ r_3 \sim \text{Lognormal}(2.15, 0.1) \\ r_4 \sim \text{Lognormal}(0.9, 0.1) \\ s_1 \sim \text{Lognormal}(1.92, 0.3) \\ s_2 \sim \text{Lognormal}(-2.7, 0.15) \\ s_3 \sim \text{Lognormal}(1.4, 0.1) \end{cases} \tag{20}$$

Profiles with $V_{S30} > 100$ m/s are used to run the Bayesian regression since it resulted in the slightest bias in computed residuals. In this Bayesian regression analysis using MCMC, six chains are considered to confirm model convergence. The posterior distributions of the hyperparameters for all six chains are shown in Figure 18. Table 1 lists the results of Bayesian regression where the quantiles are obtained using all six chains. It is worth mentioning that V_{S30} values in Equations (17) and (19) must be in km/s unit.

Table 1: Statistics of converged hyperparameter using the global Bayesian regression analysis in the second approach.

	r_1	r_2	r_3	r_4	s_1	s_2	s_3
5^{th} Percentile	2.741	$1.56e^{-4}$	8.512	2.356	6.391	0.057	3.784
25^{th} Percentile	3.141	$1.95e^{-4}$	8.758	2.404	6.900	0.066	3.915
50^{th} Percentile (Median)	3.423	$2.27e^{-4}$	8.932	2.435	7.274	0.072	4.005
75^{th} Percentile	3.715	$2.64e^{-4}$	9.111	2.467	7.659	0.078	4.097
95^{th} Percentile	4.115	$3.26e^{-4}$	9.365	2.512	8.246	0.089	4.227
Mean	3.426	$2.32e^{-4}$	8.936	2.435	7.291	0.072	4.006

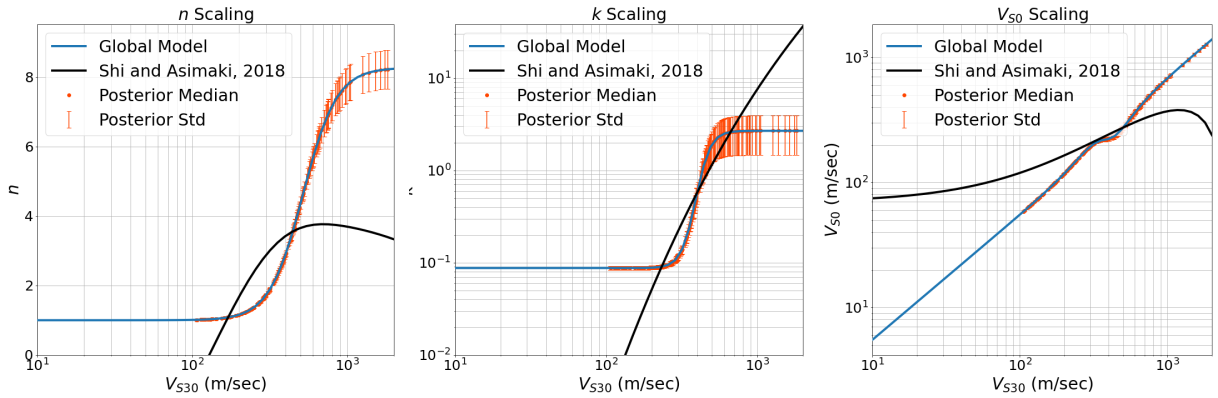


(a) r_1 (b) r_2 (c) r_3 (d) r_4

(e) s_1 (f) s_2 (g) s_3

Figure 18: Converged posterior distributions of hyper-parameters for the considered scaling relationships n and k in the second approach: (a) r_1, (b) r_2, (c) r_3, (d) r_4, (e) s_1, (f) s_2, and (g) s_3.

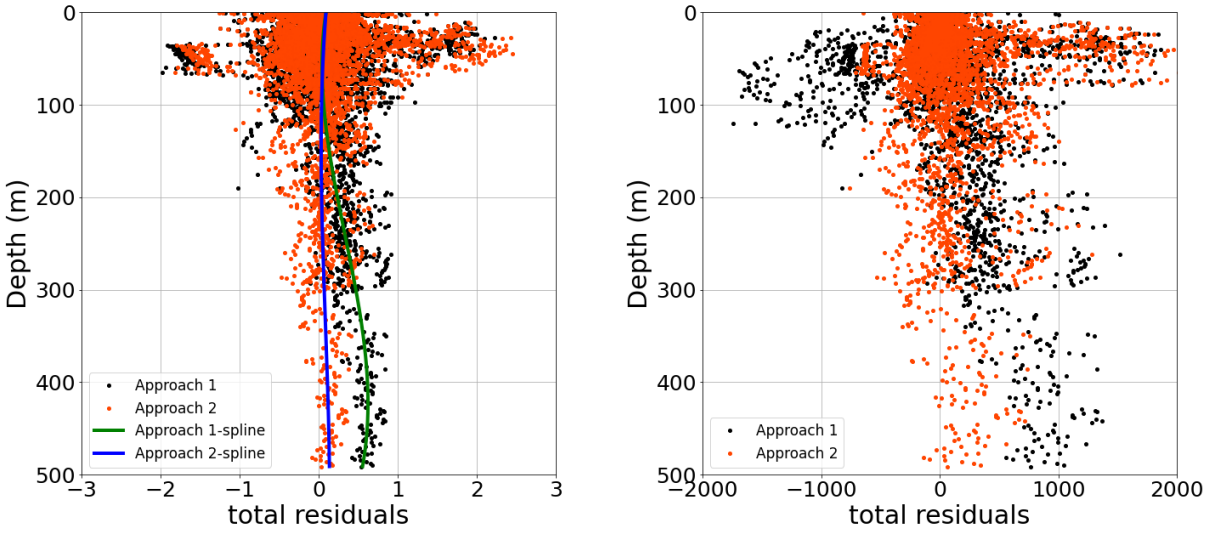
Figures 19a and 19b shows the variation of n and k as a function of V_{S30} using the Bayesian regression results. Figure 19c, on the other hand, shows the variation of V_{S0} using Equation (10). In all cases, the results of Shi and Asimaki (2018) study are plotted for comparison. It should be noted that, by design, the proposed model shows better extrapolation capacity, which has important implications for its use as a generic model for the Bay region and its future integration into SFVM.



(a) (b) (c)

Figure 19: Relationship between model parameters (a) n, (b) k, and (c) V_{S0} with V_{S30} obtained by performing global Bayesian regression analysis using the prior distributions (Equation (20)) defined in the second approach.

Figure 20 shows the computed residuals using the second approach and its comparison to those from the first approach. As expected, the second approach resulted in smaller residual values and less bias, especially at deeper layers.



(a) (b)

Figure 20: Residual comparison of two proposed models, referred to as first approach and second approach shown in (a) logarithmic (Residual $= \ln V_{S,measured} - \ln V_{S,predicted}$) (b) linear (Residual $= V_{S,measured} - V_{S,predicted}$) scales.

5.4 Modeling of Velocity Variability

In forward applications, the variability of the velocity profile around the median model is obtained by randomly sampling δV_S to create synthetic realizations:

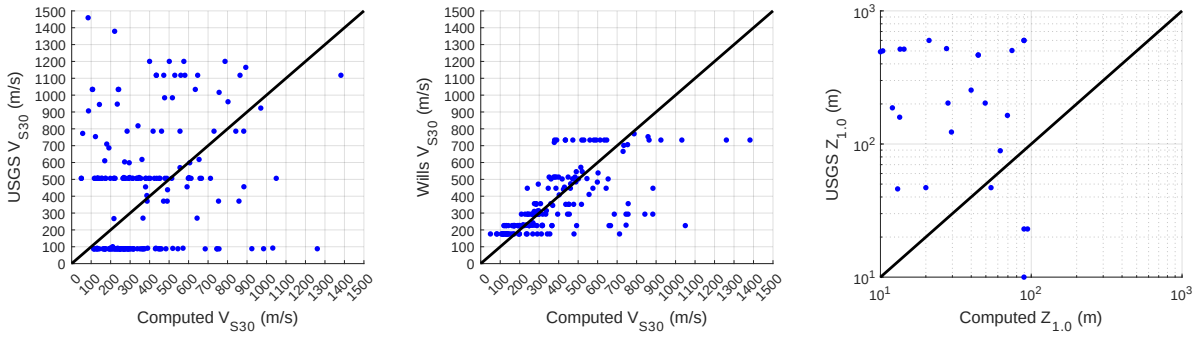
$$V_{S,realiz}(z) = V_S(z) \exp(\delta V_S) \tag{21}$$

For the along-depth correlation, the Toro (1995) model is recommended. A region-specific along-depth correlation model will be developed as part of the next phase of this project.

6 Model evaluation in the Bay region

In this section, the results of global Bayesian regression (i.e., the second approach) are used to predict shear wave velocity at different locations within the Bay region and compare them to measured profiles and those queried from the latest version of the USGS SFVM. V_{S30} is the only required parameter to compute Vs using the proposed model, and the V_{S30} map by Wills et al. (2015), a geology and topography-based map, is deemed suitable for this purpose. The V_{S30} of measured profiles are used to determine how well they correlated with Wills et al. (2015) (i.e., Wills V_{S30}) and with USGS V_{S30}, obtained using the USGS SFVM Vs profiles queried at the locations of measured Vs profiles. As shown in Figures 21a and 21b, Wills V_{S30} correlates significantly better and, therefore, will be used to compute V_{S30} at all considered locations for the proposed model's

evaluation. Additionally, locations of measured profiles with $Z_{1.0}$ information are used to compute SFVM $Z_{1.0}$, and the results are shown in Figure 21c, showing poor correlation.



(a) (b) (c)

Figure 21: Correlations of (a) USGS SFVM (v.21.1) V_{S30} values and (b) Wills V_{S30} values versus the V_{S30} values of the measured Vs profiles; (c) correlation of $Z_{1.0}$ for the measured Vs profiles and USGS SFVM (v.21.1).

Figure 22 compares Vs profiles computed using the proposed model and USGS SFVM to measured profiles at six locations shown in Figure 3. Both measured V_{S30} and Wills V_{S30} are used to compute the Vs profiles for the proposed model. Also, two configurations are considered for updating the USGS profiles using the proposed model, including up to $Z_{1.0}$ and entirely. Overall, the results of the proposed model agree well with the measured profiles, and using Wills V_{S30} provides similar profiles at these locations.

Next, two cross-sections shown in Figure 23 are considered to examine how the proposed model may modify the USGS SFVM up to $Z_{1.0}$. Figures 24 and 25 shows the background USGS Vs along each cross-section and their modifications using the proposed model. The proposed model adds more resolution to near-surface layers in both cross-sections. Additionally, they generally result in stiffer profiles in deeper layers. Future studies will focus on determining the appropriate maximum depth for which the proposed model could be used reliably. Lastly, Figures 26 and 27 show the spatial variation of Vs over the Bay region at two depths of 1m and 10m. Considering the course resolution of the USGS SFVM, the Vs contours are the same at both depths, while using the proposed model modifies the USGS model resulting in different velocity values at these two depths.

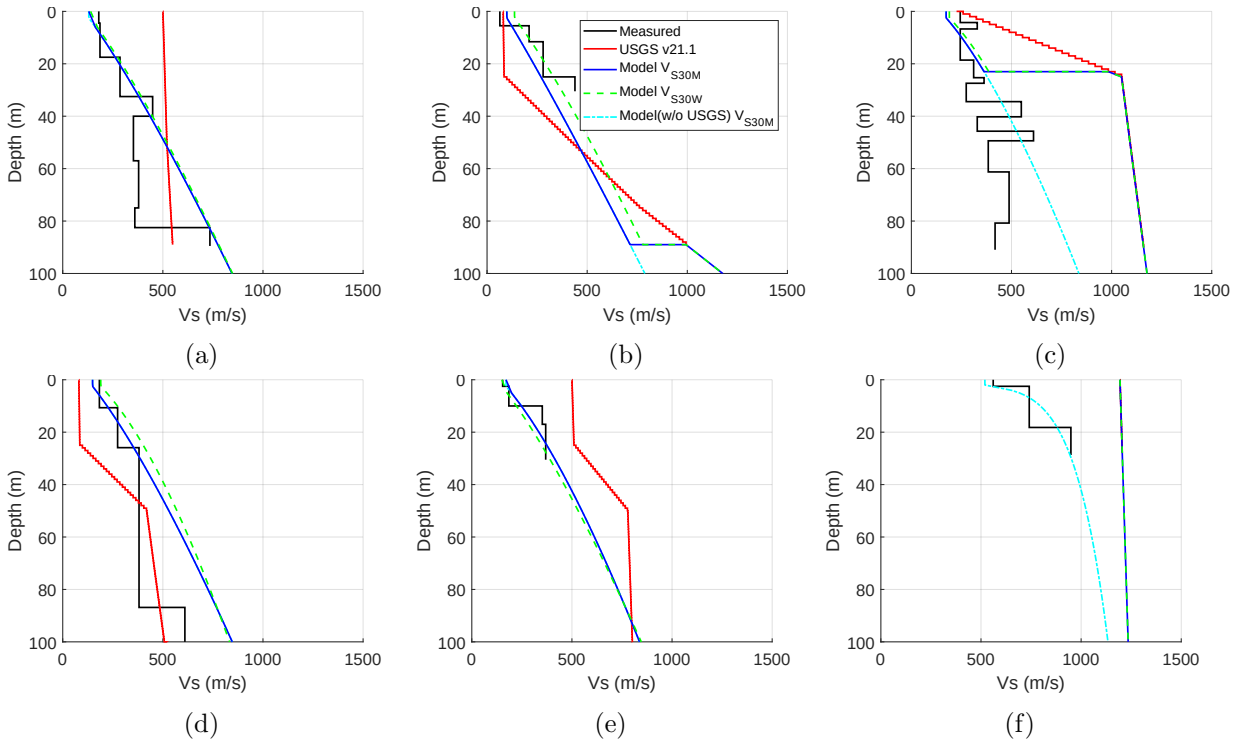


Figure 22: Comparison of measured Vs profiles to those predicted by the USGS SFVM, and the proposed model of the second approach. The measured profile and USGS SFVM profile are shown in red and black, respectively. Model V_{S30}M denotes using the measured V_{S30} to compute Vs; Model V_{S30}W denotes using the Wills V_{S30} to compute Vs. Model V_{S30}M and Model V_{S30}W profiles modify the USGS profiles up to $Z_{1.0}$. Model (w/o USGS) V_{S30}M denotes using measured V_{S30} to compute Vs over the whole depth.

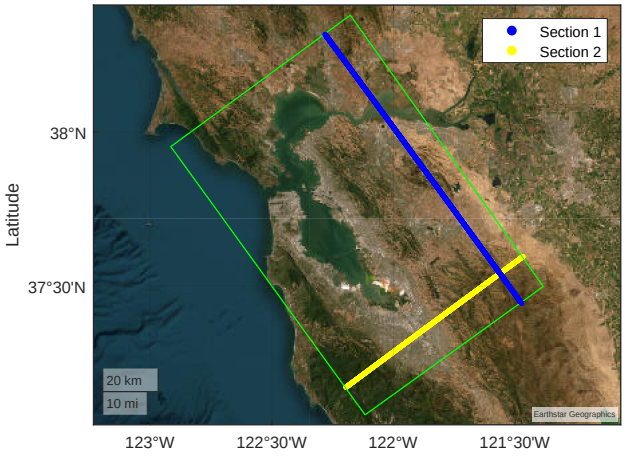
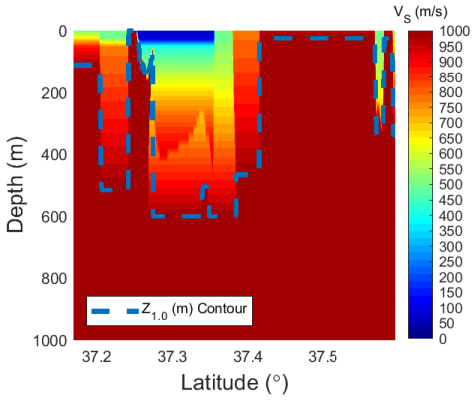
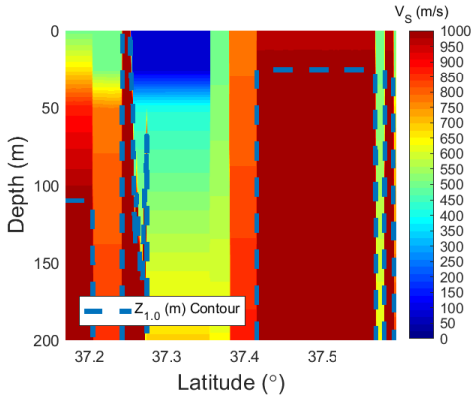


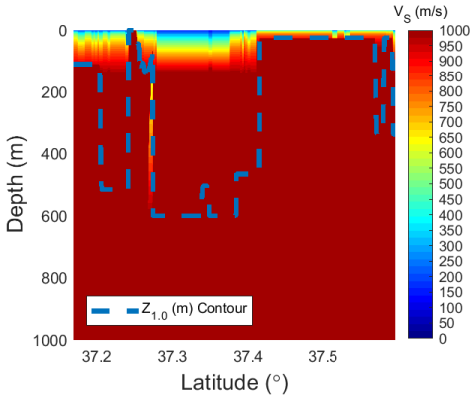
Figure 23: Locations of the two cross sections considered to develop 2D velocity structure using the proposed model and USGS SFVM.



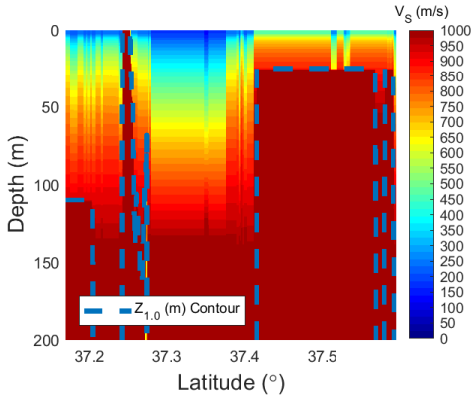
(a) USGS SFVM (up to 1 km)



(b) USGS SFVM (up to 200 m)

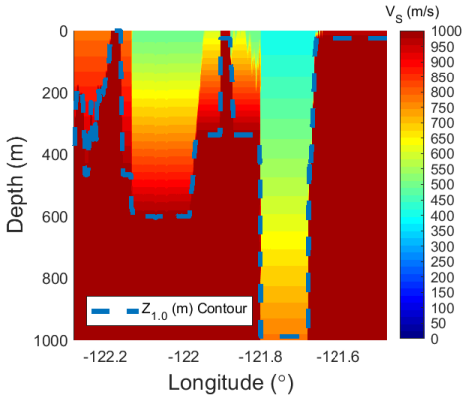


(c) Proposed model (up to 1 km)

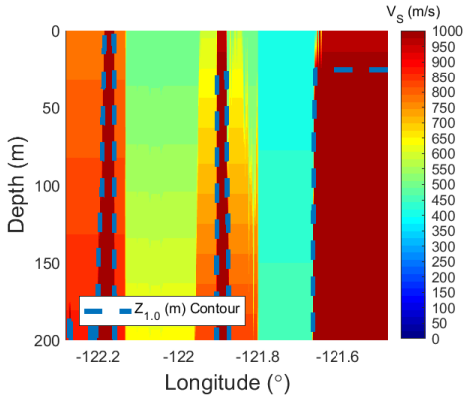


(d) Proposed model (up to 200 m)

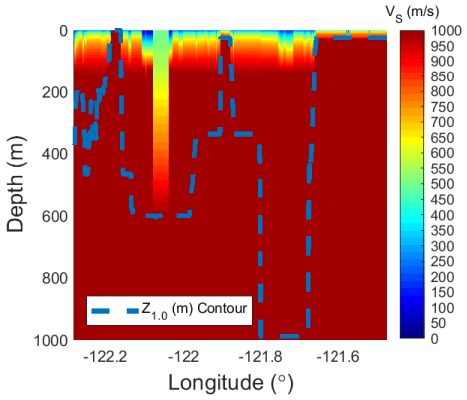
Figure 24: Comparison of 2D shear wave velocity structures for Section 1.



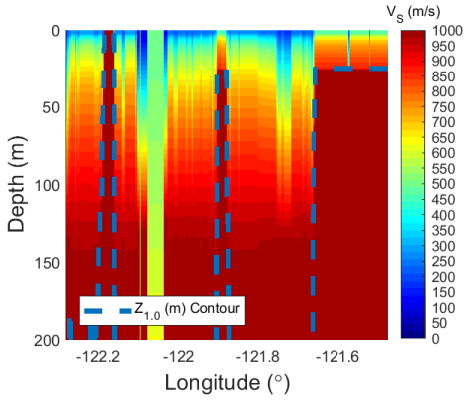
(a) USGS SFVM (up to 1 km)



(b) USGS SFVM (up to 200 m)

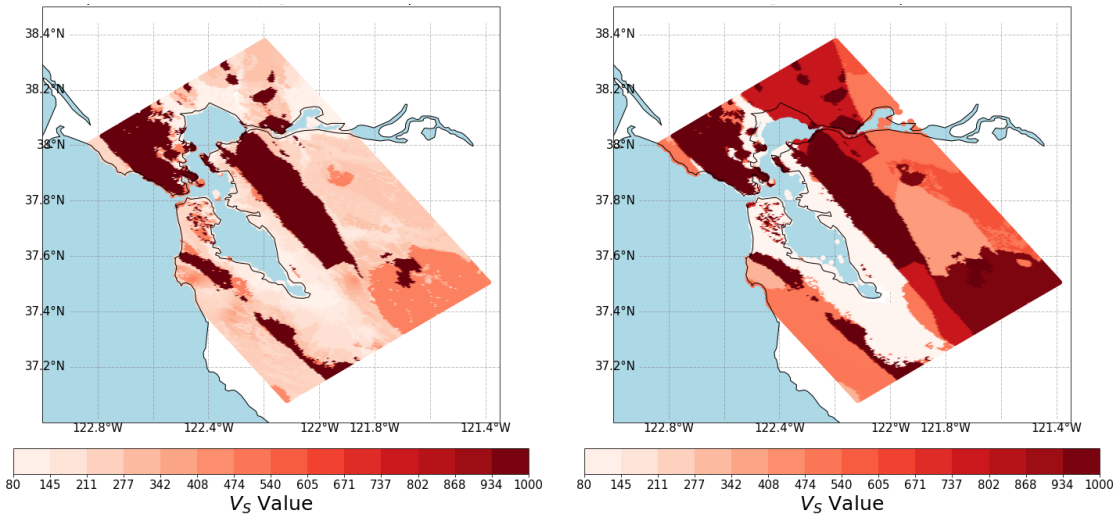


(c) Proposed model (up to 1 km)



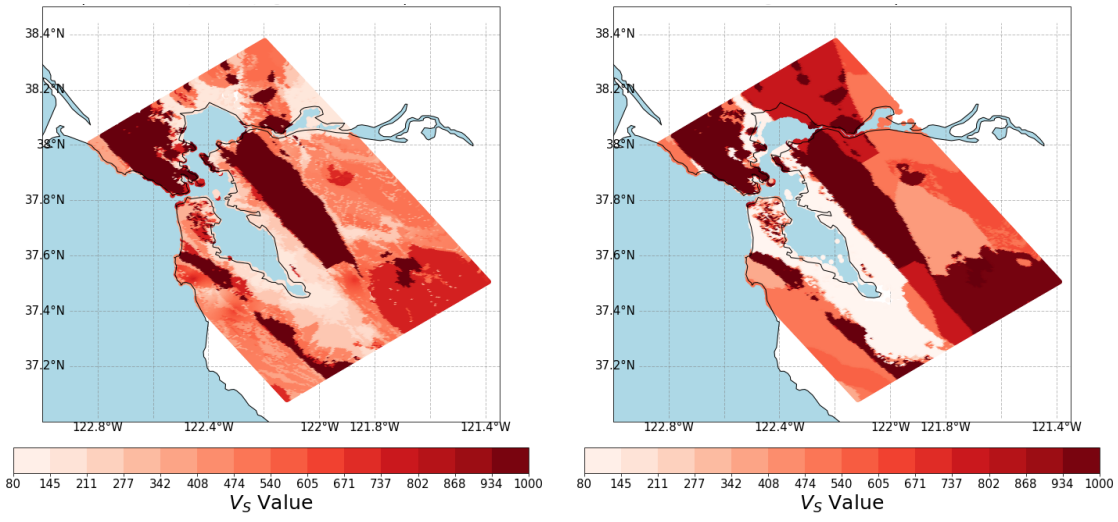
(d) Proposed model (up to 200 m)

Figure 25: Comparison of 2D shear wave velocity structures for Section 2.



(a) (b)

Figure 26: Vs map in the Bay region at $z = 1$ m using (a) the proposed model and (b) USGS SFVM.



(a) (b)

Figure 27: Vs map in the Bay region at $z = 10$ m using (a) the proposed model and (b) USGS SFVM.

7 Concluding remarks

This study presents a region-specific near-surface velocity model for the Bay Area region in Northern California. To this end, 211 profiles were gathered from different datasets, and two approaches were employed, referred to as the first and second approaches, to fit the profiles. In both approaches, the functional form (forward model) in Equation (8) proposed by Shi and Asimaki (2018) is used. Each approach aims at optimizing the hyperparameters related to the model parameters, k and n.

The first approach derived each profile independently, and the scaling relationships for k and n were estimated as a second step. This approach is similar to the method used by Marafi et al. (2021). The second approach used global Bayesian inference to determine the velocity model and directly capture the hyper-parameters of the scaling relationships. It is noted that the global Bayesian inference resulted in better constraining the hyper-parameters of k and n relationships and in less bias in the computed residuals.

The finalized proposed near-surface model is summarized as follows:

$$V_S(z) = \begin{cases} V_{S0} & z \leq 2.5 \\ V_{S0}(1 + k(z - 2.5))^{1/n} & z \geq 2.5 \end{cases} \tag{22}$$

in which

$$\begin{aligned} k &= \exp\left(\frac{r_1}{1 + r_2 {V_{S30}}^{-r_3}} + r_4\right), \\ n &= 1 + \frac{s_1}{1 + s_2 {V_{S30}}^{-s_3}}, \\ V_{S0} &= V_{S30} \frac{(1 + 27.5k)^{1/n} + 2.5k(1 - \frac{1}{n}) - 1}{30k(1 - \frac{1}{n})}. \end{aligned} \tag{23}$$

The unit of z is in m, V_{S30} is in km/s and $V_S(z)$ and V_{S0} are in m/s. Hyper-parameters s_1 to s_3 and r_1 to r_4 are repeated for convenience in the table below.

	r_1	r_2	r_3	r_4	s_1	s_2	s_3
5^{th} Percentile	2.741	$1.56e^{-4}$	8.512	2.356	6.391	0.057	3.784
25^{th} Percentile	3.141	$1.95e^{-4}$	8.758	2.404	6.900	0.066	3.915
50^{th} Percentile (Median)	3.423	$2.27e^{-4}$	8.932	2.435	7.274	0.072	4.005
75^{th} Percentile	3.715	$2.64e^{-4}$	9.111	2.467	7.659	0.078	4.097
95^{th} Percentile	4.115	$3.26e^{-4}$	9.365	2.512	8.246	0.089	4.227
Mean	3.426	$2.32e^{-4}$	8.936	2.435	7.291	0.072	4.006

A new relationship for V_{S0} is considered that ensures the calculated V_{S30} of the velocity model matches the input V_{S30} value and helps reduce the number of hyperparameters needed to be fitted. Additionally, the scaling relationships for k and n are designed to ensure well-posed extrapolation behavior for small and large V_{S30} values, and n is set to remain greater than one to avoid significantly large predictions at deep layers.

It is recommended to use the V_{S30} map by Wills et al. (2015) in computing Vs profiles using the proposed model because it shows better correlation to the V_{S30} of measured profiles in the Bay region. The maximum depths of most measured profiles were less than 100 m, and only a few profiles had $Z_{1.0}$ information. Therefore, the proposed model should be used cautiously in deeper layers. Future studies will determine the best strategy for integrating the proposed near-surface

model with the USGS SFVM. Additionally, future work will examine the proposed model further by performing 1D site response analysis and 3D ground motion simulations within the region.

References

Aagaard, B. T., Brocher, T. M., Dolenc, D., Dreger, D., Graves, R. W., Harmsen, S., Hartzell, S., Larsen, S., McCandless, K., Nilsson, S., et al. (2008a). Ground-motion modeling of the 1906 san francisco earthquake, part ii: Ground-motion estimates for the 1906 earthquake and scenario eventsground-motion modeling of the 1906 san francisco earthquake, part ii. Bulletin of the Seismological Society of America, 98(2):1012–1046.

Aagaard, B. T., Brocher, T. M., Dolenc, D., Dreger, D., Graves, R. W., Harmsen, S., Hartzell, S., Larsen, S., and Zoback, M. L. (2008b). Ground-motion modeling of the 1906 san francisco earthquake, part i: Validation using the 1989 loma prieta earthquake. Bulletin of the Seismological Society of America, 98(2):989–1011.

Aagaard, B. T., Graymer, R. W., Thurber, C. H., Rodgers, A. J., Taira, T., Catchings, R. D., Goulet, C. A., and Plesch, A. (2020a). Science plan for improving three-dimensional seismic velocity models in the san francisco bay region, 2019–24. Technical report.

Aagaard, B. T., Graymer, R. W., Thurber, C. H., Rodgers, A. J., Taira, T., Catchings, R. D., Goulet, C. A., and Plesch, A. (2020b). Science plan for improving three-dimensional seismic velocity models in the san francisco bay region, 2019–24. Technical report, US Geological Survey.

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.

Bas, E. E., Seylabi, E., Yong, A., Tehrani, H., and Asimaki, D. (2022). P-and s-wave velocity estimation by ensemble kalman inversion of dispersion data for strong motion stations in california. Geophysical Journal International, 231(1):536–551.

Bielak, J., Graves, R. W., Olsen, K. B., Taborda, R., Ramírez-Guzmán, L., Day, S. M., Ely, G. P., Roten, D., Jordan, T. H., Maechling, P. J., et al. (2010). The shakeout earthquake scenario: Verification of three simulation sets. Geophysical Journal International, 180(1):375–404.

Bielak, J., Taborda, R., Olsen, K., Graves, R., Silva, F., Khoshnevis, N., Savran, W., Roten, D., Shi, Z., Goulet, C., et al. (2016). Verification and validation of high-frequency (fmax= 5 hz) ground motion simulations of the 2014 m 5.1 la habra, california, earthquake. In AGU Fall Meeting Abstracts.

Boore, D. M. (2003). A compendium of p-and s-wave velocities from surface-to-borehole logging; summary and reanalysis of previously published data and analysis of unpublished data. Technical report, US Geological Survey.

Boore, D. M. and Asten, M. W. (2008). Comparisons of shear-wave slowness in the santa clara valley, california, using blind interpretations of data from invasive and noninvasive methods. Bulletin of the Seismological Society of America, 98(4):1983–2003.

Boore, D. M. and Brown, L. T. (1998). Comparing shear-wave velocity profiles from inversion of surface-wave phase velocities with downhole measurements: systematic differences between the

cxw method and downhole measurements at six usc strong-motion sites. Seismological Research Letters, 69:222–229.

Brocher, T., Aagaard, B., Simpson, R., and Jachens, R. C. (2006). The usgs 3d seismic velocity model for northern california. In AGU Fall Meeting Abstracts, volume 2006, pages S51B–1266.

Brocher, T. M. (2008). Compressional and shear-wave velocity versus depth relations for common rock types in northern california. Bulletin of the Seismological Society of America, 98(2):950–968.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). Handbook of markov chain monte carlo. CRC press.

Brown, L. T., Boore, D. M., and Stokoe, K. H. (2002). Comparison of shear-wave slowness profiles at 10 strong-motion sites from noninvasive sasw measurements and measurements made in boreholes. Bulletin of the Seismological Society of America, 92(8):3116–3133.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. Journal of statistical software, 76(1).

Ellsworth, W., Beroza, G., Julian, B., Klein, F., Michael, A., Oppenheimer, D., Prejean, S., Richards-Dinger, K., Ross, S., Schaff, D., et al. (2000). Seismicity of the san andreas fault system in central california: application of the double-difference location algorithm on a regional scale. Eos Trans. AGU, 81(48).

Ely, G. P., Jordan, T., Small, P., and Maechling, P. J. (2010). A vs30-derived nearsurface seismic velocity model. In Abstract S51A-1907, Fall Meeting. AGU San Francisco, CA.

Gao, F. and Han, L. (2012). Implementing the nelder-mead simplex algorithm with adaptive parameters. Computational Optimization and Applications, 51(1):259–277.

Graves, R. W. and Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. Bulletin of the Seismological Society of America, 100(5A):2095–2123.

Hanson, K. M. (2001). Markov chain monte carlo posterior sampling with the hamiltonian method. In Medical Imaging 2001: Image Processing, volume 4322, pages 456–467. SPIE.

Hirakawa, E. and Aagaard, B. (2022). Evaluation and updates for the usgs san francisco bay region 3d seismic velocity model in the east and north bay portions. Bulletin of the Seismological Society of America, 112(4):2070–2096.

Kwak, D., Ahdi, S. K., Wang, P., Zimmaro, P., Brandenberg, S. J., Stewart, J. P., et al. (2021). Web portal for shear wave velocity and hvsr databases in support of site response research and applications.

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

Moré, J. J. and Sorensen, D. C. (1983). Computing a trust region step. SIAM Journal on scientific and statistical computing, 4(3):553–572.

Olsen, K., Day, S., and Bradley, C. (2003). Estimation of q for long-period (¿ 2 sec) waves in the los angeles basin. Bulletin of the Seismological Society of America, 93(2):627–638.

Olsen, K., Day, S., Minster, J., Cui, Y., Chourasia, A., Faerman, M., Moore, R., Maechling, P., and Jordan, T. (2006). Strong shaking in los angeles expected from southern san andreas earthquake. Geophysical Research Letters, 33(7).

Rix, G. J., Hebeler, G. L., and Orozco, M. C. (2002). Near-surface vs profiling in the new madrid seismic zone using surface-wave methods. Seismological Research Letters, 73(3):380–392.

Rodgers, A. J., Anders Petersson, N., Pitarka, A., McCallen, D. B., Sjogreen, B., and Abrahamson, N. (2019). Broadband (0–5 hz) fully deterministic 3d ground-motion simulations of a magnitude 7.0 hayward fault earthquake: Comparison with empirical ground-motion models and 3d path and site effects from source normalized intensities. Seismological Research Letters, 90(3):1268–1284.

Rodgers, A. J., Pitarka, A., Petersson, N. A., Sjögreen, B., and McCallen, D. B. (2018). Broadband (0–4 hz) ground motions for a magnitude 7.0 hayward fault earthquake with three-dimensional structure and topography. Geophysical Research Letters, 45(2):739–747.

Roten, D., Olsen, K., Day, S., Cui, Y., and Fäh, D. (2014). Expected seismic shaking in los angeles reduced by san andreas fault zone plasticity. Geophysical Research Letters, 41(8):2769–2777.

Savran, W. and Olsen, K. (2016). Model for small-scale crustal heterogeneity in los angeles basin based on inversion of sonic log data. Geophysical Journal International, 205(2):856–863.

Seylabi, E., Stuart, A., and Asimaki, D. (2019). Data fusion and assimilation framework for site characterization. To be submitted to Bulletin of the Seismological Society of America.

Shi, J. and Asimaki, D. (2018). A generic velocity profile for basin sediments in california conditioned on vs30. Seismological Research Letters, 89(4):1397–1409.

Shi, Z. and Day, S. M. (2013). Rupture dynamics and ground motion from 3-d rough-fault simulations. Journal of Geophysical Research: Solid Earth, 118(3):1122–1141.

Small, P., Gill, D., Maechling, P. J., Taborda, R., Callaghan, S., Jordan, T. H., Olsen, K. B., Ely, G. P., and Goulet, C. (2017). The scec unified community velocity model software framework. Seismological Research Letters, 88(6):1539–1552.

Stephenson, W. J., Louie, J. N., Pullammanappallil, S., Williams, R., and Odum, J. K. (2005). Blind shear-wave velocity comparison of remi and masw results with boreholes to 200 m in santa clara valley: implications for earthquake ground-motion assessment. Bulletin of the Seismological Society of America, 95(6):2506–2516.

Taborda, R. and Bielak, J. (2013). Ground-motion simulation and validation of the 2008 chino hills, california, earthquake. Bulletin of the Seismological Society of America, 103(1):131–156.

Taborda, R. and Bielak, J. (2014). Ground-motion simulation and validation of the 2008 chino hills, california, earthquake using different velocity models. Bulletin of the Seismological Society of America, 104(4):1876–1898.

Taborda, R., Bielak, J., and Restrepo, D. (2012). Earthquake ground-motion simulation including nonlinear soil effects under idealized conditions with application to two case studies. Seismological Research Letters, 83(6):1047–1060.

Thompson, E., Wald, D. J., and Worden, C. (2014). A vs30 map for california with geologic and topographic constraintsa vs30 map for california with geologic and topographic constraints. Bulletin of the Seismological Society of America, 104(5):2313–2321.

Toro, G. (1995). Probabilistic models of site velocity profiles for generic and site-specific ground-motion amplification studies. Technical Rep, 779574.

Wald, D. J. and Allen, T. I. (2007). Topographic slope as a proxy for seismic site conditions and amplification. Bulletin of the Seismological Society of America, 97(5):1379–1395.

Waldhauser, F. and Ellsworth, W. L. (2000). A double-difference earthquake location algorithm: Method and application to the northern hayward fault, california. Bulletin of the seismological society of America, 90(6):1353–1368.

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(4A):1483–1501.

Wills, C., Gutierrez, C., Perez, F., and Branum, D. (2015). A next generation vs30 map for california based on geology and topographya next generation vs30 map for california based on geology and topography. Bulletin of the Seismological Society of America, 105(6):3083–3091.

Yong, A., Martin, A., Stokoe, K., and Diehl, J. (2013). Arra-funded v s30 measurements using multi-technique approach at strong-motion stations in california and central-eastern united states. Technical report, US Geological Survey.

A Bayesian Regression Overview

Bayesian regression provides the framework to combine our prior assumptions about the range of the model parameters with the evidence in the available data to come up with an updated range on the distribution of the model parameters using Bayes's rule:

$$p(\theta, \theta_{hyp}|x, y) = \frac{L(\theta, \theta_{hyp})p(\theta, \theta_{hyp})}{p(y, x)} \tag{24}$$

where $p(\theta, \theta_{hyp})$ is the prior distribution describing the uncertainty of the model parameters θ (e.g., V_{S0}, k, and n) and model hyperparameters θ_{hyp} (e.g., s_1, s_2, s_3) before observing the data. $L(\theta, \theta_{hyp})$ is the likelihood function that quantifies the evidence in the data for how plausible the different values of θ and θ_{hyp} are. $p(\theta, \theta_{hyp}|y, x)$ is the posterior distribution that combines our prior assumptions with the influence of the data.

However, to improve the computational efficiency, $p(y, x)$ is removed as it does not affect the shape of the distribution. In this case, the posterior is represented as:

$$p(\theta, \theta_{hyp}|x, y) \propto L(\theta, \theta_{hyp})p(\theta, \theta_{hyp}) \tag{25}$$

The information for defining the prior distributions may come from different sources, including prior fittings (i.e., approach 1) or previous studies such as Shi and Asimaki (2018). When less prior information is available, meaning that there is less confidence in the range of parameters before looking at the data, wider prior distributions are used, which assigns more weight to the likelihood function to determine the posterior distributions

Due to the high computational cost of analytically calculating Equation (24), multiple numerical methods have been employed for expressing $p(\theta, \theta_{hyp}|x, y)$. The most commonly used method is Markov-Chain Monte Carlo (MCMC) which draws samples from the posterior distribution to compute various statistics $p(\theta, \theta_{hyp}|x, y)$. Additional information on MCMC and its implementation on the statistical software STAN (Carpenter et al., 2017) is provided in the remainder of this section.

Bayesian Inference using MCMC

Bayesian inference performed by MCMC is based on sampling from the posterior distribution. If a set of N_k random parameters arrays (x_k) is derived from a probability density function (pdf) of $q(x)$, the expectation of each function, $f(x)$, can be estimated using:

$$\langle f(x) \rangle = \int f(x)q(x)dx \approx \frac{1}{N_k}\Sigma_{k=1}^{N_k} f(x_k)\,. \tag{26}$$

Given a set of posterior samples, the posterior mean of the parameters, $\overline{x}$ is derived using the function $f(x) = x$, and the variance of the parameters is obtained with $f(x) = (x - \bar{x})^2$.

The two main families of MCMC samplers used in Bayesian analysis are Gibbs and Metropolis algorithms. Given the simplicity in the execution of the Metropolis algorithm, it is more common, but in problems with a high number of parameters, it is not as efficient as the Gibbs sampler. A more promising MCMC implementation based on the Metropolis sampler is the Hamiltonian method, which is based on analogy and physical systems. It was first developed to simulate a system of particles, each of which is associated with a position and momentum. In this technique, each parameter, x_i, has momentum, p_i, which participates in the construction of a Hamiltonian as

a combination of both the potential energy term, $\phi = -\log(q(x))$, and the kinetic energy term, $\frac{p_i^2}{2m_i}$. This technique aims to draw random samples from the new pdf proportional to $exp(-H)$. This method is employed in the software STAN, and more detailed information can be found at: Brooks et al. (2011); Hanson (2001).

$$H = \phi(x) + \Sigma_{i=1}^{N} \frac{p_i^2}{2m_i} \tag{27}$$

where H is Hamiltonian, N is the number of particles.