EVAULATION OF THE P-WAVE SEISMOGRAM APPROACH TO ESTIMATE Vs30

FINAL TECHNICAL REPORT

SUBMITTED: August 2021

Ellen M. Rathje, Ph.D., P.E.
Meibai Li

University of Texas at Austin
Austin, TX 78712
Tel: 512-232-3683
Fax: 512-471-6548
e.rathje@mail.utexas.edu

AWARD PERIOD
September 2019-May 2021

ABSTRACT

The P-wave seismogram method is utilized to estimate the V_{S30} of 194 stations in California. Comparison between the V_{S30} measurement shows that the measured and estimated V_{S30} of 77% stations are within ±50% range of measurement, which validates the performance of the P-wave seismogram method in estimating V_{S30}. However, stations with average measured V_{S30} greater than 400 m/s may be potentially overestimated by the P-wave seismogram method. Comparison between the residual obtained by this study and by the geology and slope based V_{S30} map suggests that less dispersion relative to measured V_{S30} is achieved by the P-wave seismogram method than the proxy-based method, but the V_{S30} predicted by the P-wave seismogram method may be more overestimated when the estimated V_{S30} is greater than about 400 m/s. The effect of using a more detailed crustal model is examined, and the result shows that the V_{S30} tends to be underestimated for events with focal depth smaller than 2.5 km when the more detailed crustal model is used, which is because of the layers with low P-wave velocity and steeper gradient of P-wave velocity at shallow depths in the more detailed crustal model.

INTRODUCTION

The soils overlying bedrock can significantly amplify earthquake ground motions; therefore, characterization of the subsurface is necessary to accurately capture site effects when predicting earthquake ground motion. The parameter V_{S30}, which represents the time-averaged shear wave velocity over the top 30m, has become a standard site parameter that is widely used in a number of ground motion prediction equations (GMPEs, e.g., Boore et al. 2014, Abrahamson et al. 2014, Chiou and Youngs 2014, Idriss 2014) and building codes (e.g., American Society of Civil Engineers ASCE/SEI *7-16* 2016, European Committee for Standardization [CEN] 2013) to account for site effects.

The parameter V_{S30} is most commonly computed from a shear wave velocity profile that is measured in-situ using geophysical field tests, such as downhole testing, cross-hole testing, suspension logging, the seismic cone penetration test (SCPT), Multichannel Analysis of Surface Waves (MASW), Spectral Analysis of Surface Waves (SASW), or Microtremor array measurements (MAM). Although the database of in-situ measurements of near surface shear wave velocity is continuously expanding, V_{S30} measurements are currently not available for a large majority of the ground motion recording stations in the US. To estimate the V_{S30} for sites that lack a shear wave velocity profile, researchers have been using proxy methods to correlate V_{S30} with various relevant parameters such as topographic slope (e.g., Wald and Allen 2007, Allen and Wald 2009), terrain (e.g., Iwahashi and Pike 2007, Yong 2016), surficial geology and geotechnical indices (e.g., Wills and Clahan 2006, Kottke et al. 2012), or combinations of multiple parameters (e.g., Kwak et al. 2015, Wills et al. 2015, Parker et al. 2017). Alternatively, Ni et al. (2014) proposed the P-wave seismogram method to estimate the near surface shear wave

velocity using seismograms recorded at ground motion recording stations and the physics of wave propagation.

Compared to the proxy methods, the P-wave seismogram method has the advantage of using information that is more site-specific. With only earthquake recording from small/moderate earthquakes and the crustal velocity model of a region, the P-wave seismogram method can be rapidly applied to estimate the V_{S30} at a site. The P-wave seismogram method has been used to characterize recording stations in central and eastern North America (CENA, Kim et al. 2016, Zalachoris et al. 2017) and Japan (Miao et al. 2018, Kang et al. 2020), and it shows promising performance when compared with the in-situ measurements. To further test the reliability of the P-wave seismogram method, more studies need to be conducted in regions other than CENA and Japan.

In this study, we take advantage of the abundant V_{S30} dataset in California to evaluate the accuracy of the P-wave seismogram method to estimate V_{S30}. We estimate the V_{S30} for 194 seismic recording stations in California, statistically compare the results with existing in-situ measurements and evaluate the potential use of a more detailed crustal model.

THEORETICAL BASIS OF P-WAVE SEISMOGRAM METHOD

The physical basis of the P-wave seismogram method can be qualitatively explained by Snell's Law, which states that body waves traveling through a low-velocity soil are refracted to a more vertical position. As a result, the ratio between the radial and vertical components of the body wave becomes smaller as the site softens, and this ratio of radial to vertical components of motion at the ground surface can be used to infer the near surface velocity of a site. Generally, a

larger ratio of the radial-to-vertical components indicates a larger surface velocity, i.e., a stiffer site.

In the P-wave seismogram method, the ratio between the radial and vertical components of an incident P-wave is related to the near surface shear wave velocity. As shown in Figure 1, the motion at the ground surface caused by an incident P-wave has three components: the up-going incident P-wave, the down-going reflected P-wave and the down-going reflected SV-wave.
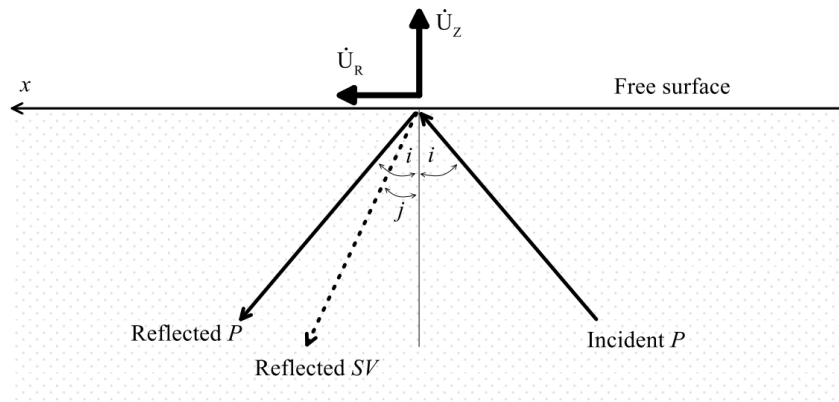

Figure 1. Geometry of the incident and reflected particle motions at the ground surface caused by an incident P-wave

Aki and Richards (2002) derived expressions for the displacements in the radial, tangential, and vertical directions (i.e., U_R, U_T, U_Z) for an incident P-wave of angular frequency ω as:

$$[U_R, U_T, U_Z] = \frac{\bar{P} \cdot [R,T,Z] \cdot \exp\left[i\omega(px-t)\right]}{\left(\frac{1}{V_S^2} - 2p^2\right)^2 + 4p^2 \frac{\cos i \cdot \cos j}{V_P \quad V_S}} \tag{1a}$$

$$R = \frac{4V_P \cdot p}{V_S^2} \frac{\cos i}{V_P} \frac{\cos j}{V_S} \tag{1b}$$

6

$$T = 0 \tag{1c}$$

$$Z = \frac{-2V_P}{V_S^2} \frac{\cos i}{V_P} \left(\frac{1}{V_S^2} - 2p^2\right) \tag{1d}$$

where $\bar{P}$ is the incident P-wave amplitude, t is time, V_S and V_P are the shear (S)-wave and

compression (P)-wave velocities of the surface layer, i is the incident angle of the P-wave from

the vertical, j is the angle of the reflected SV-wave from the vertical (Figure 1), and p is the ray

parameter. The ray parameter p can be expressed using equation (2) assuming a plane wave.

$$p = \frac{\sin i}{V_P} = \frac{\sin j}{V_S} \tag{2}$$

The equations for R and Z can be used to define the ratio between the radial component (i.e.,

direction pointing from the site to earthquake epicenter) and vertical component of motion, and

this ratio is applicable to the displacement, velocity, and acceleration. The resulting expression

written in terms of velocity is:

$$\frac{\dot{U}_R}{\dot{U}_Z} = \frac{R}{Z} = \frac{-2 \cdot V_S \cdot p \cdot \cos j}{1 - 2 \cdot p^2 V_S^2} \tag{3a}$$

Substituting $p = \frac{\sin j}{V_S}$ as defined in equation (2), equation (3a) can be eventually simplified to:

$$\frac{\dot{U}_R}{\dot{U}_Z} = \tan 2j \tag{3b}$$

Rearranging equation (2) to solve for V_S and substituting equation (3b) for j, the shear-wave

velocity of the surface layer is subsequently derived as (Park and Ishii, 2018):

$$V_S = \frac{\sin j}{p} = \frac{\sin\left(0.5 \cdot \tan^{-1}\left[\frac{\dot{U}_R}{\dot{U}_Z}\right]\right)}{p} \tag{4}$$

Thus, the shear-wave velocity of the surface layer can be computed given the ray parameter (p), which can be estimated from a crustal velocity model, and the ratio of the radial to vertical velocity amplitude of the incident P-wave $\left(\frac{\dot{U}_R}{\dot{U}_Z}\right)$, which can be obtained from the initial part of a seismic record.

As explained in Kim et al. (2016), because the radial and vertical components of the P-wave are usually measured over a finite time window, the estimated V_S is not exactly the shear-wave velocity at a depth of zero but is an average shear-wave velocity from the ground surface to a depth extent that is limited by the wavelength of the near surface shear-wave. Ni et al. (2014) conducted numerical simulations to show that the depth extent resolved by the inferred subsurface shear-wave velocity V_S is approximately one wavelength. Therefore, in this study, the same assumption is made as in Kim et al. (2016) that the estimated V_S represents the time-averaged shear-wave velocity of the upper z meter, where z is the wavelength of the shear wave, and V_S can thus be written as V_{SZ}. The wavelength z depends on the earthquake source duration (Ni et al. 2014) and can be calculated as the product of the source time function and the estimated shear wave velocity, i.e., $z = \tau_p \cdot V_{SZ}$. Kim et al. (2016) recommended that $\tau_p = 0.1s$, which is appropriate for magnitude 3-4 events. After V_{SZ} is obtained, V_{S30} can be estimated using empirical correlations between V_{SZ} and V_{S30}, as discussed in a later section.

Estimation of ray parameter p

To compute V_{SZ} from equation (4), the ray parameter p needs to be obtained. The ray parameter remains constant along a ray path and can be derived with knowledge of the location of two points on the ray path and the velocity structure. In this study, the points on the ray path are the earthquake hypocenter and the station location (Figure 2), and a one-dimensional (1-D) P-wave velocity model is used. The definition of p in equation (2) is applied to the interface

8

between each adjacent layer pair $(m, m-1)$, with the incident angle in layer (i_m) for each layer m defined geometrically from the horizontal distance traveled in the layer (R_m) and the thickness of the layer (D_m):

$$p = \frac{\sin{(i_m)}}{V_{P,m}} = \frac{\sin{(i_{m-1})}}{V_{P,m-1}} \qquad \text{for} \quad 2 \leq m \leq n \qquad (5a)$$

$$p = \frac{\sin{(\tan^{-1}\frac{R_m}{D_m})}}{V_{P,m}} = \frac{\sin{(\tan^{-1}\frac{R_{m-1}}{D_{m-1}})}}{V_{P,m-1}} \qquad \text{for} \quad 2 \leq m \leq n \qquad (5b)$$

This expression is applied to layer interfaces that occur between the ground surface and the hypocentral depth of the earthquake (D_{hyp}). Applying equation (5b) to each layer pair results in $n-1$ equations but n unknown values of R_m (note that D_m and V_{Pm} are known from the velocity model). The additional constraint that is used to solve the problem is that the sum of the R_m values is equal to the epicentral distance (R_{epi}):

$$R_{epi} = \sum_{m=1}^{n} R_m = R_1 + R_2 + R_3 + \cdots + R_n \qquad (6)$$

The unknown values of R_m for each layer are solved for numerically using equation 5b and 6. After R_m for each layer is solved, the ray parameter can be computed using equation (5b).
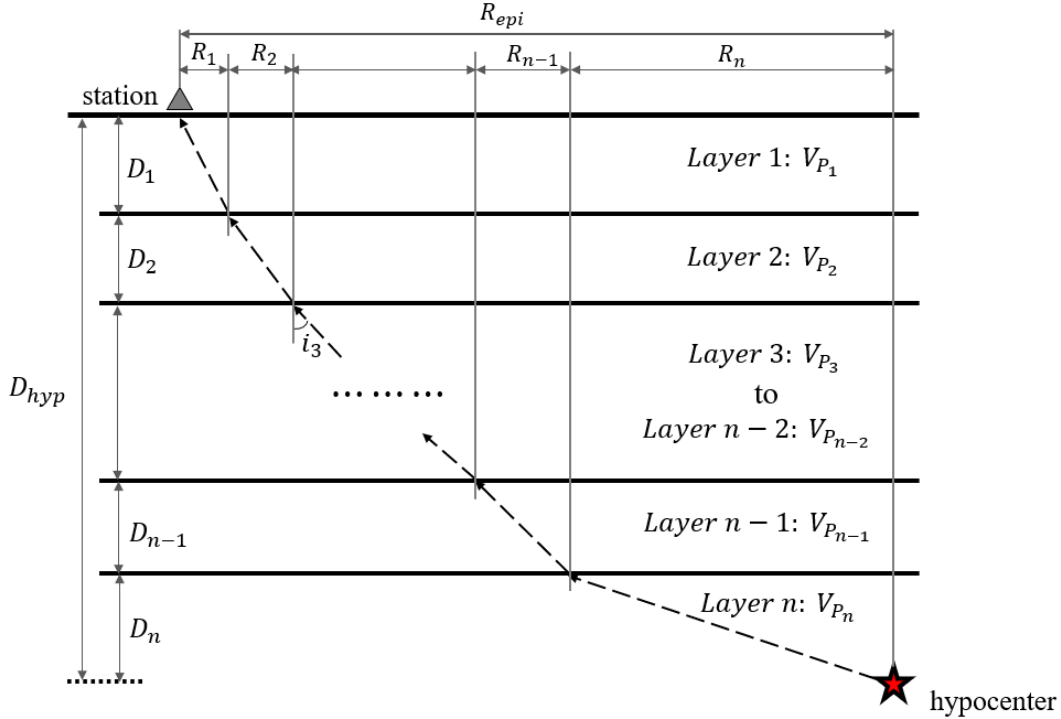
Figure 2. Ray path for earthquake-generated P-waves in a one-dimensional crustal model

DATA ANALYSIS PROCEDURES FOR P-WAVE SEISMOGRAM METHOD

In this study, 1923 recordings from 824 earthquake events are used to estimate V_{S30} of 194 ground motion recording stations in California, and the locations of the seismic stations and events analyzed are shown in Figure 3. The magnitude of the events considered varies from 2.5 to 5, so that the assumption of source time function $\tau_p = 0.1s$ is appropriate, the focal depth of the events is less than 30 km, and the epicentral distance of the events are generally less than 200 km to avoid noisy records. Ideally, at least 10 records should be analyzed for each station, but for some sites the number of records is less than 10 due to an insufficient number of records with a clear P-wave arrival. However, for all 194 seismic stations analyzed, at least 3 records are utilized to estimate the V_{S30}.
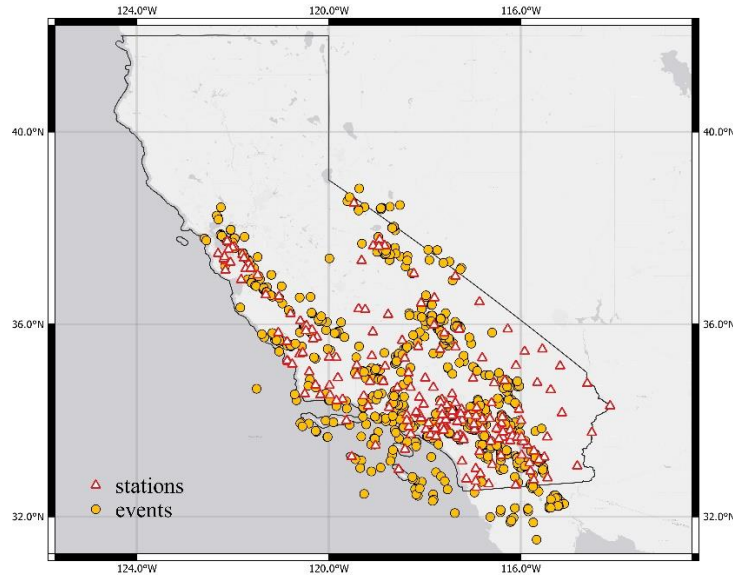
Figure 3. Location of the seismic recording stations for which V_{S30} is estimated and location of the earthquake events used in the P-wave seismogram method analysis

For most seismic stations (typically from the Southern California Seismic Network (CI) and Northern California Seismic Network (NC)), we used the data retrieving software Standing Order for Data (SOD, Owens et al. 2004) to obtain ground motion recordings from the IRIS Data Management Center (https://ds.iris.edu/ds/nodes/dmc/). For stations in the California Strong Motion Instrumentation Program (CE) and United States National Strong-Motion Network (NP), the ground motion recordings were obtained from the Center for Engineering Strong Motion Data (CESMD: https://www.strongmotioncenter.org/aboutcesmd.html). Before being used in the P-wave seismogram method, the ground motion recordings are instrument corrected, baseline corrected, and bandpass filtered with corner frequencies of 0.3 Hz and 25 Hz, which is considered appropriate for events with local magnitude below 5.5 (Jones et al. 2017). We rejected noisy records based on initial Signal to Noise Ratio check as well as visual inspection.

The P-wave seismogram method is applied in the following steps for each seismogram: (1) compute the ratio of radial to vertical components of ground motion $\left(\frac{\dot{U}_R}{\dot{U}_Z}\right)$ from the velocity time series; (2) estimate ray parameter p from event metadata and a local crustal velocity model; (3) estimate V_{SZ} from $\frac{\dot{U}_R}{\dot{U}_Z}$ and p using equation (4); and (4) convert V_{SZ} to V_{S30} with an empirical V_{SZ} to V_{S30} correlation. These four steps are described in detail in the following section.

Selection of $\dot{U}_R/\dot{U}_Z$ from recordings

The radial and vertical components of the P-wave arrival are identified from the velocity time series. The horizontal components of the ground motions are first rotated to the azimuth associated with the epicenter of the event to obtain the radial component. Then, the first peak after the P-wave arrival is selected on the vertical component, and the ratio between the radial and vertical components at this time is recorded. It should be noted that only records with acceptable signal to noise ratio and clear first peaks in both the vertical and radial components are used for analysis to minimize the uncertainty associated with the procedure of picking a peak. Some examples are provided below to illustrate the criteria used in this study for record selection.

All records used for the P-wave seismogram analysis need to be of high signal-to-noise ratio to ensure that the first peak associated with the P-wave arrival can be unambiguously identified. In addition, records with insufficient baseline correction need to be avoided, because it can be difficult to identify the P-wave arrival if the velocity time series are not about zero at the beginning of the time series. To increase the precision in computing the ratio between the radial and vertical ground motion, we required apparent first peaks to occur at approximately the same time for both radial and vertical components. The time of the first peak in the vertical component is used to select the time to compute $\frac{\dot{U}_R}{\dot{U}_Z}$.
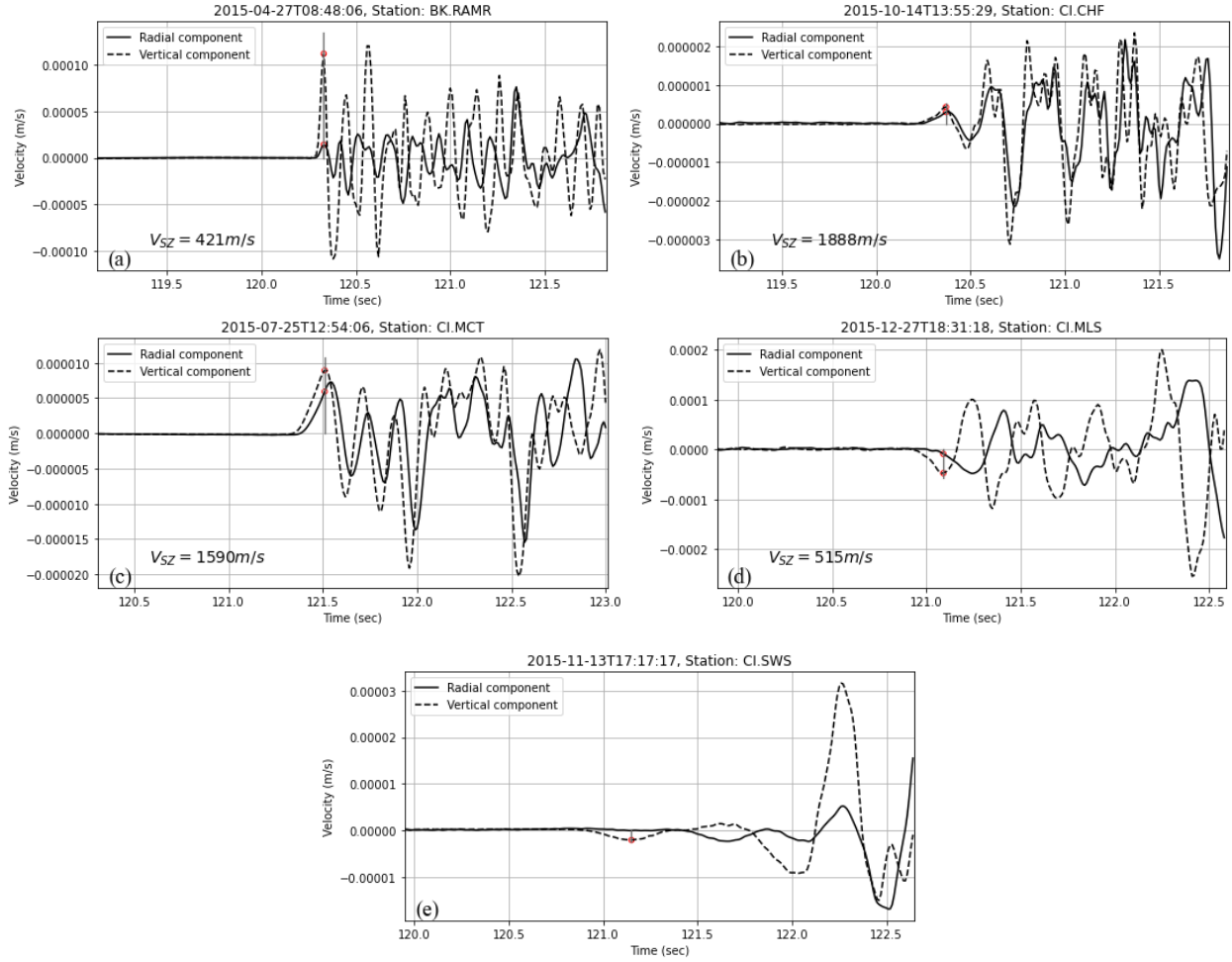
12

Figure 4. Ground motion velocity time series for the radial and vertical components recorded at stations (a) BK.RAMR, (b) CI.CHF, (c) CI.MCT, (d) CI.MLS, and (e) CI.SWS.

Figures 4a and 4b show recordings where the selection of the P-wave arrival is unambiguous. The velocity time series are shown for a magnitude 3.05 event recorded at station BK.RAMR (Figure 4a) and for a magnitude 3.47 event recorded at station CI.CHF (Figure 4b). For both recordings the peaks in the radial and vertical components occur essentially at the same time, but the relative amplitudes of the radial and vertical components are different for the two recordings. For BK.RAMR (Figure 4a) the vertical component is much larger than the radial, indicating that the P-wave has refracted to a more vertical position when traveling through softer soil layers. The selected peak amplitudes result in a smaller $\frac{\dot{U}_R}{\dot{U}_Z}$ and a V_{SZ} estimate of 421 m/s.

13

For CI.CHF (Figure 4b) the radial peak has almost the same amplitude as the vertical peak, resulting in a larger $\frac{\dot{U}_R}{\dot{U}_Z}$ and a larger V_{SZ} estimate of 1,888 m/s.

Figures 4c through 4e show cases where the initial peaks in the radial and vertical components occur at different times. In Figure 4c (station CI.MCT) the first peak in the radial component lags about 0.05 s behind the first peak in the vertical component. Because the time difference between the two peaks is relatively short compared to the width of the peaks, we accept the $\frac{\dot{U}_R}{\dot{U}_Z}$ derived from velocity time series in this case. For the motion in Figure 4c, the amplitudes of the radial and vertical peaks are similar, resulting in a V_{SZ} estimate of 1590 m/s. Figure 4d shows an example where the time difference between the first peaks in the radial and vertical components is relatively large compared to the width of the peak. Large time differences in the peaks will always result in small values of $\frac{\dot{U}_R}{\dot{U}_Z}$ and V_{SZ}. Because of the uncertainty about whether these peaks represent the same wave arrival, we reject motions that show this characteristic. Similarly, we reject motions where no identifiable peak is observed in the radial component over the time period associated with the peak in the vertical component (Figure 4e). Admittedly there is some subjectivity in deciding which motions to reject, but the examples in Figure 4 demonstrate the process used to accept and reject motions for P-wave seismogram analysis.

Estimating Ray Parameter p and V_{SZ}

To calculate V_{SZ}, the ray parameter p is needed in addition to the ratio $\frac{\dot{U}_R}{\dot{U}_Z}$. We estimate the ray parameter assuming the crustal model for southern California as utilized in Wald et al. (1995). The velocity structure of the crustal model is summarized in Table 1. With the approach

described earlier (equations 5b and 6), the ray parameter between the earthquake and recording station is estimated given the hypocentral depth (D_{hyp}), epicentral distance (R_{epi}), and velocity model. Given $\left(\frac{\dot{U}_R}{\dot{U}_Z}\right)$ and the estimated value of p, V_{SZ} is computed using equation (4) for each record.

Table 1. Crustal model of southern California region used in this study

P-wave velocity (km/s)	Depth to top of layer (km)
5.5	0.0
6.3	5.5
6.7	16.0
7.8	32.0

Relationship between V_{SZ} and V_{S30}

As noted earlier, the V_{SZ} estimated by the P-wave seismogram method is taken to represent the time-averaged shear wave velocity down to a depth of approximately $z = 0.1\,\text{s} \cdot V_{SZ}$. Therefore, a conversion between V_{SZ} and V_{S30} is needed to obtain V_{S30}. Using sites from the Kiban-Kyoshin network (KiK-net) in Japan, Boore et al. (2011) observed that smaller values of V_{S30} tend to correspond to smaller values of V_{SZ} at a given depth z. Boore et al. (2011) found that this correlation between V_{S30} and V_{SZ} is applicable for z as deep as several hundred meters and developed V_{S30} to V_{SZ} relationships for $z > 30m$ and up to 600 m using V_S profiles from KiK-net stations. V_{S30} to V_{SZ} relationships should be somewhat regional based on local and regional geology, but we expect a similar type of relationship between V_{SZ} and V_{S30} for California. Because a relationship between V_{SZ} and V_{S30} is not available for California for depths greater than 30m, we develop a relationship utilizing the shear wave velocity profiles available in the

shear-wave velocity profile database (VSPDB) developed by Ahdi et al. (2017). Similar to Boore et al. (2011), a linear relationship is used between the logarithms of V_{S30} and V_{SZ}:

$$ln(V_{S30}) = c_0 + c_1 \cdot ln(V_{SZ}) \tag{7}$$

This function is fit to the data for depths ranging from 5 to 400m.

There are 1020 V_S profiles in California that are available in the VSPDB, but the profiles at some sites do not reach a depth of 30m. Additionally, some sites have more than one measured V_S profile in the database. For example, sites measured by surface wave testing by Thompson et al. (2010) are provided with two different inversion solutions with different complexity. In that case, we followed the recommendation of the authors for which V_S profile to use, and when no recommendation is given, we use the profile that extends to the greatest depth or has more detailed layers. The V_S profiles measured by Refraction Microtremor (ReMi) are not considered due to the potential intra-method uncertainty of the method (Cox and Beekman 2011). After removing V_S profiles that extended less than 30m, excess V_S profiles at the same site, and poor quality profiles where V_S was missing in one or more layers, 680 V_S profiles were considered in the development of the V_{SZ} - V_{S30} relationship. Unfortunately, California lacks deeper V_S profiles, with only 78 V_S profiles reaching 100m and 12 V_S profiles reaching 200m. This lack of data imposes a challenge for developing the relationship between V_{SZ} and V_{S30} for depths greater than 100m. To increase the available data for developing the relationships for larger depths, we extend the V_S of the base layer to a depth of 400 m for the profiles that reach 100m. Extrapolating the V_S profiles introduces uncertainty in the V_{SZ} - V_{S30} relationship but that uncertainty is minimized by only extrapolating profiles that extend 100 m. These 100 m profiles tend to be stiffer at that depth and the depth of extrapolation is minimized. Nonetheless, it should be noted that the V_{SZ} computed from the extended profiles are likely lower bound estimates of

16

the true V_{SZ} because they assume a constant V_S below 100 m while in reality V_S usually increases with depth. The final number of profiles available for the regression as a function of depth is shown in Figure 5a. Due to the extrapolation, all 78 profiles that extend to 100 m are used to develop the V_{SZ} and V_{S30} relationships for depths between 100 and 400 m.

The fitted linear V_{S30} - V_{SZ} relationships and associated data for different depths between 50 and 400 m are shown in Figure 5b. The data for which the V_{SZ} was computed using extended V_S profiles are labeled separately from those computed without extension. As depth increases, the difference between V_{S30} and V_{SZ} increases, which is expected because V_S generally increases with depth. The standard deviation of residuals σ_{RES} (in ln units) are plotted in Figure 5a as a function of depth, with values close to 0.005 near $z = 30$ m and increasing to about 0.25 at near $z = 400$ m. The increase in σ_{RES} indicates that the variability in the predicted V_{S30} increases as depth increases.

Some additional comments are warranted regarding the use of the extended V_S profiles to develop these relationships. As noted earlier, the extended V_S profiles may underestimate the true V_{SZ}. Therefore, the true data points for deeper depths may be to the right from those shown in Figure 5b, and as a result it is possible that the V_{S30} - V_{SZ} relationships for deeper depths may be shifted or have a different slope from the true V_{S30} - V_{SZ} relationships.
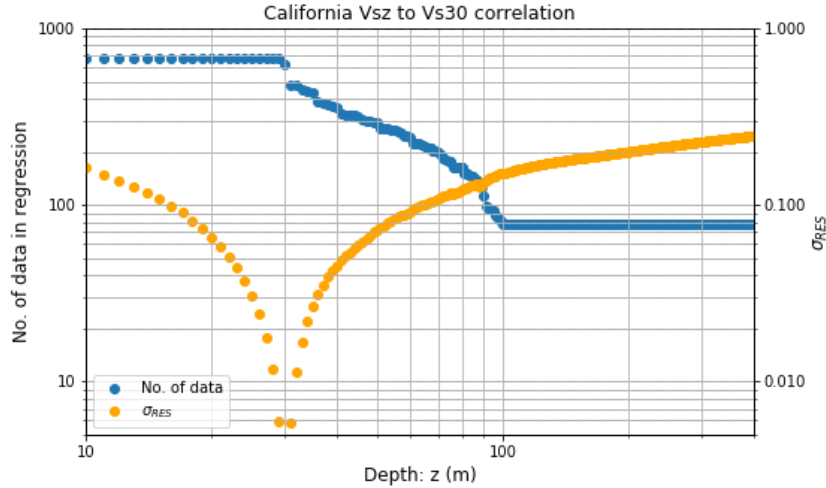
(a)
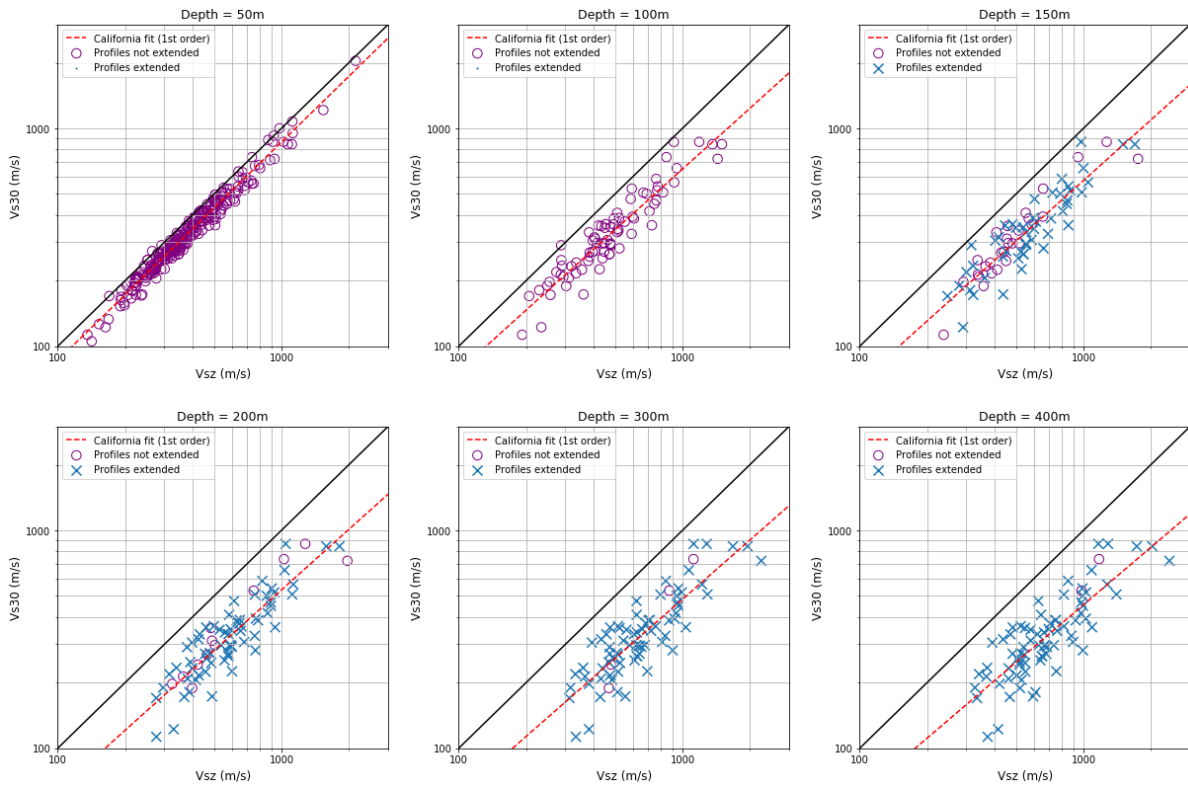
(b)

Figure 5. (a) Variation of number of points in regression and the standard deviation of residuals with depth. (b) Relationship between V_{SZ} and V_{S30} at depths z of 50, 100, 150, 20, 300, and 400 m. Californian V_S profiles from the VSPDB are used and V_S profiles that reach 100 m are extended to compute V_{SZ} for z greater than 100 m. The data computed from extended V_S profiles are labeled with crosses and the data computed from nonextended V_S profiles are labeled with crosses.

18

The V_{S30} - V_{SZ} relationships across all depths can be summarized for use in P-wave seismogram studies by considering the $z - V_{SZ}$ relationship. Because we assume that V_{SZ} represents the time averaged shear wave velocity over a depth equal to $z = 0.1 \cdot V_{SZ}$, once the V_{SZ} is specified then the depth is specified and then only a single value of V_{S30} - V_{SZ} is used from each relationship for the V_{SZ} to V_{S30} conversion. For example, if $V_{SZ} = 1000$ m/s then $z = 100$ m and the relationship in Figure 5b for $z = 100$ m is used to estimate $V_{S30} = 650$ m/s. Doing the same for a range of V_{SZ}, a single V_{SZ} to V_{S30} conversion for $z = 0.1 \cdot V_{SZ}$ can be developed. This V_{SZ} to V_{S30} conversion is shown in Figure 6a for the V_{S30} - V_{SZ} relationships developed in this study for California, as well as for V_{S30} - V_{SZ} relationships developed for Japan by Boore et al. (2011) and Miao et al. (2018), and for Central and Eastern North America (CENA) by Kim et al. (2015). Figure 6b shows the ratio between V_{S30} and V_{SZ} at various depths which is computed from Figure 6a. Please note that only depth greater than 30m are shown for Boore et al. (2011) relationship.

As shown in Figure 6a, all of the V_{SZ} to V_{S30} conversions are similar and they all indicate $V_{S30} < V_{SZ}$ for $V_{SZ} > 300$ m/s (i.e., $z = 30$ m). The California conversion developed in this study is similar to the conversion for CENA up to $V_{SZ} \sim 900$ m/s ($z < 90$ m), and similar to the Japanese relationships for $V_{SZ} \sim 1000\text{-}2000$ m/s ($z = 100\text{-}200$ m). The steeper gradient of V_S profiles in Japan relative to California (Kamai et al. 2016) likely causes the differences in the V_{SZ} to V_{S30} conversion at depths less than 100 m (i.e., $V_{SZ} < 1000$ m/s). However, Figure 6 also shows that the V_{S30} predicted using the CENA correlation is higher than the V_{S30} predicted using the California correlation for depths greater than 100 m, suggesting that by extending the V_S profiles that reach 100m constantly we may have underestimated V_{S30} for larger values of V_{SZ}.

We also observe very large predicted V_{S30} by the Japan conversions for depths greater than 200 m, which may be caused by the limited data points in this depth and V_{SZ} range.

For this study, the California V_{SZ} to V_{S30} correlations are used to predict V_{S30} from the estimated V_{SZ}. The regression coefficients can be found in the Appendix.
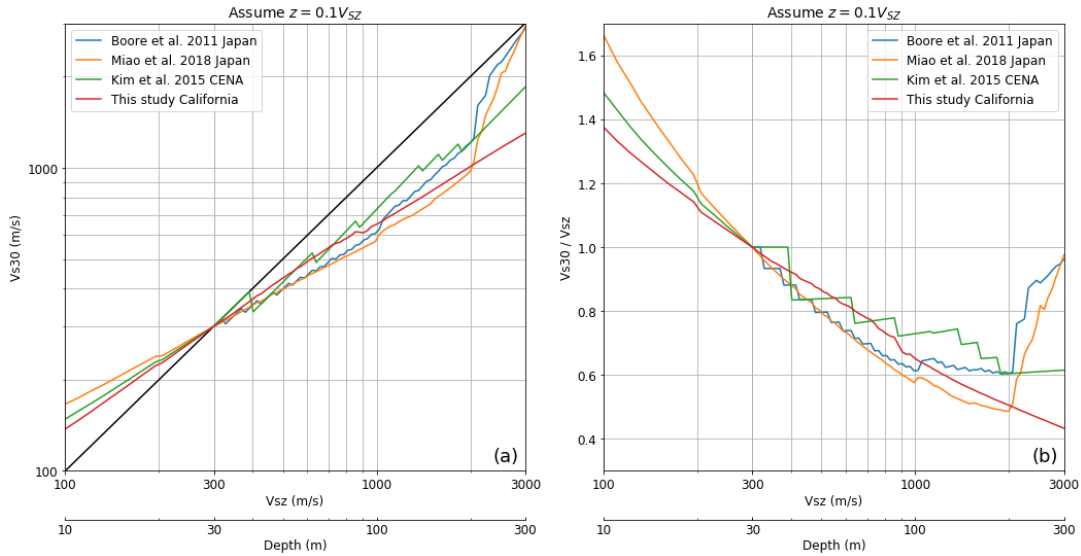

Figure 6. (a) The relationship between V_{SZ} and V_{S30} derived from the V_{SZ} to V_{S30} correlations for Japan, CENA, and California. It is assumed that $z = \tau_p \cdot V_{SZ}$ and the source time function τ_p is 0.1 second. (b) The ratio between V_{S30} and V_{SZ} computed from the relationships in (a).

COMPARISON OF V_{S30} ESTIMATES WITH V_{S30} MEASUREMENTS

The V_{S30} for 194 stations in California are estimated from 1,923 records and are compared with the V_{S30} measurements as documented in the USGS V_{S30} database (McPhillips et al. 2020) in Figure 7. Figure 7a shows the estimated V_{S30} from each record and Figure 7b shows the average estimated V_{S30} for each station. Comparing Figures 7a and 7b, it is obvious that the accuracy of the V_{S30} estimation significantly improves by averaging V_{S30} estimated from multiple

records for one site. As shown in Figure 7b, the estimated V_{S30} of 55% of stations are within ±25% of the V_{S30} measurement, and the estimated V_{S30} of 77% stations are within ±50% of the measurement. Binning the data based on the estimated V_{S30}, it is observed that in general the estimated V_{S30} is unbiased for $V_{S30} < 400$ m/s, but it tends to overestimate the measured V_{S30} for $V_{S30} > 400$ m/s and the amount of overestimation tends to increase with increasing estimated V_{S30}. For instance, the V_{S30} is in general overestimated by about 20% for sites with average estimated V_{S30} between 500 and 700 m/s; while for sites with average estimated V_{S30} greater than 1000 m/s, the amount of overestimation increases to 33%.

In an effort to remove the uncertainty in the V_{SZ} to V_{S30} correlations in our assessment of the P-wave seismogram method, we also compared the estimated V_{SZ} with the V_{SZ} computed from the 154 sites with measured V_S profiles documented by the VSPDB database. The depth extent (i.e., z) of the measured V_{SZ} is taken as 0.1s times the mean estimated V_{SZ}, and the measured V_S profile is extrapolated with a constant velocity if it does not extend to the required depth z. The comparison between estimated and measured V_{SZ} for each record and each station are shown in Figure 7c and 7d, respectively. The data points are binned based on the average estimated V_{SZ} in Figure 7d, and the binned data suggests that the V_{SZ} tends to be overestimated for sites with average estimated V_{SZ} greater than about 400 m/s. Similar to what is observed for V_{S30}, the overestimation in V_{SZ} increases with increasing average estimated V_{SZ}: the V_{SZ} of sites with average estimated V_{SZ} between 500 and 700 m/s are overestimated by about 20% whereas sites with average estimated V_{SZ} over 1000 m/s are overestimated by about 45%. It is considered that constant extrapolation of the V_S profiles may cause the computed V_{SZ} from the V_S measurements to be smaller than the actual values because shear wave velocity typically

increases with depth. However, the similar overestimation observed for V_{S30} in Figure 7b indicates that the overestimation of V_{SZ} for larger V_{SZ} may be genuine.
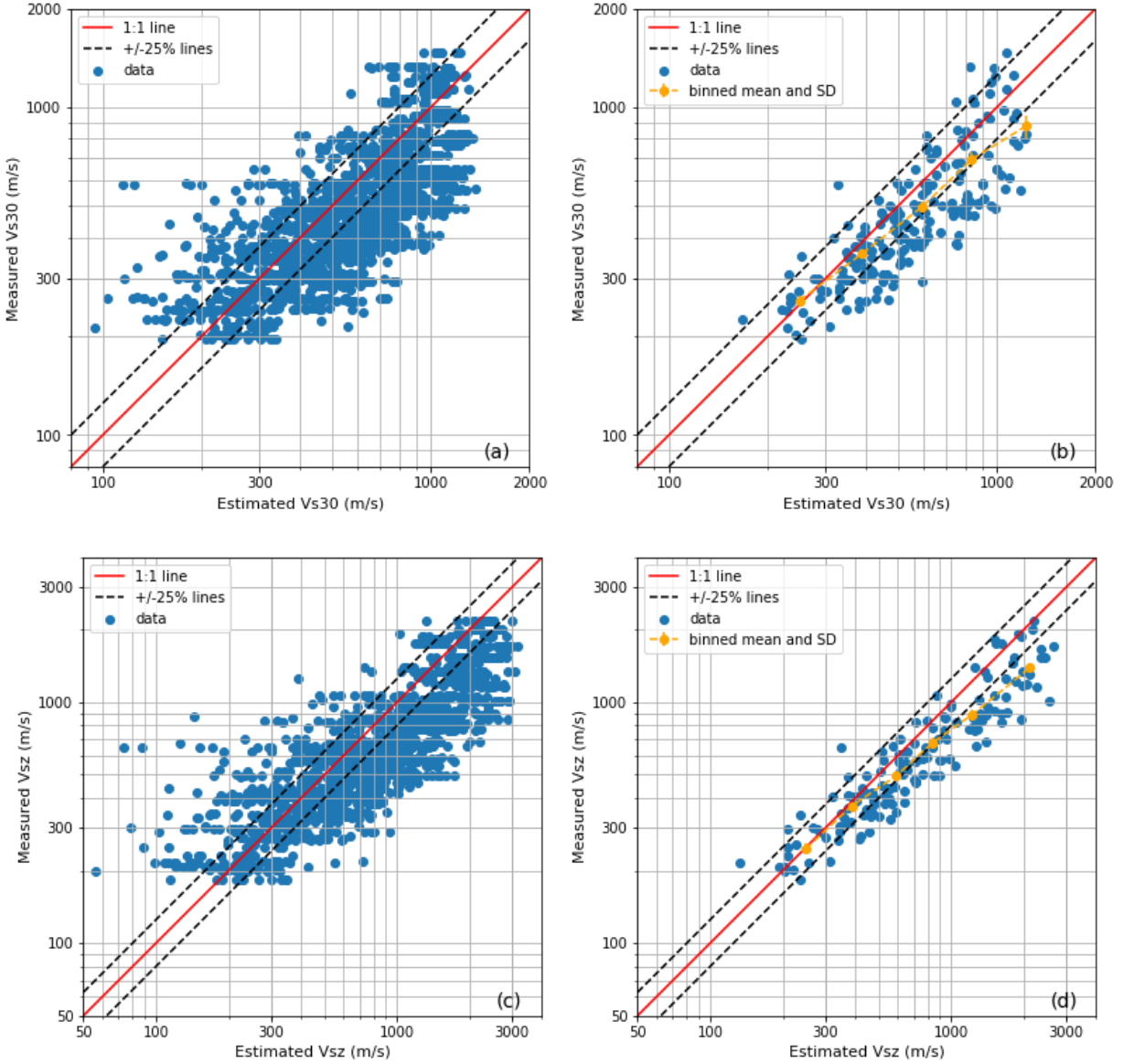

Figure 7. (a) Measured V_{S30} and the V_{S30} estimated from 1923 records; (b) Measured V_{S30} and the average estimated V_{S30} for the 194 stations; (c) Measured V_{SZ} and the V_{SZ} estimated from 1568 records; (d) Measured V_{SZ} and the average estimated V_{SZ} for the 154 stations with V_S profile available on VSPDB.

The residuals ($y_{i,j}$) between the measured V_{S30} ($V_{S30,mea}$) and estimated V_{S30} ($V_{S30,est}$) for each record j and each site i are computed using equation (8). We applied the same mixed-effect model (Pinheiro and Bates 2000) as Kim et al. (2015) to partition the residual $y_{i,j}$ into three components: the overall mean residual a, the average site residual η_{Si}, and the remaining within-site residual $\varepsilon_{i,j}$. The equations used to compute the residuals are given below, with $\bar{y}_i$ representing the average residual for site i.

$$y_{i,j} = \ln\left(V^i_{S30,mea}/V^{i,j}_{S30,est}\right) \tag{8}$$

$$y_{i,j} = a + \eta_{Si} + \varepsilon_{i,j} \tag{9}$$

$$\eta_{Si} = \bar{y}_i - a \tag{10}$$

$$\varepsilon_{i,j} = y_{i,j} - \eta_{Si} - a \tag{11}$$

The mean (a) of the total residuals ($y_{i,j}$) is computed as -0.179. The negative value of a is consistent with the observation of overestimation of V_{S30} by the P-wave seismogram method and represents an average overestimation of about 20%. The standard deviation of the residuals can be partitioned into the between-site variability (τ) and within-site variability (ϕ) by computing the standard deviations of η_{Si} and $\varepsilon_{i,j}$, respectively. The resulting values are $\tau = 0.260$ and $\phi = 0.232$, and these values can be combined to compute the overall standard deviation (σ) using:

$$\sigma = \sqrt{\tau^2 + \phi^2} \tag{12}$$

The value of σ computed from τ and ϕ is 0.348.

To examine the factors that may cause bias in the estimated V_{S30}, we plot the total $(y_{i,j})$ and within-site $(\varepsilon_{i,j})$ residuals against epicentral distance, focal depth, earthquake magnitude, and the take-off angle (i.e., angle i in layer n, Figure 2), as shown in Figure 8. The data only show a significant trend between the residuals (both total and within-site) and take-off angle, and it is observed that for take-off angles smaller than about 30° the V_{S30} can be severely overestimated. This trend was also observed by Kang et al. (2020). A small take-off angle is associated with a small epicentral distance coupled with a deeper focal depth, and thus the waves take-off more steeply to travel to the site. The reason for the bias is likely that for a ray path with a small take-off angle, the radial component is too small compared to the vertical component to allow for Snell's Law to show a significant effect on $\frac{\dot{U}_R}{\dot{U}_Z}$ and thus the radial to vertical ratio does not reduce sufficiently to reflect the stiffness of the soil layers that the wave travelled through. It is therefore suggested that records with take-off angle less than 30° should be avoided when applying the P-wave seismogram method. For this study, there are only 15 records falling in this category. Alternatively, a bias model can be utilized to correct the trend in residual with take-off angle, as in Kang et al. (2020).

In general, the average negative bias indicates about 20% overestimation of V_{S30}, on average, and this bias cannot be explicitly explained by factors such as epicentral distance, focal depth, or earthquake magnitude. We suspect that the V_{SZ} estimated by the P-wave seismogram method may be influenced by the presence of high shear-wave velocity layers that extend relatively close to the surface but are not reflected in the value of V_{S30}. In this case, the small thickness of soft material near the surface may not generate enough refraction for on $\frac{\dot{U}_R}{\dot{U}_Z}$ to reflect the stiffness in the top 30 m. HVSR (horizontal to vertical spectral ratio) data may provide

information that can be used to detect the existence of a shallow velocity contrast and identify sites that may potentially be overestimated by the P-wave seismogram method. This is an avenue for future research.
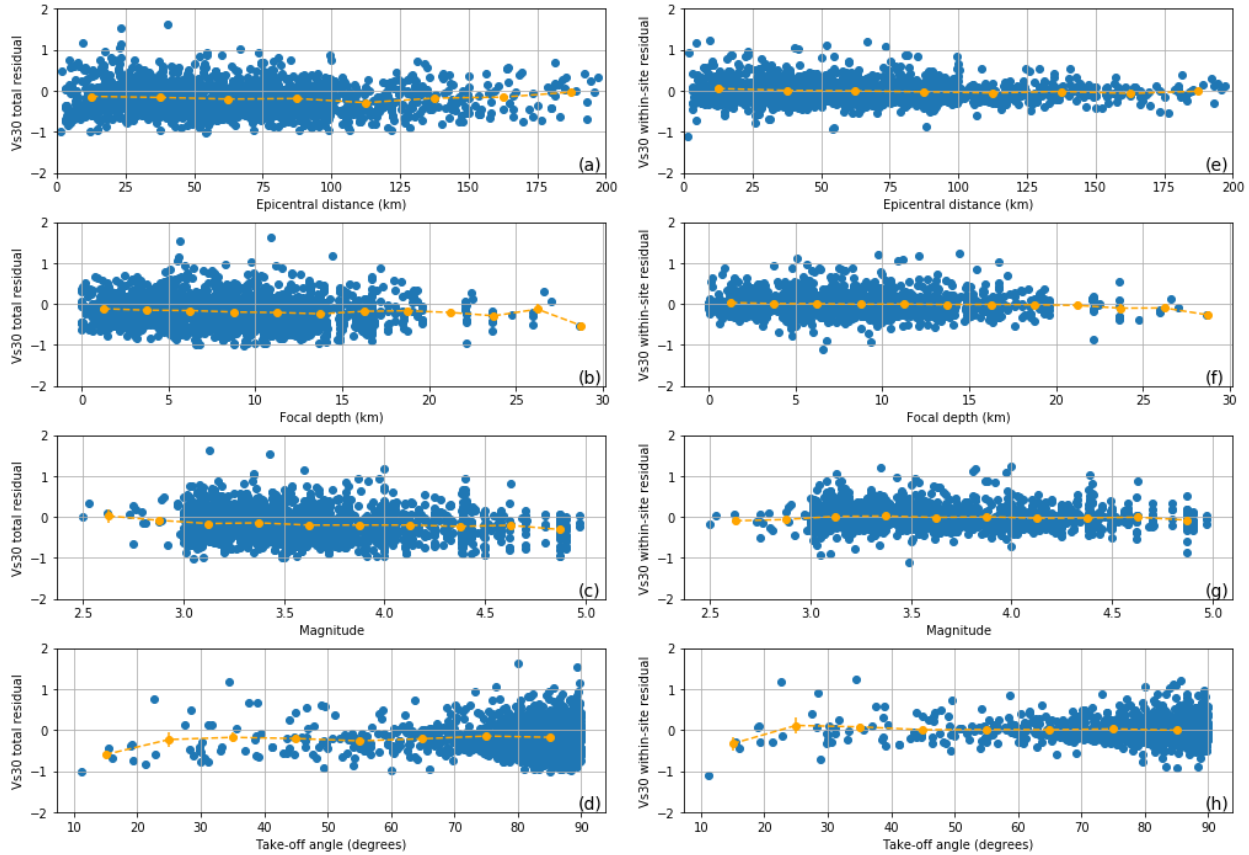

Figure 8. Total and within-site residuals against epicentral distance (a, e), focal depth (b, f), magnitude (c, g), and take-off angle (d, h) of the event

EFFECT OF USING A MORE DETAILED CRUSTAL MODEL

Because crustal velocity models with more detailed layers at shallow depth are available for California, we investigate the effect of utilizing a detailed crustal velocity model on the V_{S30} results. We consider the crustal velocity models that are available on the Broadband Platform (BBP, https://github.com/SCECcode/bbp/wiki/File-Format-Guide) of the Southern California

Earthquake Center (SCEC), which divides California into four regions: Southern California (SoCal), Mojave, Northern California (NoCal), and Central Coast, as shown in Figure 9a. Four different crustal models are adopted for the four regions, and the P-wave velocity profiles for these crustal models are plotted in Figure 9b together with the four-layer simplified crustal model used in this study. The main difference between the detailed and simplified crustal models is the velocity structure in the top 5 km, with the detailed velocity models showing much smaller velocities over this depth range.
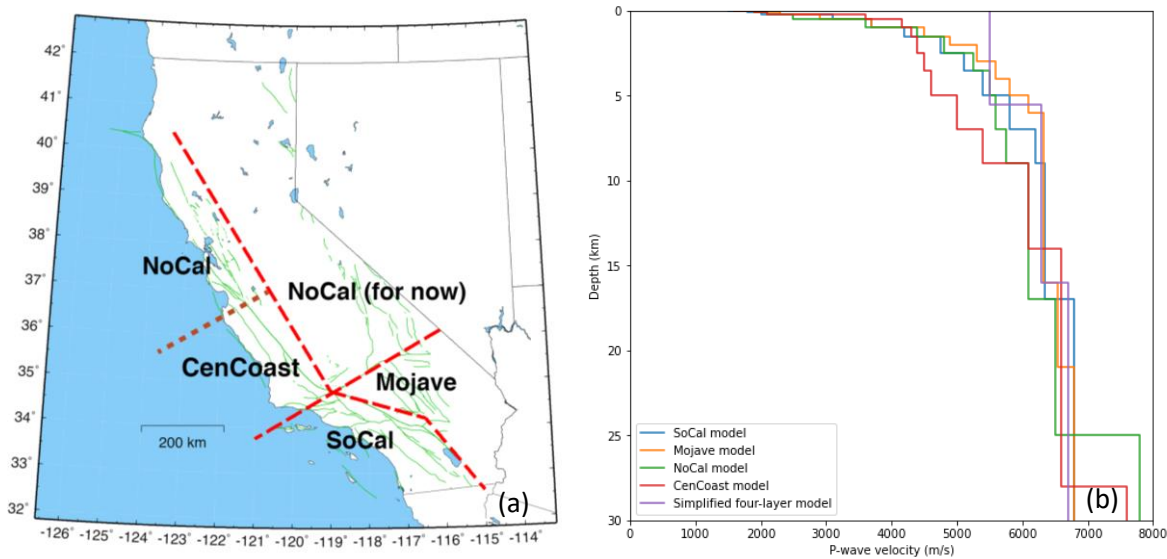

Figure 9. (a) Four regions that are divided from California where different crustal models are used; (b) P-wave velocity models for the four different regions and the simplified P-wave velocity crustal model used by this study

Applying a different crustal velocity model affects the estimation of the ray parameter p. We therefore assign crustal models based on the location of the ground motion recording station being analyzed and compute the ray parameter with the appropriate BBP crustal model. The V_{S30} residuals from the analysis using the BBP and simplified crustal models are plotted versus focal depth in Figure 10a. The residuals are similar for focal depths greater than about 2.5 km, but

become more positive for the BBP crustal models at focal depths less than 2.5 km, which indicates that the BBP crustal models are producing smaller values of V_{S30}. The ray parameter p estimated using the BBP crustal models is compared with the ray parameter estimated using the simplified crustal model used in this study in Figure 10b. The data clearly show that p estimated from the BBP crustal models is significantly greater than p estimated from the simplified crustal model for focal depths less than about 2.5 km. A larger p for the same $\frac{\dot{U}_R}{\dot{U}_Z}$ results in a smaller V_{SZ} (equation 4). The unreasonably large values of p at shallow focal depths is caused mainly by the smaller P-wave velocities and significant change in velocity over the top 3 km in the more detailed BBP crustal models. These characteristics the BBP crustal models result in larger computed take-off angles and larger ray parameters (see equation 2) than for the simplified crustal model due to significant refraction that occurs as the velocities decease towards the surface.

Due to the significant positive bias observed in total residuals for V_{S30} when the more detailed BBP crustal models are used, a more simplified crustal model is recommended for use in the P-wave seismogram method. The more detailed BPP crustal velocity models do not improve the V_{S30} estimates and may bias V_{S30} estimates significantly for focal depths less than 2.5 km.
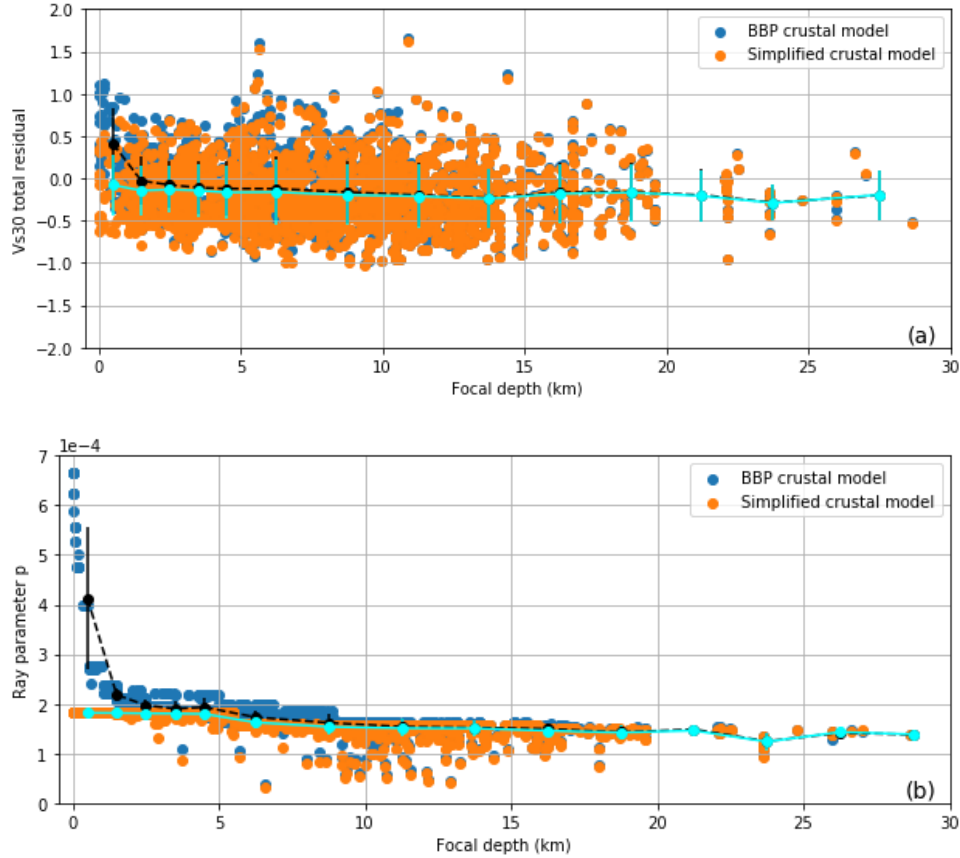
Figure 10. (a) Total residuals of V_{S30} computed assuming the more detailed BBP crustal model and assuming the simplified southern California crustal model against focal depth. (b) Ray parameters computed assuming the more detailed BBP crustal model and assuming the simplified southern California crustal model against focal depth.

COMPARISON WITH OTHER V_{S30} ESTIMATION METHODS

The mean (a) and standard deviation (σ) of the total residuals obtained in this study for California are -0.179 and 0.350, respectively. The mean residuals for the P-wave seismogram method reported by others include -0.073 for Japan (Miao et al. 2018), -0.26 for automatic procedure when SNR threshold of 2 is applied for Japan (Kang et al. 2020) and -0.10 for CENA (Kim et al. 2015). This study indicated slightly more overestimation for California (i.e., ~20%) than Miao et al. (2018) and Kim et al. (2015) that indicate about 10% overall overestimation of V_{S30}. The overestimation of Kang et al. (2020) is about 30%, and they proposed a bias correction

approach to reduce the bias. The standard deviation of 0.350 obtained in this study for California is similar to that reported for Japan ($\sigma = 0.36$, Miao et al. 2018) and slightly smaller than for CENA ($\sigma = 0.43$, Kim et al. 2016) and for automatic procedure in Japan ($\sigma = 0.69$ after bias correction, Kang et al. 2020).

Finally, the V_{S30} estimated using the P-wave seismogram method in this study is compared with the V_{S30} estimated by Wills et al. (2015) based on geology and slope (Figure 11). The Wills et al. (2015) V_{S30} estimates represent just a few different values because values are assigned based predominantly on geology and each geologic unit is assigned the same V_{S30}. The P-wave seismogram estimates are more evenly distributed across the range of measured V_{S30}. As shown in Figure 11, the largest values of V_{S30} predicted by Wills et al. (2015) are 733 m/s for unit KJf (Marine sedimentary and metasedimentary rocks Cretaceous-Jurassic) and 710 m/s for crystalline, despite measurements as larger as 1312 m/s. Also, some stations with measured V_{S30} greater than 800 m/s are assigned V_{S30} of 350-390 m/s by the Wills et al. (2015) proxy based on the mapped geology of Qal3 (young alluvium with slope greater than 2%) or Qoa (Older Pleistocene alluvium). The discrepancy here may be associated with errors in the mapped geologic units for the stations, but the P-wave seismogram method is able to estimate the V_{S30} of those stations with better accuracy. The V_{S30} residuals (in ln units) associated with each station are computed for the Wills et al. (2015) geology/slope proxy estimates of V_{S30}. The mean residual is 0.002, indicating overall an unbiased estimate, and the standard deviation is 0.369, which represents the between-site variability (τ). As noted earlier, the P-wave seismogram method provides slightly biased V_{S30} estimates (average residual -0.174) but the resulting site-wise variability is considerably smaller ($\tau = 0.258$).
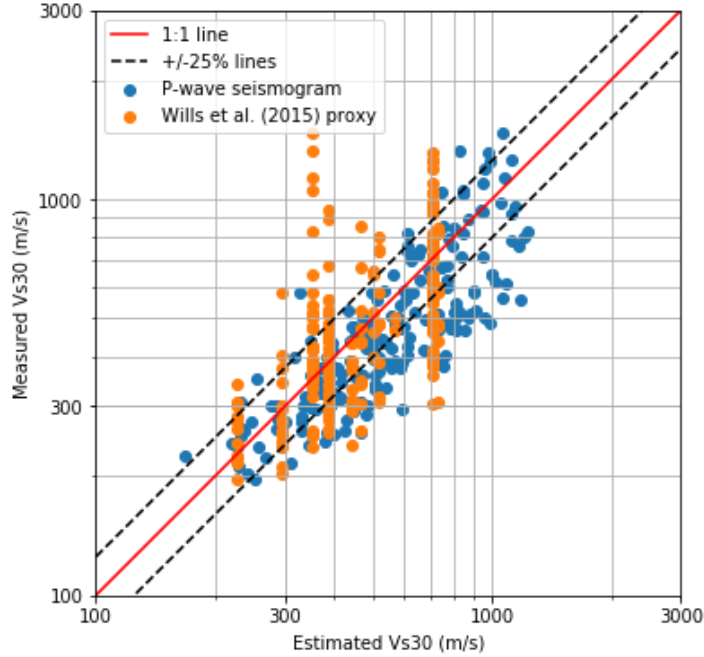
Figure 11. Measured V_{S30} and the average estimated V_{S30} by the P-wave seismogram method and the estimated V_{S30} by the Wills et al. (2015) proxy-based method.

CONCLUSIONS

In this study, the P-wave seismogram method was applied to estimate the V_{S30} of 194 stations in California utilizing 1,923 ground motion recordings. The P-wave seismogram method first estimates the average shear wave velocity over depth z (V_{SZ}) from the ratio of the radial to vertical components of the P-wave arrival on recorded seismograms $\left(\frac{\dot{U}_R}{\dot{U}_Z}\right)$ and the ray parameter p associated with the P-wave travel path from the hypocenter to recording station. The associated z is taken as $z = 0.1 \text{ s} \cdot V_{SZ}$, and then the V_{SZ} is converted to V_{S30} using a regional V_{SZ}-V_{S30} relationship that is a function of z.

A California-specific correlation between V_{SZ} and V_{S30} was developed for z between 5 and 400 m utilizing 680 measured V_S profiles documented in the VSPDB database. To enhance

30

the number of datapoints available for deeper depths, V_S profiles that reached 100m depth were extended to 400 m using the V_S at the base of the profile. The V_{SZ} to V_{S30} conversion for California was compared with published relationships for Japan and CENA, and the results show that the relationship for California is more similar to the relationship for CENA for depths less than 90 m and more similar to the relationship for Japan for depth 100-200m. Nonetheless, the three relationships are relatively similar.

Comparison between the V_{S30} measurements and P-wave seismogram estimates shows that for 55% of stations the estimated and measured V_{S30} are within ±25% of each other, and for 77% stations the estimated and measured V_{S30} are within ±50% of each other. The results validate the reliability of the P-wave seismogram method in estimating the near surface shear wave velocity relatively quickly and with low cost at seismic recording stations that have earthquake records available. The mean residual (a) and standard deviation of the residuals (σ) for the natural log of V_{S30} for this data set were estimated to be -0.179 and 0.348, respectively. The mean residual indicates a potential overestimation of V_{S30} (about 20%, on average), especially for stations with estimated V_{S30} between 500 to 1300 m/s. This mean residual is slightly smaller (i.e. more negative) than those reported for studies in Japan and CENA, with these other studies indicting about 10% overestimation from the P-wave seismogram method. The standard deviation obtained for California in this study is similar to the standard deviations reported for Japan and CENA.

The effect of using a more detailed crustal velocity model in estimating the ray parameter p was analyzed, and the results show that the smaller low P-wave velocities in shallow layers in the detailed crustal models lead to underestimation of V_{S30} for earthquake events with focal depths shallower than 2.5km, and this is caused by an overestimation in the ray parameter. For

events with focal depth greater than about 5 km, there is no influence of the more detailed crustal velocity models of the similarity between the detailed crustal model and the simplified crustal model at greater depths.

The V_{S30} at the 194 study sites were also estimated using the geologic/slope proxy of Wills et al. (2015), and the V_{S30} residuals for the proxy predictions relative to the measurements were compared with those for the P-wave seismogram method. The geologic/slope proxy estimates are generally unbiased (mean residual ~ 0) for the sites evaluated, but the standard deviation of the residuals was larger for the geologic/slope proxy than for the P-wave seismogram method.

PROJECT DATA

The data analyzed as part of this project is being prepared for a data publication in the DesignSafe Data Depot (www.designsafe-ci.org). The data will be published electronically with a Digital Object Identifier, facilitating data re-use and citation by others.

REFERENCES

Abrahamson, N. A., Silva, W. J., and Kamai, R., 2014. Summary of the Abrahamson, Silva, and Kamai NGA-West 2 ground-motion relations for active crustal regions, Earthquake Spectra 30, 1025–1055.

Ahdi, S.K., J.P. Stewart, T.D. Ancheta, D.Y. Kwak, and D. Mitra (2017). Development of Vs profile database and proxy-based models for Vs30 prediction in the Pacific Northwest region of North America, Bull. Seis. Soc. Am., 107(4), 1781-1801.

Aki, K. and P. G. Richards (2002). Quantitative Seismology. J. Ellis. Sausalito, CA. University Science Books. 700 pages.

Allen, T. I., and Wald, D. J., 2009. On the use of high-resolution topographic data as a proxy for seismic site conditions (Vs_{30}), *Bulletin of the Seismological Society of America* 99, 935–943.

American Society of Civil Engineers (ASCE), 2016. Minimum Design Loads for Buildings and Other Structures, ASCE Standard ASCE/SEI 7-16, Reston, VA.

David M. Boore, Eric M. Thompson, Héloïse Cadet; Regional Correlations of *VS*30 and Velocities Averaged Over Depths Less Than and Greater Than 30 Meters. *Bulletin of the Seismological Society of America* 2011; 101 (6): 3046–3059. doi: https://doi.org/10.1785/0120110071.

Boore, D. M., Stewart, J. P., Seyhan, E., and Atkinson, G. M., 2014. NGA-West2 Equations for Predicting PGA, PGV, and 5% Damped PSA for Shallow Crustal Earthquakes, Earthquake Spectra 30, 1057–1085.

Chiou, B. S.-J., and Youngs, R. R., 2014. Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra, Earthquake Spectra 30(3), 1117–1153.

Cox, B. R., and Beekman, A. N., 2011. Intramethod Variability in ReMi Dispersion Measurements and V_S Estimates at Shallow Bedrock Sites, Journal of Geotechnical and Geoenvironmental Engineering 137(4).

European Committee for Standardization (CEN), 2013. Eurocode 8: Design of structures for earthquake resistance, part 1: General rules, seismic actions and rules for buildings, EN 1998-1, Brussels, Belgium.

Jones, J., Kalkan, E. and Stephens, C. (2017). Processing and Review Interface for Strong Motion Data (PRISM)—Methodology and Automated Processing, Version 1.0.0: U.S. Geological Survey Open-File Report, 2017–1008, 81 p.

Idriss, I. M., 2014. An NGA-West2 empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes, Earthquake Spectra 30(3), 1155–1177.

Iwahashi, J., and R. J. Pike (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three part geometric signature, Geomorphology 86, nos. 3/4, 409–440.

Kang S, Kim B, Park H, Lee J (2020) Automated procedure for estimating V_{S30} utilizing P-wave seismograms and its application to Japan. *Engineering Geology* 264:1–16, DOI: https://doi.org/10.1016/j.enggeo.2019.105388

Kim, B., Hashash, Y. M. A., Rathje, E. M., Stewart, J. P., Ni, S., Somerville, P. G., Kottke, A. R., Silva, W. J., and Campbell, K. W., 2016. Subsurface shear-wave velocity characterization using P-wave seismograms in Central, and Eastern North America, Earthquake Spectra 32(1), 143–169.

Kottke, A. R., Hashash, Y. M. A., Stewart, J. P., Moss, C. J., Nikolaou, S., Rathje, E. M., Silva, W.J.,and Campbell, K. W., 2012. Development of geologic site classes for seismic site amplification for central, and eastern North America, Paper No. 4557, in Proc. 15th World Conf. on Earthquake Eng., 24–28 September 2012, Lisbon, Portugal.

Kwak D., S. J. Brandenberg, A. Mikami, J. P. Stewart, 2015. Prediction Equations for Estimating Shear-Wave Velocity from Combined Geotechnical and Geomorphic Indexes Based on Japanese Data Set. BSSA; 105 (4): 1919–1930.

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

Miao, Y., Shi, Y., & Wang, S.-Y. (2018). Estimating Near-Surface Shear Wave Velocity Using the P-Wave Seismograms Method in Japan. Earthquake Spectra, 34(4), 1955–1971. https://doi.org/10.1193/011818EQS015M

Ni, S., Li, Z., and Somerville, P., 2014. Estimating subsurface shear velocity with radial to vertical ratio of local P waves, Seismological Research Letters 85(1), 82–89.

Owens, T. J., Crotwell, H. P., Groves, C., and Oliver-Paul., P., SOD: Standing Order for Data. Seismological Research Letters, 75:515–520, 2004.

Parker, G. A., Harmon, J. A., Stewart, J. P., Hashash, Y. M. A., Kottke, A. R., Rathje, E. M., Silva, W., and Campbell, K. W., 2017. Proxy-based Vs_{30} estimation in central, and eastern North America, *Bull. Seisol. Soc. Am.*, 107 (1): 117–131. doi: https://doi.org/10.1785/012016010.

Thompson, Eric M., Kayen, Robert E., Carkin, Brad, and Tanaka, Hajime, 2010, Surface-wave site characterization at 52 strong-motion recording stations affected by the Parkfield, California, M6.0 earthquake of 28 September 2004, U.S. Geological Survey Open-File Report 2010-1168, 117 p. [http://pubs.usgs.gov/of/2010/1168/].

Wald, D. J., and Allen, T. I., 2007. Topographic Slope as a Proxy for Seismic Site Conditions and Amplification, *Bull. Seisol. Soc. Am.*, 97, 1379–1395.

Wald, L. A., Hutton, L. K., and Given, D. D., 1995. The southern California network bulletin: 1990-1993 summary, Seismological Research Letters 66(1), 9-19.

Wills, C. J., and K. B. Clahan (2006). Developing a map of geologically defined site-condition categories for California, Bull. Seismol. Soc.Am. 96, 1483–1501.

Wills, C. J., Gutierrez, C. I., Perez, F. G., Branum, D. M., 2015. A Next Generation *VS*30 Map for California Based on Geology and Topography, *Bull. Seisol. Soc. Am.*, 105 (6): 3083–3091. doi: https://doi.org/10.1785/0120150105.

Yong, A., 2016. Comparison of measured and proxy-based Vs_{30} values in California, *Earthquake Spectra,* 32(1), 171–192.

Zalachoris, G., Rathje, E. M., and Paine, J. G., 2017. VS30 characterization of Texas, Oklahoma, and Kansas using the P-wave seismogram method, Earthquake Spectra 33(3), 943–961.

Appendix: Regression coefficients for V_{SZ} - V_{S30} relationships using equation (7)

Depth (m)	C1	C0	No. of data	σ_{RES}
5	0.8705	1.1195	680	0.253
10	0.9645	0.4817	680	0.162
11	0.9723	0.4166	680	0.149
12	0.9777	0.3668	680	0.137
13	0.9811	0.3294	680	0.127
14	0.9847	0.2921	680	0.117
15	0.9883	0.2550	680	0.108
16	0.9914	0.2218	680	0.098
17	0.9929	0.1991	680	0.090
18	0.9948	0.1739	680	0.081
19	0.9968	0.1483	680	0.073
20	0.9983	0.1268	680	0.066
22	1.0006	0.0874	680	0.051
24	1.0011	0.0597	680	0.037
26	1.0014	0.0348	680	0.024
28	1.0013	0.0128	680	0.012
30	1.0000	0.0000	631	0.000
32	1.0005	-0.0225	475	0.011
34	1.0011	-0.0437	445	0.022
36	1.0054	-0.0841	385	0.031
38	1.0042	-0.0930	370	0.039
40	1.0017	-0.0927	358	0.045
45	1.0050	-0.1459	319	0.059
50	1.0022	-0.1606	293	0.072
55	0.9984	-0.1635	267	0.084
60	0.9954	-0.1718	242	0.091
65	0.9934	-0.1851	213	0.101
70	0.9959	-0.2239	200	0.109
75	0.9877	-0.1972	176	0.116
80	0.9682	-0.1000	162	0.121
85	0.9662	-0.1025	145	0.130
90	0.9382	0.0376	112	0.134
95	0.9196	0.1347	92	0.144
100	0.9260	0.0807	78	0.150
110	0.9263	0.0507	78	0.160
120	0.9260	0.0264	78	0.167
130	0.9239	0.0160	78	0.173
140	0.9229	0.0014	78	0.178
150	0.9224	-0.0153	78	0.182

160	0.9226	-0.0346	78	0.186
170	0.9228	-0.0527	78	0.190
180	0.9228	-0.0684	78	0.193
190	0.9227	-0.0820	78	0.197
200	0.9220	-0.0903	78	0.200
225	0.9178	-0.0919	78	0.208
250	0.9130	-0.0848	78	0.215
275	0.9073	-0.0678	78	0.222
300	0.9012	-0.0450	78	0.227
325	0.8947	-0.0179	78	0.233
350	0.8883	0.0114	78	0.238
375	0.8818	0.0420	78	0.242
400	0.8755	0.0732	78	0.246