PRELIMINARY ERGODIC SITE RESPONSE MODEL FOR CALIFORNIA CONDITIONED ON mHVSR

Francisco Javier G. Ornelas[1], Christopher A. de la Torre[2], Tristan E. Buckreis[1], Chukwuebuka C. Nweke[3], Scott J. Brandenberg[1], and Jonathan P. Stewart[1]

[1] *Department of Civil & Environmental Engineering*
University of California, Los Angeles, USA
[2] *Department of Civil & Natural Resources Engineering*
University of Canterbury, Christchurch, NZ
[3] *Department of Civil & Environmental Engineering*
University of Southern California, Los Angeles, USA

Abstract

Traditional ergodic ground motion models (GMMs), typically conditioned on the time-averaged shear-wave velocity in the upper 30 m of soil (V_{S30}), exhibit high site-to-site variability due to their inability to capture site-specific features such as resonances caused by impedance contrasts. Microtremor-based horizontal-to-vertical spectral ratios (mHVSR), which can contain peaks linked to site resonances, offer a useful supplement to V_{S30} for site response prediction. We use peak features from mHVSR spectra from 685 sites across California and neighboring regions (e.g., Nevada, Oregon) to develop mHVSR-conditioned site response models, which reduce site-to-site variability compared to a baseline linear site amplification model.

Keywords: horizontal-to-vertical spectral ratios, ground motion models, site response

Introduction

Ground motion models (GMMs) developed for active tectonic regions in the NGA-West2 project (Bozorgnia et al. 2014) and the ongoing NGA-West3 project include site amplification models derived using global datasets, but with regional customization as warranted by the data. Such models are referred to as *ergodic* (Anderson and Brune, 1999), as they assume that ground motions for different sites and events across the globe or a region can be used to characterize ground motions for a single site, albeit with large uncertainty. The site amplification models are typically conditioned on simplified site parameters, such as the time-averaged shear-wave velocity in the upper 30 m (V_{S30}) and depths to shear-wave velocity isosurfaces (e.g., $z_{1.0}$ or $z_{2.5}$). These parameters only partially characterize the underlying site conditions and are generally unable to capture site-specific effects such as resonances caused by large impedance contrasts.

Horizontal-to-vertical spectral ratios (HVSR) can be used to identify site resonances from peaks in HVSR-frequency curves. HVSR can be employed in site characterization at low cost (Molnar et al., 2018). Microtremor HVSR (denoted mHVSR) is computed from ambient vibration recordings from a three-component seismometer typically about 1–2 hours in duration. Following Fourier amplitude spectrum (FAS) computation, mHVSR curves are obtained as the ratio of the combination of horizontal FAS components to the vertical component. First introduced by Nogoshi

and Igarashi (1970; 1971) and later formalized by Nakamura (1989), HVSR can reveal site resonance effects when horizontal ground motion is amplified relative to the vertical. The fundamental resonance frequency (f_0) has been found to be associated with the lowest-frequency peak in mHVSR curves (e.g., Lermo and Chávez-García, 1993; Lachet et al., 1996; Ghofrani et al., 2013). Spectral shape and peak amplitudes of mHVSR have also been shown to correlate with site-specific ground motion characteristics, as reflected in residuals from ground motion models (e.g., Senna et al., 2008; Pinilla-Ramos et al., 2022; Wang et al., 2022a; Buckreis et al., 2024).

This study investigates the use of mHVSR features as conditioning parameters for an ergodic site response model. We utilize mHVSR data from 685 sites across California and compare them to site terms derived from residuals of the Boore et al. (2014) GMM (hereafter BSSA14), which uses the Seyhan and Stewart (2014) ergodic site amplification model (hereafter SS14). We anticipate developing two models: one based on peak features extracted from mHVSR to capture site-specific resonance behavior, and another that incorporates the full spectral shape to model linear site amplification. An initial version of the first of these is presented here. We evaluate performance from changes to site-to-site variability relative to SS14.

mHVSR and Site Term Data

Database and mHVSR Data Processing

We extract mHVSR data from the shear-wave velocity profile database (VSPDB) (Kwak et al., 2021; https://vspdb.org). The VSPDB contains mHVSR data from 1400 sites in California and surrounding regions (Nevada, Oregon, Mexico). The data sources are: (1) vibrations recorded by permanently installed seismometers, (2) vibrations recorded during temporary deployments (typically ~1-2 hr). Sources of the vibrations recorded in these surveys include a mix of shear and surface waves induced by ambient natural (e.g., wind, ocean waves) and anthropogenic (e.g., traffic, machinery) sources (e.g., Lermo and Chavez-Garcia, 1994). The data were generally processed using procedures outlined in Wang et al. (2022b), which are modified from SESAME (2004). Table 1 summarizes the sources of the mHVSR data retrieved from the VSPDB.

Table 1. mHVSR data sources from the VSPDB.

Citation ID	Citation	Count
1	Yong et al. (2013)	45
115	GEOVision (2018)	6
116	GEOVision (2016)	10
130	ENGEO	10
196	Buckreis et al. (2021)	19
197	IRIS	48
578	Ornelas et al. (2023)	7
584	Hallal and Cox, pers. comm. (2023)	1
725	Nweke et al. (2024a)	1
727	Nweke et al. (2024b)	1
728	Ornelas et al. (202X)	1
735	Ornelas et al. (202X)	10
742	Ornelas et al. (2025a)	508
743	Ornelas et al. (2025b)	4
744	Ornelas et al. (2025c)	2
745	Ornelas et al. (2025d)	12

Earthquake ground motions were queried from the ground motion database (GMDB) (Buckreis et al., 2025). We utilize 83,644 ground motion recordings from 4,469 stations and 873 events. Data were screened as follows:

1. We remove sites with less than four recordings of **M** > 4.0 earthquakes, and we remove events recorded by less than five stations. This was done so that site and event terms derived from earthquake data are reasonably well-defined.
2. We apply maximum-distance (R_{max}) cutoffs appropriate for sensor type (Mohammed et al., 2024 for CSN stations, Boore et al., 2014 otherwise).
3. We only utilize station types of "ground" or "free-field" based on COSMOS codes.
4. We only use spectral ordinates for periods shorter than the longest usable period defined as $T_{max} = 1/(1.25 \times f_{c,hp})$, where $f_{c,hp}$ is the high-pass corner frequency (the screen uses the maximum value for the two-horizontal components).

After screening, the dataset consists of 63,768 recordings from 2,573 stations and 589 events. Among the 2453 stations, we then identified a subset of 685 with co-located mHVSR surveys within a 150 m radius at the locations shown in Figure 1 that are considered in this study.

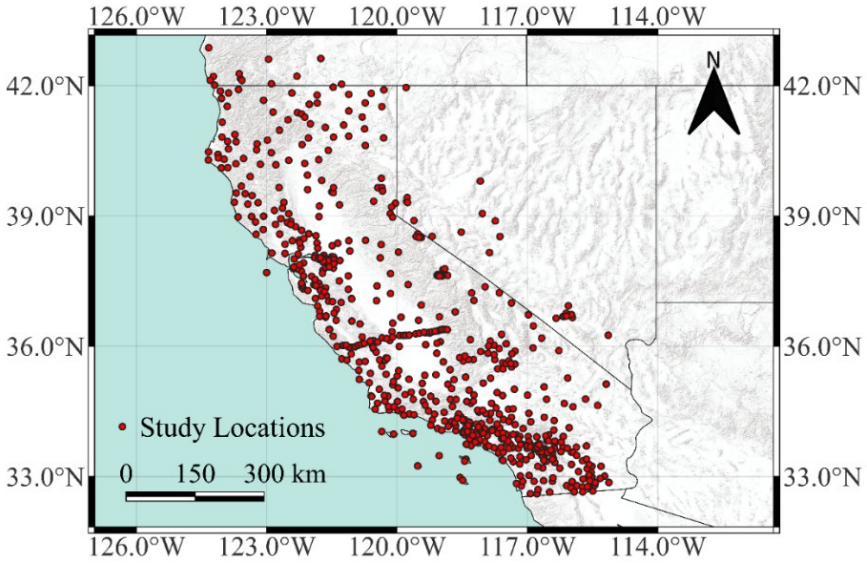

Figure 1. Locations of ground motion stations with mHVSR curves considered in the present analyses.

Site Response Estimates Derived from Ground Motion Data

For each of the 685 seismic recording station sites we compute site-specific linear site amplification to compare to mHVSR ordinates. That process begins by computing total residuals δ_{ij} for site j in event i, as the difference between the natural log of an observed intensity measure (Y_{ij}) and the median predictions of a GMM (μ_{ij}) (Boore et al., 2014),

$$\delta_{ij} = \ln(Y_{ij}) - \mu_{ij}(\mathbf{M_i}, R_{jb,ij}, V_{S30,j}, z_{1.0,j}) \quad (1)$$

where $\mathbf{M_i}$ is the event magnitude, $R_{jb,ij}$ is the distance from the site to the ground surface projection of the fault plane, $V_{S30,j}$ and $z_{1.0,j}$ are site parameters, and Y_{ij} is an observed ground motion intensity measure (peak acceleration (PGA), peak velocity (PGV) or pseudo-acceleration response spectral ordinate). Each intensity measure is defined as the median horizontal-component (RotD50; Boore, 2010). Total residuals were partitioned into fixed and random effects using linear mixed effects analyses (Abrahamson and Youngs, 1992; Sahakian et al., 2018):

$$\delta_{ij} = c_k + \eta_{E,i} + \eta_{S,j} + \varepsilon_{ij} \quad (2)$$

where c_k represents the model bias for GMM k, $\eta_{E,i}$ and $\eta_{S,j}$ are random effects that quantify event- and site-specific biases (referred to as event terms and site terms, respectively), and ε_{ij} is the remaining portion of the within-event residual. The total linear site response, f_1, relative to a reference condition of V_{S30} = 760 m/s is given as (Stewart et al., 2017):

$$(f_1)_j = F_{lin}^{erg} + \eta_{S,j} \quad (3)$$

where F_{lin}^{erg} is the ergodic site response model in the GMM. Uncertainties in site responses (i.e., f_1) are equivalent to the standard error of η_S. These site response uncertainties are not considered in the analyses presented in this paper. Site-to-site aleatory variability (denoted ϕ_{S2S}) is closely related to the standard deviation of $\eta_{S,j}$ across all sites and hence depends on the predictive power of the ergodic model. The benefits of utilizing mHVSR are quantified by their effect on ϕ_{S2S}.

Site Attributes Used in Model Development

Sub-Regionalization of California

The model development in this paper considers the potential for sub-regional differences in mHVSR-η_S relationships. Subregions are defined using the physiographic province map of Buckreis et al. (2023), which was modified from Chiou et al. (2010). Figure 2 shows the subregion locations. While the subregions in Figure 2 were developed for path investigations, because they represent different physiographic conditions, they may also influence site response.

The map differentiates several coastal geologic domains: the Coast Range, aligned with relatively straight portions of the San Andreas fault; the east–west–trending Transverse Range south of the San Andreas “big bend”; and the Peninsular Range of coastal southern California. Major mountain provinces include the Klamath Mountains, Cascades, and Sierra Nevada. Significant basin structures include the Central Valley, the Colorado Desert (including the Imperial Valley), and large basins within the Peninsular Range.

Figure 3 compares subregional site terms. Arithmetic means of site terms are relatively flat with period and near or slightly below zero in the South Coast, Eastern California Shear Zone, and the Basin and Range. The North Coast, Bay Area, and Central Valley exhibit overprediction at short periods and underprediction of long periods (the Central Coast and Colorado Desert shapes are similar but are unbiased at short periods). The Sierra Nevada and Northeastern California regions tend to be unbiased at short periods and overpredicted at long periods.

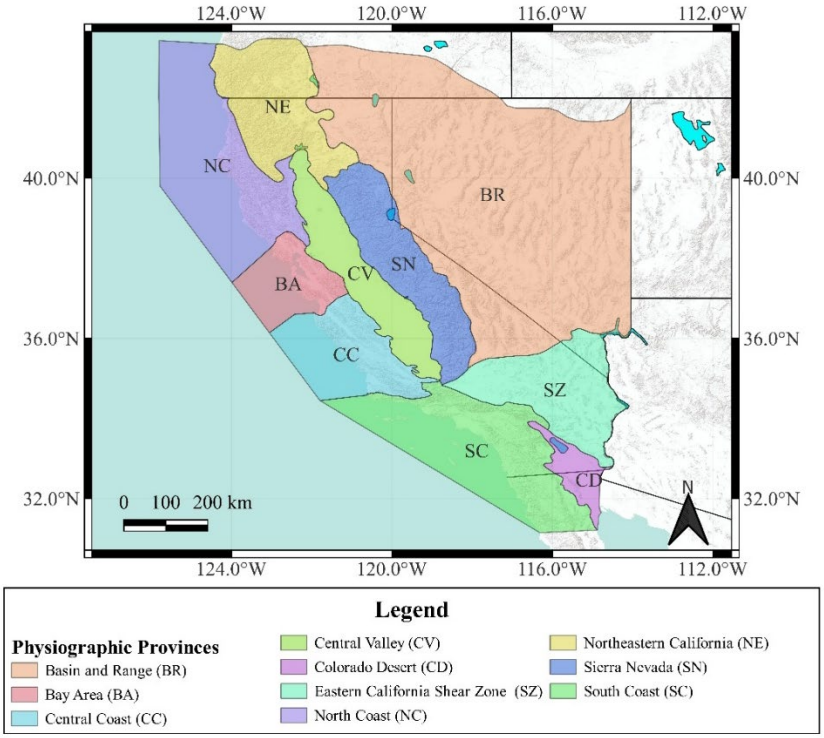

Figure 2. California sub-regions defined from physiographic provinces (from Buckreis et al. 2023).

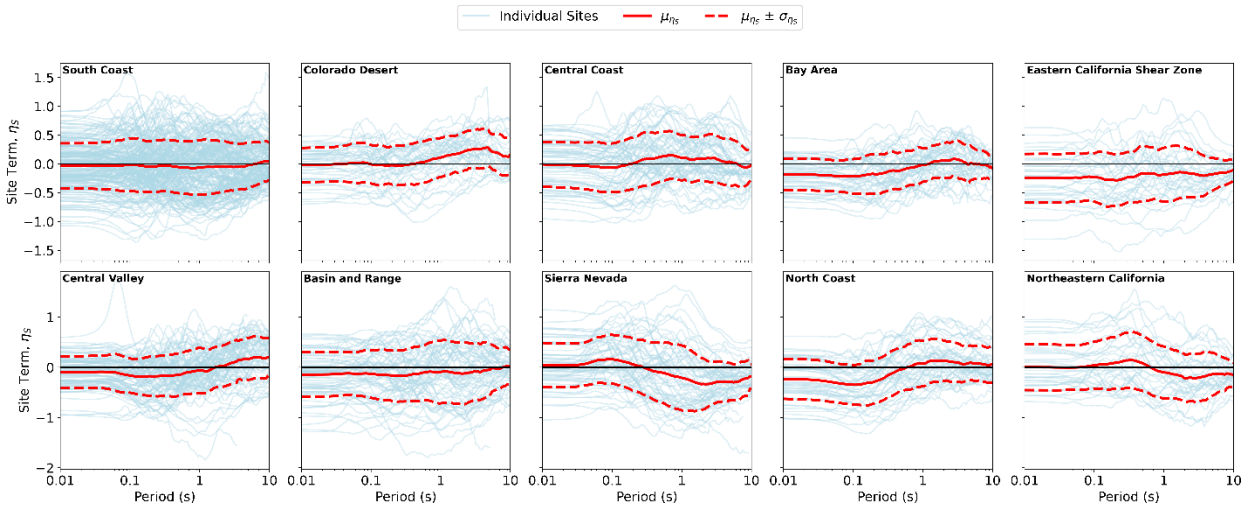

Figure 3. Individual site terms and their means for 10 California sub-regions.

mHVSR Peak Identification

We assessed whether peaks were present in the mHVSR for each of the 685 sites. Peak detection used the regression tree algorithm developed by Wang et al. (2023) that models mHVSR curves as a stepwise function. A peak is identified when its amplitude from tree regression, relative to neighboring amplitudes, exceeds a threshold (typically a factor of two), and when amplitude variability is sufficiently low. Peaks are fitted using a Gaussian pulse function (Ghofrani and Atkinson, 2014):

$$F_{H/V,i}(f) = c_{0,HV,i} + c_{1,HV,i}\exp\left[-\frac{1}{2}\left(\frac{\ln(f/f_{p,HV,i})}{2w_{p,HV,i}}\right)^2\right] \quad (4)$$

where $f_{p,HV,i}$ is the fitted peak frequency for the i^{th} HVSR peak, $c_{1,HV,i}$ is the peak amplitude relative to neighboring ordinates, $w_{p,HV,i}$ is the peak width, $c_{0,HV,i}$ is a frequency-independent constant, and f is frequency in Hz. Absolute amplitudes of peaks are computed as, $a_{p,HV,i} = c_{0,HV,i} + c_{1,HV,i}$. Figure 4a shows example tree regression results (steps) fitted to a mean mHVSR curve and the peak fitting using Eq. (4). The peak parameters in Eq. (4) are considered in site response model development.

The lowest identified peak frequency may not correspond to the fundamental site frequency (f_0) for the following reasons: (1) for deep basin sites, f_0 may fall below the usable frequency range of mHVSR (~ < 0.1 Hz), where ordinates tend to be unstable (Ornelas et al., 2024a); and (2) horizontal component amplitudes near f_0 may not exceed the instrument noise threshold. We do not consider mHVSR data below 0.1 Hz, but ramps at low frequencies can still obscure low-frequency resonances. For this reason, we instead utilize the dominant peak (f_d), defined as the largest amplitude mHVSR peak relative to other peaks, the frequency of which is denoted as ($f_{p,HV}$) (amplitudes and widths are denoted similarly).

Approximately 50% of sites exhibited no mHVSR peaks, 42% exhibited a single peak, and 8% exhibited multiple peaks—consistent with the findings of Wang et al. (2023) for California. The average V_{S30} values for these groups were 514 m/s (no peak), 417 m/s (one peak), and 342 m/s (multiple peaks). This trend is consistent with expectations, as stiffer sites are generally less prone to resonance compared to softer near-surface soils overlying stiffer materials.

Site Terms Peak Identification

We identify resonant features in site terms utilizing a similar tree regression algorithm (*srPeak* developed by Buckreis, 2022). Criteria considered for peak identification are (Buckreis et al., 2024):

1. Peak should be relatively localized (i.e., period range not too wide).
2. Peak should have sufficiently large mean amplitude relative to adjacent periods.
3. The peak is meaningful (i.e. sufficiently small uncertainties in amplitudes and periods).

These criteria were applied in the manner presented by Buckreis et al. (2024) for a soft soil region with the exception of thresholds related to peak width and amplitude that were modified for this application. The modifications were made because the original algorithm identified peaks for site response features that we considered to be too large in width and too small in amplitude to classify as peaks. Figure 4b shows an example mean site response – period plot for a site with a peak.

Approximately 51% of sites do not exhibit a site term peak, while 49% have at least one peak, and 7% have two peaks. Among the sites with an mHVSR peak, 60% have a site term peak, whereas among the sites without an mHVSR peak, 40% have a site term peak. Viewed a different way, 40% of sites without a site term peak exhibit an mHVSR peak whereas 60% of sites with a site term peak also have an mHVSR peak. It is likely possible to improve these performance metrics with refinements of the peak identification protocols.

Site term peaks were fit using a Gaussian pulse function, with slight modifications applied on either side of the peak period. This fitting utilized the function *minimize* from the Python package *scipy.optimize* (Virtanen et al. 2020). The initial fitted Gaussian functional form is:

$$F_{\eta_s,i}(T) = c_{0,\eta_s,i} + c_{1,\eta_s,i} \exp\left[-\left(\frac{ln(T f_{p,\eta_s,i})}{\omega_{p,\eta_s,i}}\right)^2\right] \tag{5}$$

The fit parameters align with those previously defined for use with Eq. (1) except that T denotes oscillator period, and $\omega_{p,\eta_s,i}$ denotes peak width. Total amplitude, $\alpha_{p,\eta_s,i} = c_{0,\eta_s,i} + c_{1,\eta_s,i}$. Modifications around the peak period is introduced using a smooth Gaussian transition function,

$$\Phi(T) = \exp\left(-\left(\frac{T - T_{p,i}}{\omega_{p,\eta_s,i}}\right)^2\right) \tag{6}$$

where $T_{p,i} = 1/f_{p,i}$ represents the period at which the peak is located, and $\omega_{p,i}$ represents the width of the Gaussian pulse. Eqs. (5) and (6) are combined as follows:

$$\hat{\eta}_S(T) = \begin{cases} F_{\eta_s,i}(T)\Phi(T) + P_L\big(1 - \Phi(T)\big), & T < T_p \\ F_{\eta_s,i}(T)\Phi(T) + P_R\big(1 - \Phi(T)\big), & T > T_p \end{cases} \tag{7}$$

where P_L denotes the plateau value of the site term on the left side of the peak in oscillator-period space, while P_R denotes the corresponding value on the right side. Figure 5 provides an example of a site-term curve peak fitted with the function in Eq. (7). The resulting peak parameters are subsequently related to mHVSR, as discussed in the following sections.

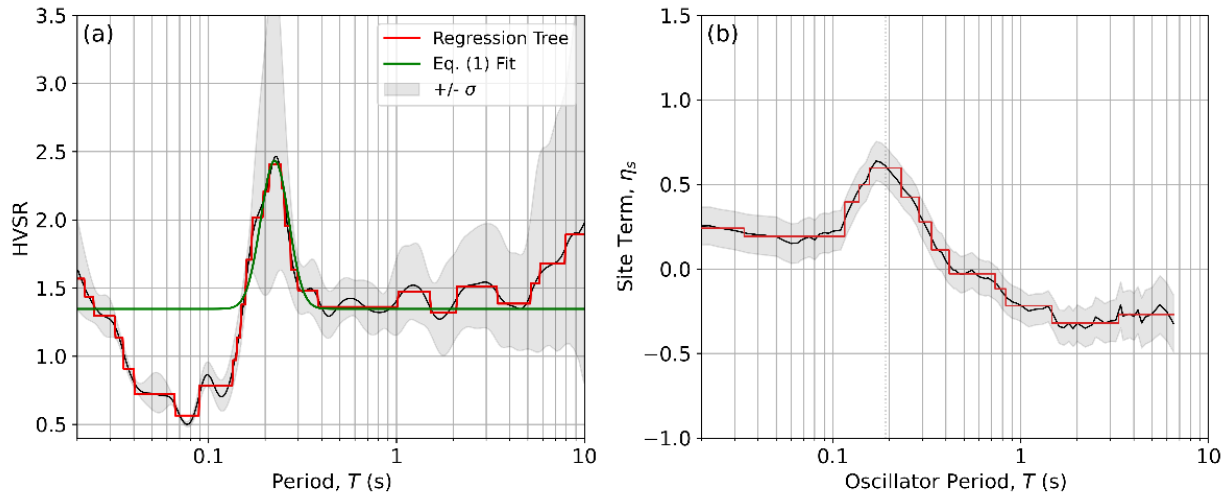

Figure 4. Tree regression results and fitted peaks for (a) mHVSR and (b) site terms.

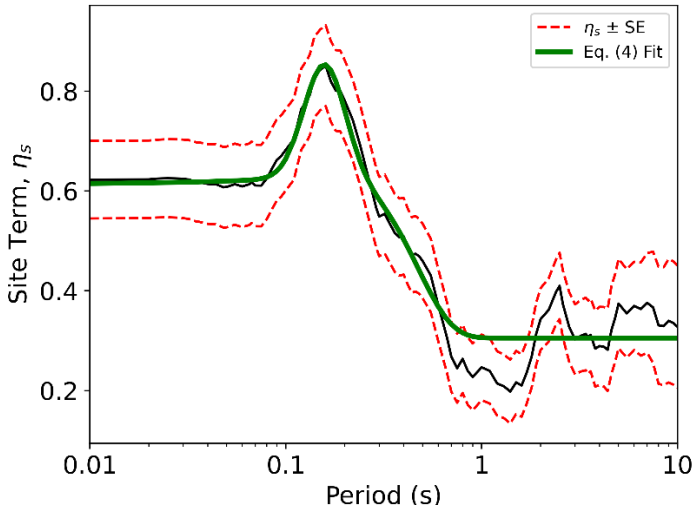

Figure 5. Example of a site term peak fit by the modified Gaussian pulse Eq. (4).

Peak-Based mHVSR Model

Sites with mHVSR Peaks

In this section, we develop a model for predicting resonant features in site terms given peak parameters from mHVSR. Site response plateau and ramp effects, which are described further in Ornelas et al. (2024b), are not modeled but represent a potential extension for future work. Moreover, we also do not model second peaks for simplicity, but we could integrate modeling related to second peaks from mHVSR in future work. We only consider peaks in mHVSR and site response that are in the frequency range of 0.1-20 Hz.

The evaluation begins by partitioning the data into a training set (80%) and a testing or validation set (20%). The evaluation is performed separately for each of the California subregions in Figure 2. Stronger correlations were observed between mHVSR and site response peak parameters within sub-regions than when the data was analyzed as a whole for the state. Subregions that exhibit similar behaviors are combined. Specifically, we combine the regions in the eastern half of California, including the Sierra Nevada, Eastern California Shear Zone, Basin and Range, and Northeastern California, under the label "Eastern California." We combine the northwestern portion of California, encompassing the North Coast, Central Coast, and Bay Area, under the label "Western California". The South Coast and Colorado Desert subregions are combined, under the label "Southern California". Similarly, the Central Valley is not combined with adjacent regions due to its distinct physiographic characteristics.

Figure 6 presents the results for the Eastern California region. In Figure 6a, the mHVSR dominant frequencies ($f_{p,HV}$) are strongly correlated with the site term peak frequencies $f_{p,\eta_s} = 1/T_p$ with no outliers. Following precedent from Kwak et al. (2017), a linear expression was used to relate these frequencies to each other,

$$log_{10}(\hat{f}_{p,\eta_S}(f_{p,HV})) = p_0 + p_1\, log_{10}(f_{p,HV}) \tag{5}$$

where p_0 and p_1 are regression coefficients. The fit was constrained to avoid impacts from outliers (while not present for Eastern California, they are present for other subregions). The constraint was applied by limiting p_1 to a range of [0.9 -- 1.1] and p_0 to a range of [0 -- 0.02]. We acknowledge that careful consideration of outliers may not need the use of model constraints, which we plan to do in future work. Model fitting was performed using the *curve_fit* function from the *scipy.optimize* package (Virtanen et al., 2020).

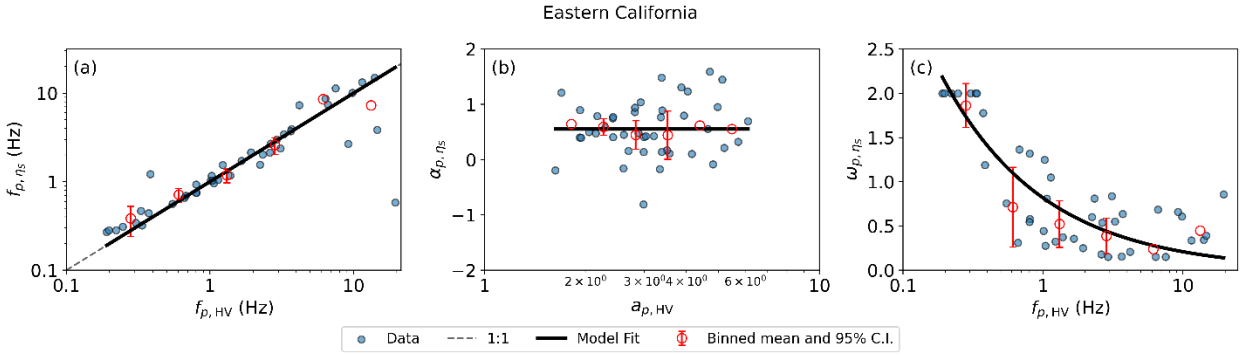

Figure 6. Comparisons of mHVSR and site term peak parameters for the Eastern California combined sub-regions: (a) peak frequencies, (b) peak amplitudes, and (c) widths

Figure 6b shows that the mHVSR absolute peak $a_{p,HV}$ has no predictive power for the site term absolute peak α_{p,η_s}. We investigated correlations with other parameters such as relative peak amplitude ($c_{1,HV}$), but found a relatively flat trend, similar to what was found using $a_{p,HV}$. As a result, a fixed value of α_{p,η_s}= 0.55 natural log units is recommended regardless of mHVSR parameters. Figure 6c shows an exponential decay in site term peak width (ω_{p,η_s}) as $f_{p,HV}$ increases (i.e., higher frequency peaks are narrower). This relationship was fit as,

$$\widehat{\omega}_{p,\eta_S}(f_{p,HV}) = w_0 \cdot f_{p,HV}^{-w_1} \tag{6}$$

where w_0 represents the scaling coefficient and w_1 denotes the exponential decay rate. Eq. (6) shows that high frequency peaks are narrower than low frequency peaks. The mHVSR peak width, $w_{p,HV}$, was considered as a conditioning variable for site response peak width, $\widehat{\omega}_{p,\eta_S}$, but had weaker correlation. Similar results have been obtained previously using data from the California Delta region (Buckreis et al., 2024) using a different functional form. The fitting of Eq. (6) utilized the same curve fit package as with Eq. (5). For the analyses presented in this paper, constraints were implemented for the Southern California and Western California regions to capture the larger width values at lower peak frequencies, which would otherwise be obscured by outliers. The values used to constrain the models were, w_0: [0.5 -- 0.58] and [0.85 -- 0.95]; w_1: [0.2 -- 1.0] and [0.4 -- 0.6] for Western and Southern California, respectively.

We recognize that the use of regression constraints is non-standard, and some explanation is warranted here. The outliers that are screened by the application of these constraints are anticipated to result from sub-optimal peak identification in both mHVSR and site response. Identifying the features where mis-identifications occur is a time-consuming task that has yet to be completed. The preparation of this paper could not wait for that process to be completed due to

a contractual deadline. As a result, to allow preliminary results to be prepared at the current stage of model development, constraints are applied to allow physically meaningful results to be obtained. These results will be refined in ongoing/future work.

Figures 7-9 show similarly formatted results for Western California, Southern California, and Central Valley, respectively. The modeling approach for each was the same as discussed above for Eastern California, but the data trends have differences:

- Western California and Southern California have greater scatter in the peak frequency relationships (Figure 7a, 8a) and peak width relationships (Figure 7c, 8c). The fit results follow similar trends, but the peak widths on average are lower for Western California.
- Central Valley results (Figure 9) are similar to those for Eastern California except for smaller peak widths for Central Valley sites.

For each of the regions, the relationships between peak frequencies are very similar. This is likely attributed to the same constraints used to identify a peak that were utilized for each of the 685 sites in the dataset. However, the Western and Southern California subregions have more outliers than Eastern California or the Central Valley, and those outliers are more commonly encountered at low frequencies (< 0.5 Hz) and high frequencies (> 3 Hz). Relatively few outliers occur at intermediate frequencies for all subregions.

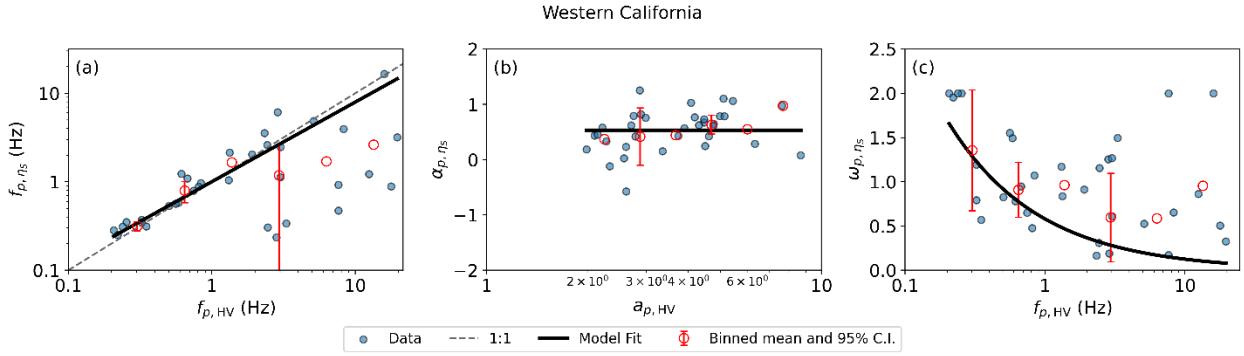

Figure 7. Comparisons of mHVSR and site term peak parameters for the Western California combined sub-regions: (a) peak frequencies, (b) peak amplitudes, and (c) widths.

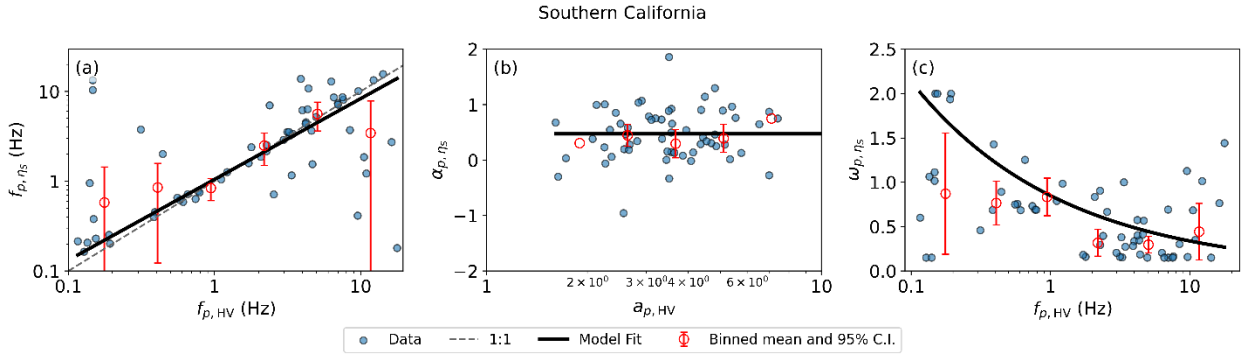

Figure 8. Comparisons of mHVSR and site term peak parameters for the Southern California combined sub-regions: (a) peak frequencies, (b) peak amplitudes, and (c) widths.

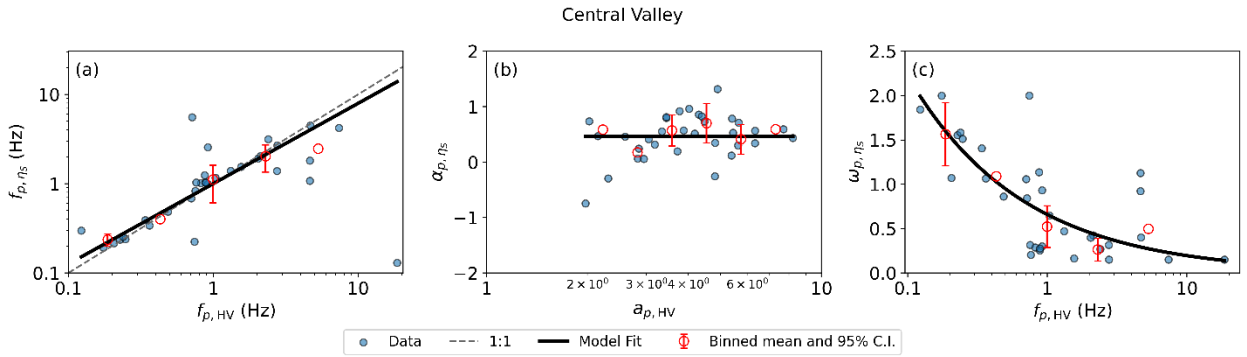

Figure 9. Comparisons of mHVSR and site term peak parameters for the Central Valley sub-region: (a) peak frequencies, (b) peak amplitudes, and (c) widths

The resulting coefficients from fitting the three different models across the various subregions are summarized in Table 2. The table also presents coefficients if the same modeling framework is applied to the full statewide data. The peak frequency ($\hat{f}_{p,\eta_S}$) parameters exhibit minimal variations between subregions, which is likely influenced by the use of bounding parameters in the regressions. The lower peak widths in Western California are evident by its smallest w_0 value and its highest decay parameter (w_1). The amplitude parameter ranges from 0.48-0.66 natural log units. The amplitude coefficient for All California is most similar to the value for Southern California, both of which are low compared to other regions. The similarity of these results is due to the large amount of data from stations in Southern California.

Table 2. Summary of the sub-regional coefficients.

Site Region	p_0	p_1	$\hat{\alpha}_p$	w_0	w_1
Eastern California	1.62e-10	1.00	0.555	0.818	0.593
Western California	7.71e-15	0.90	0.535	0.580	0.666
Southern California	0.019	0.90	0.475	0.850	0.400
Central Valley	5.87e-15	0.90	0.659	0.659	0.528
All California	1.05e-13	1.00	0.505	0.750	0.550

The model parameters were used to predict a Gaussian pulse function that characterizes the resonant features of the site terms. This predictive model follows a similar functional form to that used in Equations (1) and (2), and is defined as follows:

$$G(T, f_{p,HV},\ site\ region) = \hat{\alpha}_{p,\eta_S} \exp\left[-\left(\frac{ln(T\hat{f}_{p,\eta_S}))}{\hat{\omega}_{p,,\eta_S}}\right)^2\right] \tag{7}$$

where $\hat{\alpha}_{p,\eta_S}$ represents the peak amplitude in natural log units, $\hat{f}_{p,\eta_S}$ is the peak frequency derived from Eq. (5), and $\hat{\omega}_{p,\eta_S}$ is the peak width derived from Eq. (6), and T represents the 105 oscillator periods from BSSA14 ranging from 0.01 sec to 10 sec. The full functional form of the site component of the GMM is then defined as:

$$F_{S,ij}(T, V_{S30,j}, \delta_{z_1,j}\ f_{p,HV},\ site\ region) = F_{lin,j}(V_{S30,j}, \delta_{z_1,j}) + F_{nl,ij} + G(T, f_{p,HV}, site\ region) \tag{8}$$

where $V_{S30,j}$ represents the time-averaged shear-wave velocity in the upper 30 meters, δ_{z_1} denotes the differential depth parameter, and $F_{lin,j}$ and $F_{nl,ij}$ represent the linear and nonlinear site amplification terms, respectively. The function in Eq. (8) applies only in the case that mHVSR identifies at least one peak. The next section presents a model for cases where no peak is identified.

Sub-Regional Site Response Adjustments for Non-Peak Sites

As noted previously 51% of sites do not have an identified mHVSR peak using the Wang et al. (2023) algorithm. The model in the previous section, which is conditioned on mHVSR peak parameters, cannot be applied to non-peak sites. In this section we present adjustments to the SS14 model that can be applied conditionally on two pieces of information: (1) the lack of an mHVSR peak and (2) the location of the site within one of the subregions shown in Figure 2 (without use of the combined subregions from the previous section).

Figure 10 presents the site terms and their corresponding means (indicated in red) for each of the sub-regions from Figure 2, over the period range of 0.01–10 sec, for sites without mHVSR peaks. Additionally, the means for all sites in the respective subregions, regardless of whether peaks were observed, are shown in green. Subregional adjustments are calculated as the difference between the subregional natural log mean and the natural log mean for all sites with mHVSR data. A similar approach was used by Kwak et al. (2017), but without partitioning the data into subregions. This methodology is designed to isolate subregional variations under the specific condition of "no-peak sites," while accounting for potential small discrepancies between the current dataset and the overall GMDB, as described by Buckreis et al. (2025).

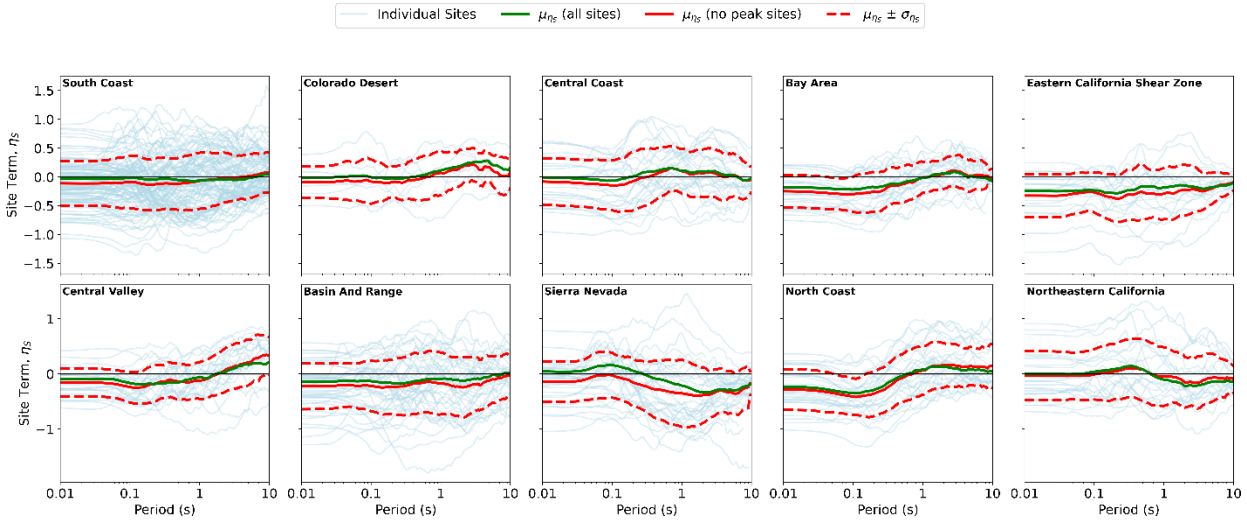

Figure 10. Individual site terms, their means, and means ± one standard deviation for each subregion for non-peak sites. The green line represents the mean of all the data in the study.

The South Coast subregion has the most sites and the trends with period are relatively flat. One observed trend is negative means at short periods and positive means at long periods (e.g., Central Valley, Bay Area, Central Coast, North Coast), which is expected for regions with sedimentary basins. The corollary is positive and negative means at short and long periods,

respectively, which is expected for regions predominantly with shallow soils (e.g., Sierra Nevada, Northeastern California). Based on similar trends, we combined subregions as follows:

- Sierra Nevada and Northeastern California are combined into Northeastern California (positive and negative means at short and long periods, respectively).
- Central Coast, Bay Area, and North Coast are combined into Western California (negative and positive means at short and long periods, respectively).
- South Coast, Colorado Desert, Eastern California Shear Zone, and Basin and Range into Southeastern California (Generally flat across periods).
- Central Valley was not combined.

Figure 11 repeats this information from Figure 10 but for the combined subregions.

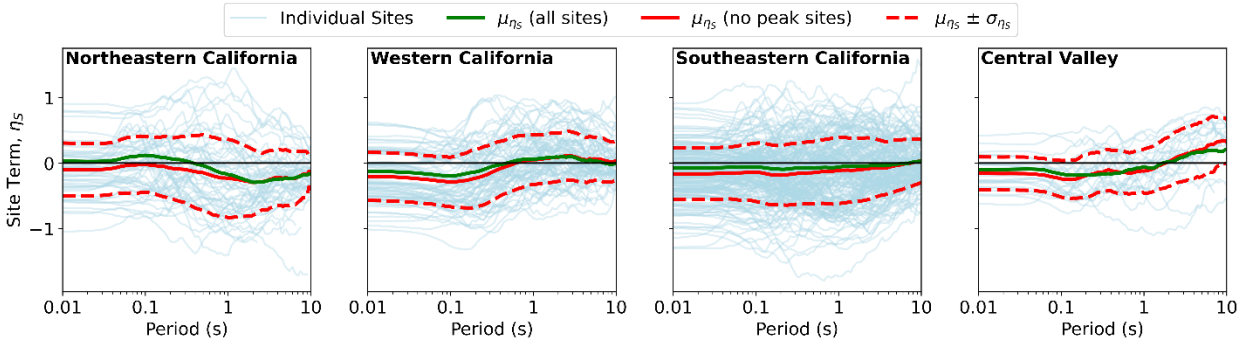

Figure 11. Individual site terms, their means, and means ± one standard deviation for combined subregions for non-peak sites. The green line represents the mean of all the data within the respective combined subregions.

Combined Model

The models described in the two previous subsections, conditioned on peaks being identified or not in mHVSR, can be synthesized as follows:

$$F_{peak}(f_{p,HV}, T, subregion) = \begin{cases} G(f_{p,HV}, T, subregion) & if\ mHVSR\ peak \\ Coeff(T, subregion) & otherwise \end{cases} \quad (9)$$

where $G(f_{p,HV}, T)$ is from Eq. (7) and *Coeff.* Represent the coefficients across the 105 BSSA14 oscillator periods based on the subregion where the site is located.

Evaluation of Models

Residuals Analysis

We compute residuals and site terms using the two models described in the previous sections. The results are shown in Figure 12a for the training data and in Figure 12b for the testing data. Overall, the trend is generally flat across periods, and close to zero. However, for the training dataset, there is overprediction of the model at long periods. For the testing dataset there is slight overprediction at short periods.

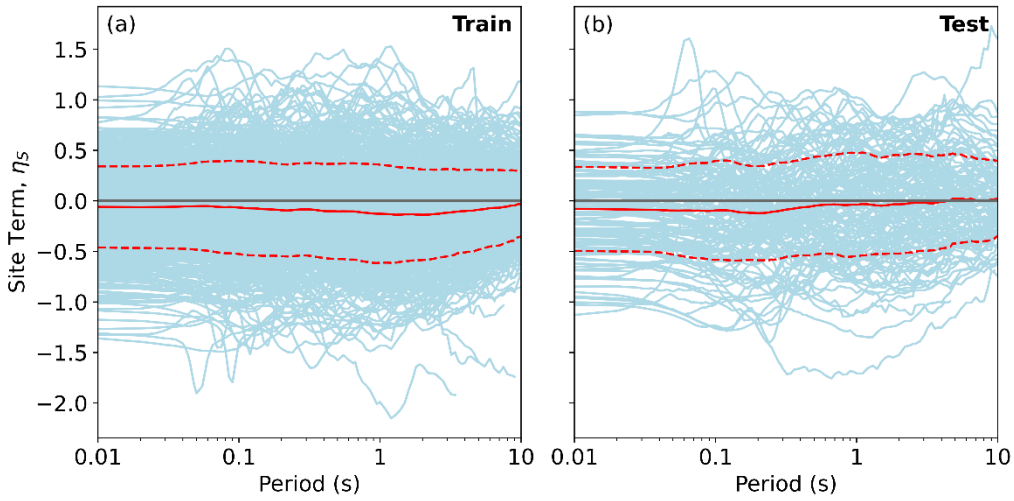

Figure 12. Comparison between residuals for training and testing datasets. (a) training set, (b) testing set. The red lines show means and means ± one standard deviations of the site terms.

Figure 13 shows plots of residuals vs period for the subsets of sites with and without mHVSR peaks using both the testing and training data. The peak sites are over-predicted at long periods, with the maximum over-prediction near 1.0 sec. The non-peak sites are over-predicted at short periods. Figure 14 shows similar residuals plots for both training and testing sites with peaks in the frequency ranges of 0.3-0.5 Hz (Figure 14a) and 3-5 Hz (Figure 14b). The sites with low-frequency peaks (0.3-0.5 Hz) are over-predicted at long periods, suggesting that the mHVSR peak correction may be too strong. The sites with mid-frequency peaks (3-5 Hz) do not exhibit bias in the corresponding periods range (0.2-0.35 sec), but are biased at shorter and longer periods, suggesting that model refinements to address performance problems away from the peaks could be fruitful.

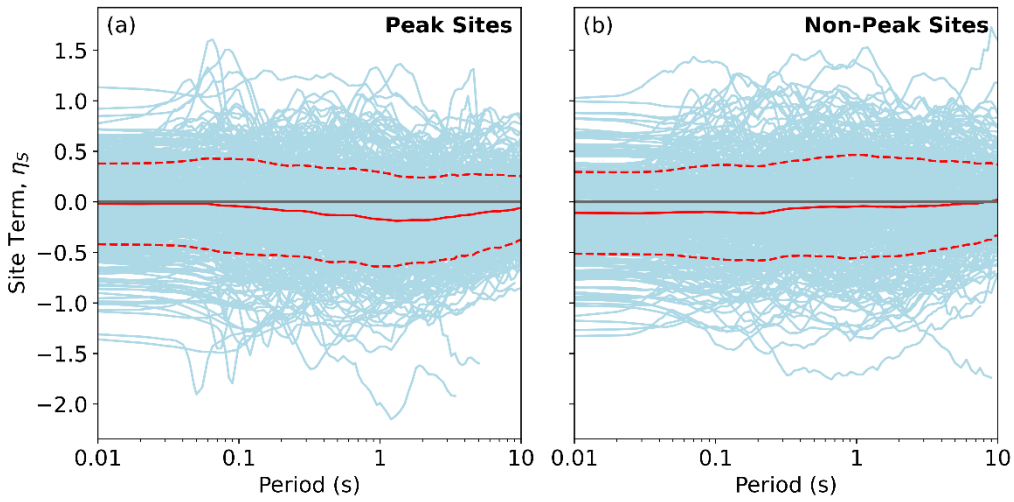

Figure 13. Comparison between residuals for peak and non-peak sites. (a) peak sites, (b) non-peak sites. The red lines show means and means ± one standard deviations of the site terms.

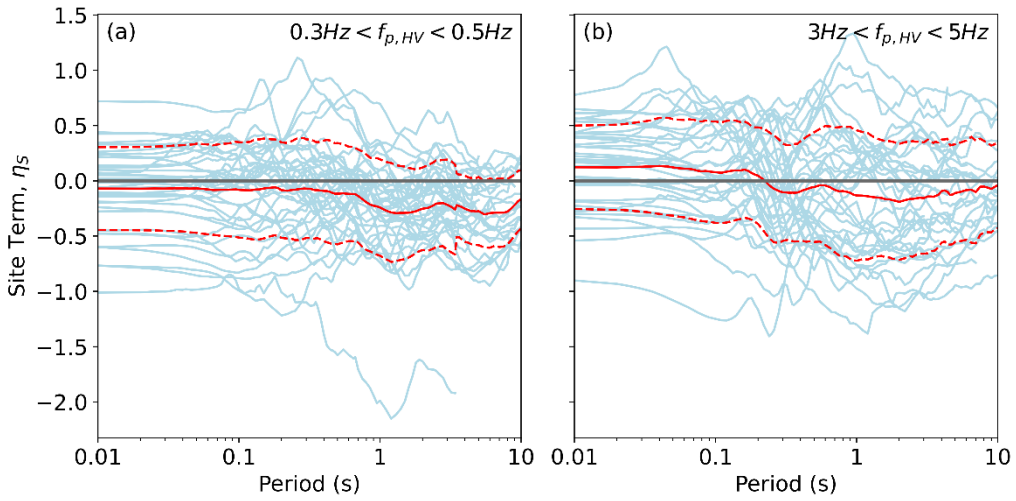

Figure 14. Comparison between residuals for peak sites at certain frequency ranges. (a) 0.3 Hz < $f_{p,HV}$ < 0.5 Hz, (b) 3 Hz < $f_{p,HV}$ < 5 Hz. The red lines show means and means ± one standard deviations of the site terms.

Site-to-Site Aleatory Variability

We evaluate site-to-site aleatory variability (ϕ_{S2S}) as the standard deviation of site terms derived using different models and different groups of sites. Figure 15 shows ϕ_{S2S} values for the combined training and testing data sets for the subsets of sites with and without peaks, using both the mHVSR-informed models (this paper) and models conditioned only on V_{S30} (SS14). For sites with peaks, the reduction of ϕ_{S2S} ranges from negligible for periods < 0.2 sec to values as large as 0.05 to 0.06 at long periods (up to about 13% reduction). Uncertainty reduction is minimal for no-peak sites. The increased uncertainty reduction at long periods for peak sites is a feature that has been observed in prior work (e.g., Esteghamati et al. 2022; Pinilla-Ramos et al. 2022).

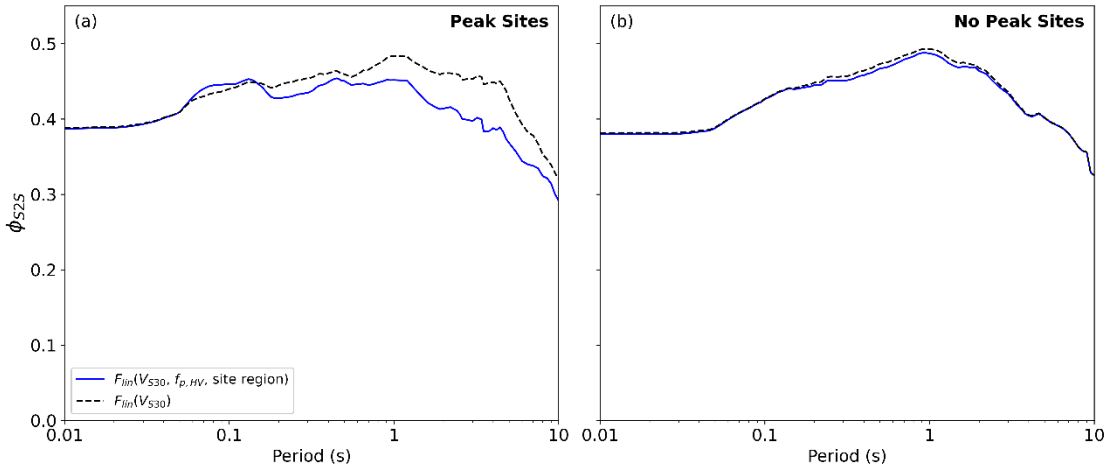

Figure 15. Comparison of site-to-site variability (ϕ_{S2S}) for the models presented in this paper to those derived using models that only consider V_{S30}-scaling, (a) sites with mHVSR peaks; (b) sites without peaks

Conclusions

In this paper, we have presented and applied a framework for mHVSR modeling conditioned on the presence or lack of peaks. Our motivation for developing a model structured in this manner is two-fold. First, it has proven to be effective in soft soil areas with strong resonances (Buckreis et al. 2024), although that leaves open the question of broader applicability. Second, the alternative of using mHVSR ordinates directly faces the challenge of those ordinates generally considered to be unreliable by experts on mHVSR measurement.

Machine learning based approaches, using a similar dataset, indicate that models conditioned on mHVSR can achieve lower site-to-site standard deviations than provided by SS14 or achieved in the present work (Ornelas et al. 2026). We are actively working on developing an alternative to the model in this paper that is conditioned directly on mHVSR ordinates rather than on peaks fit to the mHVSR data. Preliminary versions of these high-order models provide better accuracy for the training dataset but have not been thoroughly evaluated against testing data. High order models are difficult to explain, may be prone to overfitting and poor extrapolation, and must be treated with skepticism and carefully vetted. The aim is to evaluate the performance of two competing models derived using the same data set.

The model presented in this paper is preliminary and is not ready for application. The preliminary models presented here are in the process of being refined in the following respects: (1) improve peak selection protocols for mHVSR and site response to potentially reduce outliers; (2) consider geomorphic provinces as an alternative to physiographic subregions for site categorization in the modeling framework; (3) incorporate ramps into the modeling approach.

Acknowledgements

This study was funded under CSMIP Agreement #1023-018, and we sincerely acknowledge their generous support. Additionally, we would like to express our gratitude to Pacific Gas and Electric (PG&E) for their partial funding contribution to this research.

References

Abrahamson NA, Youngs RR 1992. A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America*, 82, 505–510.

Anderson JG, Brune JN 1999. Probabilistic seismic hazard assessment without the ergodic assumption, *Seismological Research Letters,* 70, 19–28.

Boore DM, Stewart JP, Seyhan E, Atkinson GM 2014. NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthquake Spectra,* 30, 1057-1085.

Bozorgnia Y, Abrahamson NA, Al Atik L, et al. 2014. NGA-West2 research project. *Earthquake Spectra*. 30, 973-987.

Buckreis T, Winders A, Wang P, Brandenberg S, Stewart J 2021. Microtremor Data Collected in Sacramento-San Joaquin Delta region of California*, DesignSafe-CI* PRJ-3106.

Buckreis TE 2022. Customization of path and site response components of global ground motion models for application in Sacramento-San Joaquin Delta region of California, *Doctoral dissertation*, University of California, Los Angeles, Los Angeles, CA.

Buckreis TE, Stewart JP, Brandenberg SJ and Wang P. 2023. Subregional anelastic attenuation model for California. *Bulletin of the Seismological Society of America* 113(6), 2292–2310.

Buckreis TE, Stewart JP, Brandenberg SJ, Wang P 2024. Small-strain site response of soft soils in the Sacramento-San Joaquin Delta region of California conditioned on V_{S30} and mHVSR. *Earthquake Spectra*. 40, 230-260.

Buckreis TE, Nweke CC, Wang P, Brandenberg SJ, Ramos-Sepulveda ME, Shams R, Mohammed S, Pretell R, Mazzoni S, Zimmaro P, and Stewart JP 2025. A global Application Programming Interface–enabled earthquake ground motion relational database for engineering applications. *Earthquake Spectra*. doi:10.1177/87552930251344978

Campbell KW and Bozorgnia Y 2014. NGA-West2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthquake Spectra* 30, 1087–1115.

Chiou BSJ, Youngs RR, Abrahamson NA, and Addo K 2010. Ground motion attenuation model for small-to-moderate shallow crustal earthquakes in California and its implications on regionalization of ground-motion prediction models, *Earthquake Spectra*, 26, 907–926.

Esteghamati MZ, Kottke AR, Rodriguez-Marek A 2022. A data-driven approach to evaluate site amplification of ground-motion models using vector proxies derived from Horizontal-to-Vertical spectral ratios. *Bulletin of the Seismological Society of America* 112, 3001–3015.

GEOVision 2016. Surface wave measurements report, Riverside County, California, Report 16192-01 Rev 2. Prepared for State of California Department of Conservation, California Geological Survey, *Strong Motion Instrumentation Program*, 167.

GEOVision 2018. Surface wave measurements, Santa Clara, Santa Cruz, San Benito, and Monterey counties, Report 18045-01. Prepared for State of California Department of Conservation, California Geological Survey, *Strong Motion Instrumentation Program*, 211.

Ghofrani H, Atkinson GM, Goda K 2013. Implications of the 2011 M9.0 Tohoku Japan earthquake for the treatment of site effects in large earthquakes. *Bulletin of Earthquake Engineering*, 11, 171–203.

Ghofrani H, Atkinson GM 2014. Site condition evaluation using horizontal-to-vertical response spectral ratios of earthquakes in the NGA-West2 and Japanese databases, *Soil Dynamics and Earthquake Engineering* 67, 30–43.

Incorporated Research Institutions for Seismology (IRIS) DMC 2020. Web Services, available at https://service.iris.edu/ (last accessed May 2020).

Kwak DY, Stewart JP, Mandokhail SJ, and Park D 2017. Supplementing V_{S30} with H/V spectral ratios for predicting site effects. *Bulletin of the Seismological Society of America*, 107, 2028–2042.

Kwak DY, Ahdi SK, Wang P, Zimmaro P, Brandenberg SJ, Stewart JP 2021. Web portal for shear wave velocity and HVSR databases in support of site response research and applications. *UCLA Geotechnical Engineering Group*. https://doi.org/10.21222/C27H0V

Lachet D, Hatzfeld C, Bard P-Y, Theodulis N, Papaioannou C, Savvaidis A 1996. Site effects and microzonation in the city of Thessaloniki (Greece): Comparison of different approaches. *Bulletin of the Seismological Society of America*, 86, 1692–1703.

Lermo J, Chávez-García FJ 1993. Site effect evaluation using spectral ratios with only one station. *Bulletin of the Seismological Society of America* 83, 1574–1594.

Lermo J, Chávez-García FJ 1994. Are microtremors useful in site response evaluation? *Bulletin of the Seismological Society of America,* 84, 1350–1364.

Molnar S, Cassidy JF, Castellaro S, et al. 2018. Application of microtremor horizontal-to-vertical spectral ratio (MHVSR) analysis for site characterization: State of the art. *Surveys in Geophysics*, 39, 613–631.

Mohammed S, Shams R, Nweke CC, Buckreis TE, Kohler MD, Bozorgnia Y, Stewart JP 2024. Usability of Community Seismic Network recordings for ground-motion modeling, *Earthquake Spectra*, 40, 2598-2622.

Nakamura Y 1989. A method for dynamic characteristics estimation of subsurface using microtremors on the ground surface. *Quarterly Report of Railway Technical Research Institute*, 30, 25-33.

Nogoshi M, Igarashi T 1970. On the amplitude characteristics of microtremor (part 1). *Journal of Seismological Society of Japan*, 24, 26–40.

Nogoshi M, Igarashi T 1971. On the amplitude characteristics of microtremor (part 2). *Journal of Seismological Society of Japan*, 24, 26–40.

Nweke CC, Stewart JP, Wang P and Brandenberg SJ. 2022. Site response of sedimentary basins and other geomorphic provinces in southern California. *Earthquake Spectra* 38, 2341–2370.

Nweke C, Ornelas F, Shams R, Vyas P 2024. Microtremor Horizontal-to-Vertical Spectral Ratio (mHVSR) Data Collection at Baldwin Hills, 2023, in Microtremor Measurements Performed Within the Southern California Region. *DesignSafe-CI*.

Nweke C, Buckreis T, Ornelas F 2024. Microtremor Horizontal-to-Vertical Spectral Ratio (mHVSR) Data Collection at Brentwood, 2022, in *Microtremor Measurements Performed Within Southern California*. *DesignSafe-CI*. https://doi.org/10.17603/ds2-d56j-h615

Ornelas F, de la Torre C, Nweke C, Buckreis T, Wang P, Bradley B, Brandenberg S, Stewart J 2023. Microtremor Horizontal-to-Vertical Spectral Ratio (mHVSR) Data Collection at California Downhole Vertical Array Sites, 2022, in *Microtremor Horizontal-to-Vertical Spectral Ratio (mHVSR) Site Characterization of California Vertical Arrays* [Version 2]. *DesignSafe-CI*. https://doi.org/10.17603/ds2-by4m-ed67

Ornelas FJ, Nweke CC, de la Torre CA, Wang P, Mai TD, Cox BR, Brandenberg SJ, Stewart JP 2024a. Reliability of low frequency mHVSR ordinates. *18th World Conference on Earthquake Engineering Conference*, Italy.

Ornelas, FJG, Buckreis TE, Nweke CC, Wang P, delaTorre C, Brandenberg SJ, Stewart JP 2024b. Preliminary observations in support of the development of an ergodic site response model in California conditioned on V_{S30} and HVSR parameters, *8th International Conference on Earthquake Geotechnical Engineering,* Osaka, Japan, May 2024, Paper OS-36-06, 6 pages

Ornelas, F, Buckreis T, Brandenberg S, Stewart J. 2025a. Microtremor measurements collected from permanent stations in and around California and Nevada between 2000 and 2025, in *Microtremor Horizontal-to-Vertical Spectral Ratios (mHVSR) from Permanently Installed Broadband Seismometers*. *DesignSafe-CI.* https://doi.org/10.17603/ds2-4ehv-c597

Ornelas F, Stapleton J, Buckreis T, Nweke C, Stewart J 2025b. Microtremor Data Collection in the San Fernando Valley Basin, 2024, in *Microtremor Measurements Performed in Proximity to Strong Motion Stations Across California*. DesignSafe-CI. https://doi.org/10.17603/ds2-yxm7-2p26

Ornelas F, Stapleton J, Buckreis T, Nweke C, Stewart J 2025c. Microtremor Data Collection in West LA Region, 2024, in *Microtremor Measurements Performed in Proximity to Strong Motion Stations Across California*. *DesignSafe-CI*. https://doi.org/10.17603/ds2-kz21-za86

Ornelas F, Stapleton J, Buckreis T, Nweke C, Stewart J 2025d. Microtremor Data Collection in the Imperial Valley Basin, 2024, in *Microtremor Measurements Performed in Proximity to Strong Motion Stations Across California*. DesignSafe-CI. https://doi.org/10.17603/ds2-6wpb-0r31

Ornelas FJG, de la Torre CA, Zhang ZS, Buckreis TE, Brandenberg SJ, Stewart JP 2026. Preliminary findings on the application of mHVSR for data-driven site response prediction in California. In *Proceedings of GeoCongress 2026* (pp. xx–xx). American Society of Civil Engineers.

Pinilla-Ramos C, Abrahamson NA, Kayen RE 2022. Estimation of site terms in ground-motion models for California using horizontal-to-vertical spectral ratios from microtremor. *Bulletin of the Seismological Society of America,* 112, 3016–3036.

Sahakian V, Baltay A, Hanks TC, Buehler J, Vernon F, Kilb D, & Abrahamson NA 2018. Decomposing leftovers: Event, path, and site residuals for a small-magnitude Anza region GMPE. *Bulletin of the Seismological Society of America*, 108(5A), 2478–2492.

Site Effects Assessment using Ambient Excitations (SESAME). 2004. Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations—Measurements, processing and interpretation. *European Commission Project* No. EVG1-CT- 2000-00026, 62. Available at http://sesame.geopsy.org/Papers/HV_User_Guidelines.pdf (last accessed September 2012).

Senna S, Midorikawa S and Wakamatsu K. 2008. Estimation of spectral amplification of ground using H/V spectral ratio of microtremors and geomorphological land classification. *Journal of Japan Association for Earthquake Engineering* 8, 1–15.

Seyhan E, Stewart JP 2014. Semi-empirical nonlinear site amplification from NGA-West2 data and simulations. *Earthquake Spectra,* 30, 1241–1256.

Stewart JP, Afshari K, and Goulet CA (2017) Non-ergodic site response in seismic hazard analysis. *Earthquake Spectra* 33, 1385-1414.

Virtanen P, et al. 2020. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods,* 17, 261–272.

Wang P, Tsai YT, Stewart JP, Mikami A, Brandenberg SJ 2022a. Region-specific linear site amplification model for peaty organic soil sites in Hokkaido, Japan. *Earthquake Spectra,* 38, 2207-2234.

Wang P, Zimmaro P, Buckreis TE, Gospe T, Brandenberg SJ, Ahdi SK, Yong A, Stewart JP 2022b. Relational database for horizontal-to-vertical spectral ratios. *Seismological Research Letters,* 93, 1075-1088.

Wang P, Zimmaro P, Ahdi SK, Yong A, Stewart JP 2023. Identification protocols for horizontal-to-vertical spectral ratio peaks. *Bulletin of the Seismological Society of America*, 113, 782 – 803.

Yong A, Martin A, Stokoe K, Diehl J 2013. ARRA-funded V_{S30} measurements using multi-technique approach at strong-motion stations in California and central-eastern United States, *U.S. Geol. Surv*. Open-File Rept. 2013-1102, 60 pp. and data files.