

UC Berkeley

UC Berkeley Previously Published Works

Title

Evaluation of Site Terms for Ground-Motion Models for South America

Permalink

<https://escholarship.org/uc/item/59d6g4b4>

Journal

Bulletin of the Seismological Society of America, 114(5)

ISSN

0037-1106 1943-3573

Authors

Pinilla-Ramos, Camilo I

Abrahamson, Norman A

Kayen, Robert E

Publication Date

2024-07-03

DOI

10.1785/0120240020

Peer reviewed

Evaluation of Site Terms for Ground-Motion Models for South America

Camilo I. Pinilla-Ramos^{*1}, Norman A. Abrahamson¹, and Robert E. Kayen¹

ABSTRACT

The time-averaged shear-wave velocity in the top 30 m, V_{S30} , is used to represent the site condition in many ground-motion models (GMMs). Regionalized GMMs account for regional differences in the $\ln(V_{S30})$ scaling by including region-specific coefficients for the site term. For example, recent GMMs developed for subduction zone earthquakes as part of the Next Generation Attenuation-Subduction (NGA-Sub) project include region-specific V_{S30} scaling for seven regions: South America, Central America, Japan, New Zealand, Taiwan, Cascadia (Pacific Northwest America), and Alaska. Of these seven regions, South America has the largest within-event standard deviation. One cause for the larger within-event standard deviation is a weaker relation between V_{S30} and the site term in South America compared to other regions. A cause for this weak correlation is that most of the V_{S30} values for the South American region in the NGA-Sub data set are based on proxies. The relation between V_{S30} and the site term improves significantly for periods less than 1 s if only stations with measured V_{S30} are considered, but the correlation is weak for periods greater than 1 s even for stations with measured V_{S30} , indicating that, for the South America region, V_{S30} is not well correlated with the deeper shear-wave velocity profile that controls the long-period amplification. To improve the site terms for South America, we develop a model based on the horizontal-to-vertical spectral ratio (HVSr) from microtremors, similar to the approach used by Pinilla-Ramos *et al.* (2022) for California. The data set includes 660 recordings from 51 earthquakes recorded at 274 sites. The site-term model includes the period-dependent HVSr amplitude and the geometric mean of the average HVSr amplitude over the frequency band of 0.25 to 15 Hz. Including these two parameters reduces the standard deviation of site-specific site terms over the period range 0.3–3 s, with the largest reduction in the period range 1–2 s. This site-term model can be incorporated into subduction GMMs and implemented in probabilistic seismic hazard analyses for South America.

KEY POINTS

- What is the cause of the large within-event standard deviation of ground-motion models in South America?
- Horizontal-to-vertical spectral ratio (HVSr) is a better parameter than V_{S30} for predicting site amplification in South America.
- Use of HVSr for modeling site amplification may improve the accuracy of site terms in ground-motion models (GMMs) for other regions.

[Supplemental Material](#)

INTRODUCTION

The traditional approach to developing ground-motion models (GMMs) uses the ergodic assumption, which assumes that the median and standard deviation for the combined set of worldwide data applies to any site in the same tectonic category worldwide (Anderson and Brune, 1999). Because the ground-motion

data sets have grown, it has become clear that the ergodic assumption does not work well for specific regions due to significant systematic differences in the ground-motion scaling for different regions. Some GMMs address the regional differences using data limited to a single region (e.g., Bragato and Slejko, 2005; Danciu and Tselentis, 2007; Akkar and Çağnan, 2010; Bindi *et al.*, 2011). An alternative approach includes region-specific terms within a global GMM (e.g., Abrahamson *et al.*, 2014). Both approaches account for the systematic regional differences in the scaling of the median ground motion for different regions.

One of the important regional differences is the scaling for the site term. V_{S30} is a widely used site parameter; however, it is

1. Department of Civil Engineering, University of California, Berkeley, California, U.S.A., <https://orcid.org/0000-0002-4034-6618> (CIP-R); <https://orcid.org/0000-0001-7900-6023> (NAA); <https://orcid.org/0000-0002-0356-072X> (REK)

*Corresponding author: camilo.pinilla@berkeley.edu

Cite this article as Pinilla-Ramos, C. I., N. A. Abrahamson, and R. E. Kayen (2024). Evaluation of Site Terms for Ground-Motion Models for South America, *Bull. Seismol. Soc. Am.* **XX**, 1–13, doi: [10.1785/0120240020](https://doi.org/10.1785/0120240020)

© Seismological Society of America

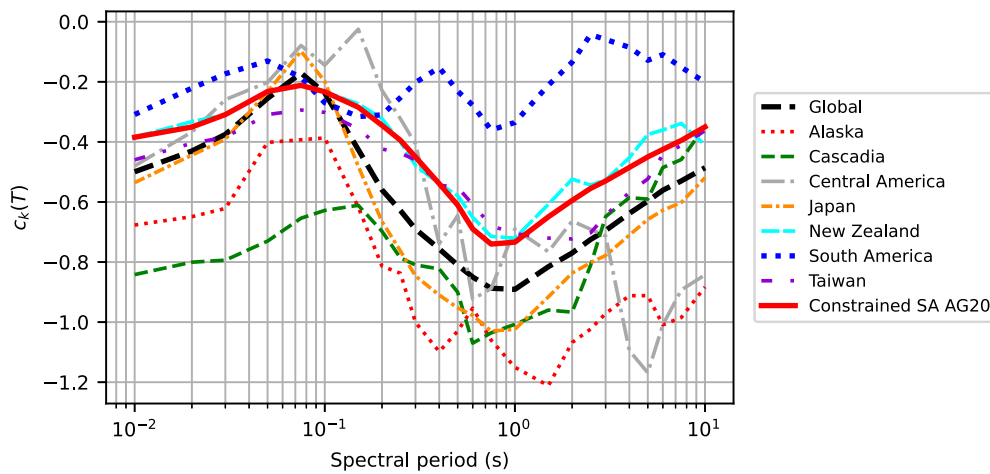


Figure 1. Scaling coefficient, $c_k(T)$, for V_{S30} (time-averaged shear-wave velocity in the top 30 m depth) for the seven subduction regions on the Next Generation Attenuation-Subduction (NGA-Sub) data set from the AG22 ground-motion model (GMM). The red solid line corresponds to the constrained South America V_{S30} coefficient adopted in the AG22 GMM. The blue dashed line is the scaling between site terms and $\ln(V_{S30})$ considering only sites with measured V_{S30} in the NGA-Sub. The color version of this figure is available only in the electronic edition.

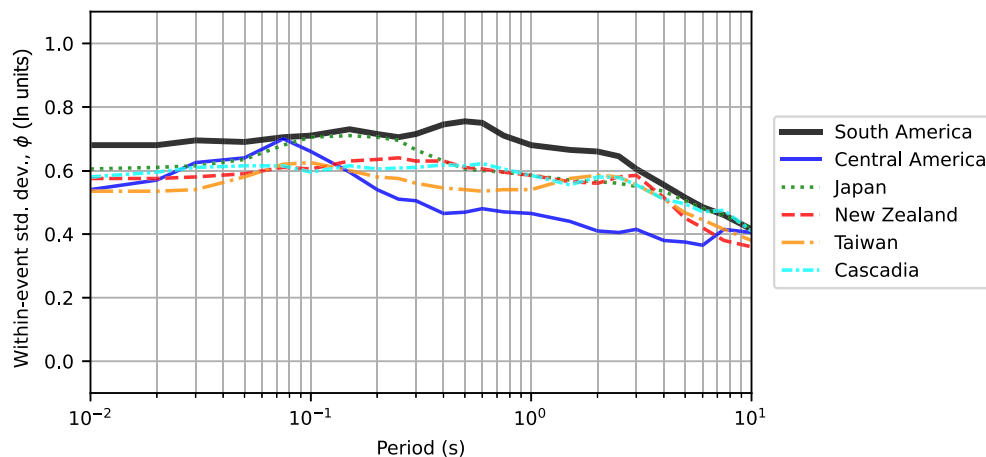


Figure 2. Within-event standard deviation observed for different subductive tectonic environments around the world. The color version of this figure is available only in the electronic edition.

not a fundamental physical parameter that determines the site amplification. Instead, for some regions, the V_{S30} values for natural deposits are correlated with the deep shear-wave velocity profile, $V_S(Z)$, that controls the site amplification (Boore *et al.*, 2011). The correlation between the V_{S30} and the $V_S(Z)$ profile varies for different regions (Kamai *et al.*, 2016). The effect of the regional dependence of the correlation is captured in GMMs by including a region-specific coefficient for the $\ln(V_{S30})$ scaling. The linear site term for region k is typically modeled by a linear function of the logarithm of V_{S30} :

$$f_{\text{site},k}(T) = c_k(T) \ln(V_{S30}). \quad (1)$$

The Next Generation of Ground-Motion Attenuation for Subduction Zones Project (NGA-Sub) (Bozorgnia and Stewart, 2020) developed a set of regionalized GMMs that include the region-specific $c_k(T)$ terms for seven regions: South America, Central America, Japan, New Zealand, Taiwan, Cascadia (Pacific Northwest America), and Alaska. As an example, the raw (before smoothing over periods) $c_k(T)$ coefficients for the seven regions from the Abrahamson and Gulerce (2022) GMM (AG22) are compared in Figure 1. In most regions, there is a characteristic shape of the period dependence of the $c_k(T)$ term with values near -0.45 for PGA, decreasing to about -0.9 near a period of 1 s, and increasing to about -0.5 at a period of 10 s. In contrast, the raw $c_k(T)$ for the South American data show much weaker scaling with V_{S30} and do not show the characteristic period-dependent shape seen for other regions.

For the AG22 GMM, the within-event standard deviation, ϕ , was calculated for six out of the seven regions as shown in Figure 2. The Alaska region was omitted due to metadata errors in the 2019 version of the flat file used in the development of the AG22

GMM, which led to unreliable residuals. Of the six regions analyzed, South America exhibits the largest standard deviation of the within-event residuals for all spectral period ranges considered. As we show later, a key cause of the large within-event variability for the South American region is the weak correlations between the V_{S30} values in the NGA-Sub data set and the site terms.

The weaker correlation between V_{S30} and the site term for the South America data could be due to either poor estimates of the V_{S30} values due to the use of proxies, such as topographic slope, or to a weaker correlation between V_{S30} and the $V_S(Z)$ profile that controls the site amplification for the geologic

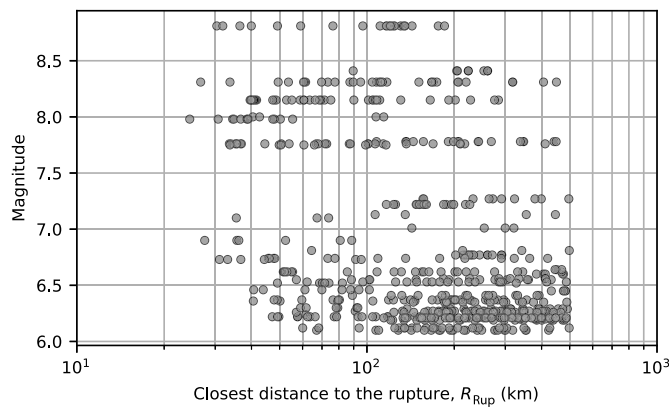


Figure 3. Distribution between the closest distance to the rupture, R_{Rup} and magnitude for the South American subset.

environment in South America. In this article, we evaluate these two potential causes of the weak correlation of V_{S30} with the site terms. As an alternative to V_{S30} , a site term based on the horizontal-to-vertical spectral ratio (HVSr) from microtremors is proposed for use with the AG22 GMM for South America.

DATA SET

For our analysis, we adopted the subset of the NGA-Sub data set for South America used in the development of the AG22 GMM. The magnitude and distance distribution for this subset is shown in Figure 3. The AG22 subset for South America includes 660 records from 51 earthquakes, with 83% of the recordings in Chile, 12% in Colombia, 3% in Peru, and 2% in Ecuador. The data set includes 274 stations. The spatial distribution of the stations in the South America region is shown in Figure 4. The stations are concentrated in Chile (75%), followed by Colombia (18%), Peru (5%), and Ecuador (2%). Whereas the Chilean territory has the densest seismometer network, the spatial sampling is sparser for other countries in the region.

V_{S30} values

The NGA-Sub for South America has measured V_{S30} values for 26% of the stations (72 of the 274 stations), and 21% of the recordings. Of the 72 stations with measured V_{S30} , 31 of the V_{S30} come from *Kayen et al. (2014)*, who performed surface analysis of surface waves (SASW) studies for Chilean seismic stations that recorded the 2010 M 8.8 Maule earthquake. *Leyton et al. (2018)* measured V_{S30} at 142 stations of the Chilean Seismic network. Of these 142 sites, only 37 were in the selected NGA-Sub data set, and only 11 were sites with unmeasured V_{S30} in the NGA-Sub data set. The V_{S30} values for these 11 sites were added to the data set of our analysis for a total of 83 of 274 sites with measured V_{S30} (30%). For the 191 stations without measured V_{S30} , a topographic-slope-gradient proxy was used to estimate the V_{S30} (*Contreras et al., 2018*).

The distribution of V_{S30} in our data set is shown in Figure 5. The distribution shown in Figure 5b is skewed toward large values with a geometric mean of 526 m/s and a standard deviation of 0.39 ln units. The geometric mean V_{S30} for sites with measured V_{S30} is 520 m/s. The rock sites ($V_{S30} > 700$ m/s) are all located in Chile (Fig. 5c). The histogram in Figure 5d uses a bin size of 1 m/s. The sharp peaks with the same V_{S30} values indicate the V_{S30} values estimated by the topographic-slope-gradient proxy. These proxy estimations can lead to poor estimates of V_{S30} . For example, the V_{S30} values for stations PB13 and PB16 were estimated by proxies, with a reported V_{S30} of 605 m/s for both sites. These sites are located on the summit of a mountain (~4760 and 4475 m above sea level) upon a human-made plateau, over Oligocene–Miocene volcanosedimentary sequences constituted by basaltic and dacitic lava flows (*SERNAGEOMIN, 2003*). Human-made plateaus are commonly observed in facilities situated on top of mountains, where a section of the summit is excavated through blasting to create space for the project. The site conditions of these areas are unlikely to have a V_{S30} as low as 605 m/s, indicating that the topographic-slope proxy-based V_{S30} is not working well for these sites.

HVSr from microtremors

The HVSr for the Fourier amplitude spectra from microtremors was estimated for a subset of the sites. We collected triaxial microtremors signals using the mass-downloading module from the open-source ObsPy package written in Python (*Beyreuther et al., 2010*) for the 78 of the 274 stations in our data set that have continuous recordings. For each station, we retrieved 1 hr samples of microtremors recorded during the night to reduce the contamination due to anthropogenic activity. This data set was augmented by including microtremor recordings from the pre-event measurements at six sites that do not have continuous recordings, but which have adequate pre-event memory to record the microtremors before the arrival of the P waves at the site. The total HVSr data set has 84 samples, all of which are from stations located in Chile.

Using the HVSr data, three site metrics are computed: the period-dependent HVSr amplitude, $HVSr(T)$, the normalized spectrum, $HVSr^*(T)$, proposed by *Pinilla-Ramos et al. (2022)*, and the normalization value, $\overline{HVSr}$. The normalized $HVSr^*(T)$ is defined as

$$HVSr^*(T) = \frac{HVSr(T)}{\overline{HVSr}}. \quad (2)$$

The normalization value, $\overline{HVSr}$, is defined as the geometric mean of the HVSr over the Fourier frequency range 0.25–15 Hz:

$$\overline{HVSr} = \left(\prod_{i=1}^{N_f} HVSr(f_i) \right)^{1/N_f}, \quad (3)$$

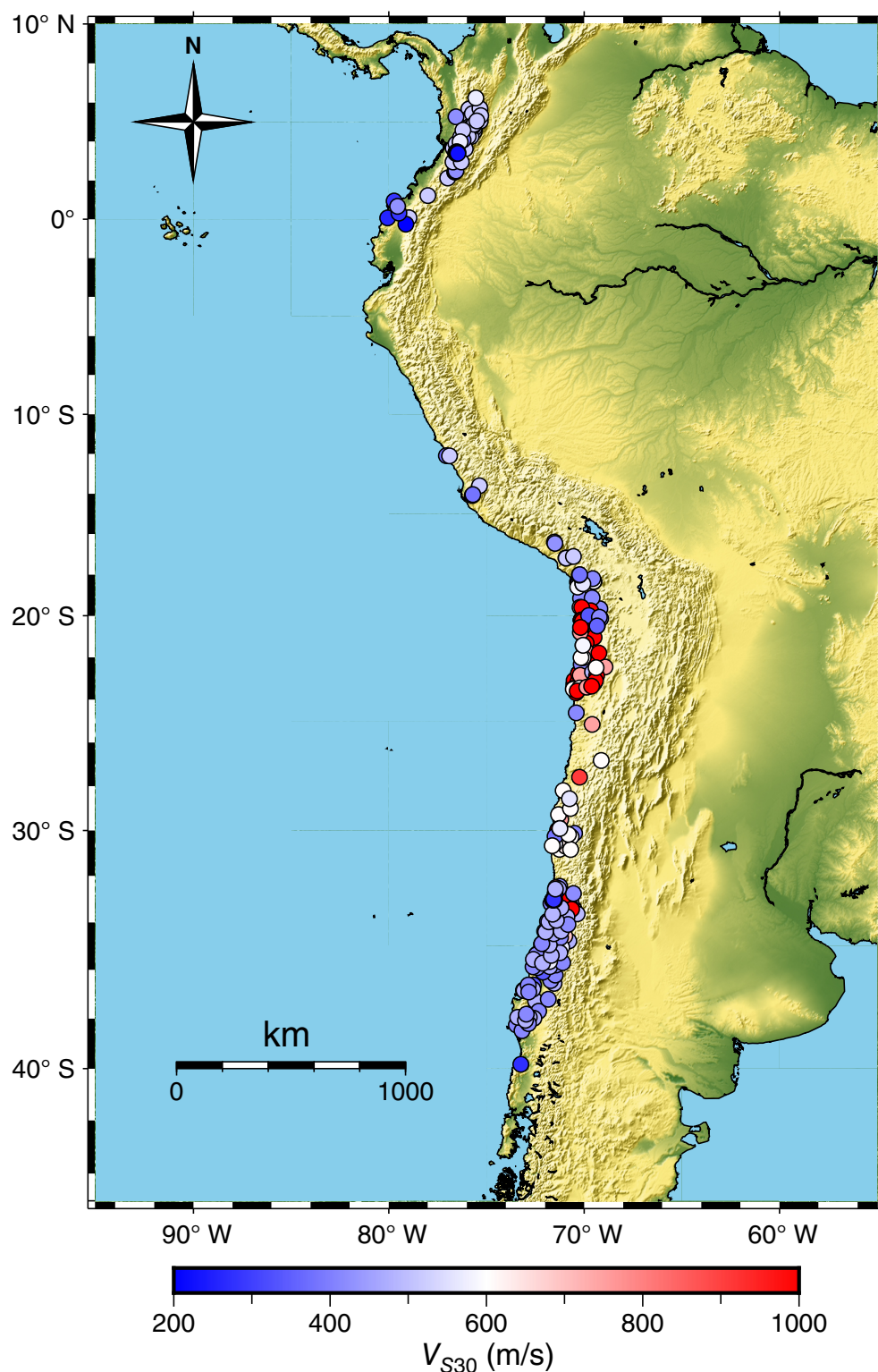


Figure 4. Map showing the available stations in the NGA-Sub data set for South America. The color map presents the V_{S30} . The color version of this figure is available only in the electronic edition.

in which N_f is the number of frequencies between 0.25 and 15 Hz with an equal spacing in log frequency. This frequency band was used by Pinilla-Ramos *et al.* (2022) based on their

exception of the HVSR amplitude at 0.2 s and $\overline{\text{HVSR}}$, the HVSR-based site parameters have a weak correlation with $\ln(V_{S30})$ (correlation less than 0.21). The correlation between

evaluation of site terms from California ground-motion data. Given the small number of recordings per site in the South America data set, we kept the frequency band fixed to the range developed for the larger California data set.

f_0 from microtremors

In addition to the HVSR amplitude, we identified the predominant frequency in the HVSR spectrum, f_0 , if a clear peak was present. The f_0 was estimated following the method proposed by SESAME (2004). Of the 84 stations with available HVSR from microtremor, 46 stations have a clear peak for determining f_0 . The f_0 distribution of this subset is displayed in Figure 6. The mean f_0 is 4.2 Hz, and there are only four sites with f_0 below 1 Hz. This indicates that most of the sites with clear f_0 are stiff sites.

Correlation of parameters

We analyzed the correlation structure between $\ln(V_{S30})$, $\text{HVSR}(T)$, $\overline{\text{HVSR}}$, and the predominant frequency, f_0 . The correlogram is shown in Figure 7. For periods above 0.75 s, there is a moderate positive correlation between the $\text{HVSR}(T)$ at different periods. The interperiod correlations are larger at long periods because the HVSR spectra tend to be flat above 0.75 s. The normalization term, $\overline{\text{HVSR}}$, has a moderate positive correlation with $\text{HVSR}(T)$ and a negative correlation with $\ln(V_{S30})$, implying that softer sites tend to have larger mean values of HVSR amplitude. With the

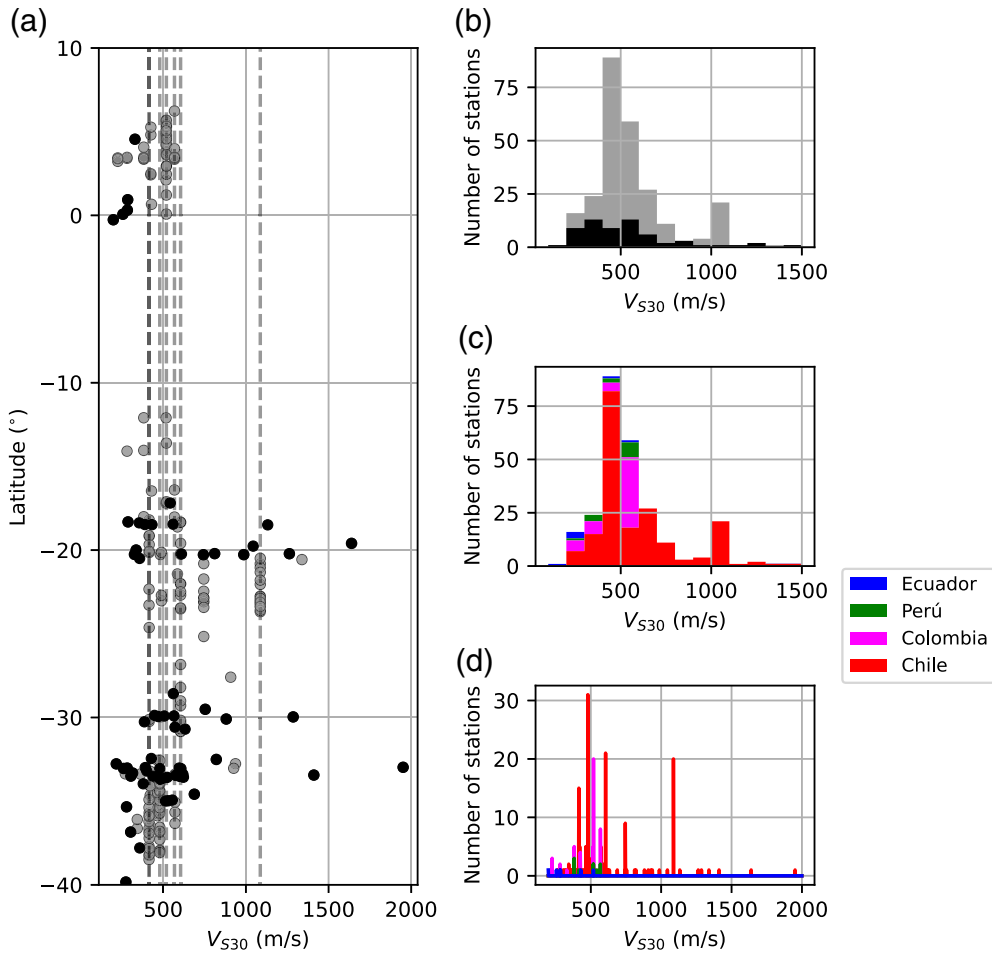


Figure 5. Summary of the V_{S30} distribution of the NGA-Sub data for South America. (a) V_{S30} distribution as a function of latitude. The black dots are sites with measured V_{S30} . (b) Histograms of the V_{S30} for all the data (gray bars) and for the subset of sites with measured V_{S30} (black bars). (c) Histogram of the V_{S30} by country. (d) Histogram of V_{S30} values using a narrow bin size of 1 m/s. The spikes show that there are several V_{S30} repeated values due to using proxies to estimate the V_{S30} . The color version of this figure is available only in the electronic edition.

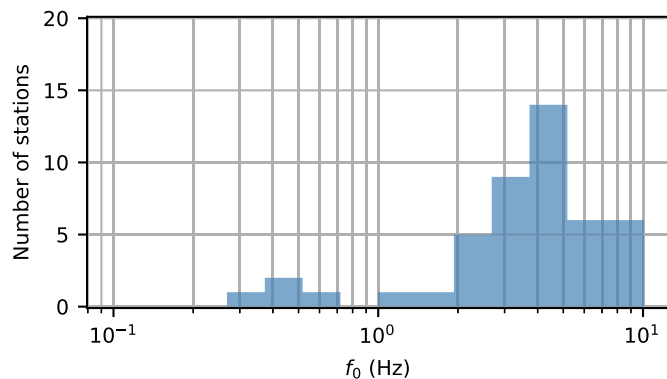


Figure 6. Distribution of f_0 in our data set. The color version of this figure is available only in the electronic edition.

f_0 and $\ln(V_{S30})$ for the 84 South America sites is also low ($r = 0.14$). For comparison, using the NGA-West2 data set for California (Bozorgnia *et al.*, 2014), the correlation between f_0 and $\ln(V_{S30})$ is much larger ($r = 0.52$) (Pinilla-Ramos *et al.*, 2022). The low correlation for the South America data is likely due to the use of proxy-based $\ln(V_{S30})$ values for most of the sites in our data set with HVSR data: only ten of the 84 sites with HVSR data have measured V_{S30} , and of those ten sites, only three sites have measured V_{S30} and a clear f_0 .

SITE TERMS

A partially nonergodic GMM can be written with the following statistical form:

$$\ln(Y_{es}) = \bar{Y}_{AG22}(M, R_{Rup}, V_{S30}, \dots) + \delta B_e + \delta S2S_s + \delta WS_{es}, \quad (4)$$

in which Y_{es} is the subduction ground-motion observation from event “e” at site “s”, $\bar{Y}_{AG22}$ is the median ground-motion from the AG22 GMM; δB_e is the between-event residual; $\delta S2S_s$ is the between-site residual; and δWS_{es} is the within-site residual

(Al-Atik *et al.*, 2010). The δB_e , $\delta S2S_s$, and δWS_{es} are assumed to be normally distributed random variables with means of zero and standard deviations of τ , ϕ_{S2S} , and ϕ_{SS} , respectively.

In equation (4), the $\delta S2S_s$ term is the average site term relative to the V_{S30} scaling in the GMM. As an alternative, we define a site term, $\delta S2S'_s$, that is relative to the AG22 GMM without the V_{S30} scaling as

$$\ln(Y_{es}) = [\bar{Y}_{AG22}(M, R_{Rup}, V_{S30}, \dots) - f_{\text{site,AG22}}(V_{S30}, \text{PSA}_{1100})] + a_1 + \delta S2S'_s + \delta B_e + \delta WS_{es}, \quad (5)$$

in which $f_{\text{site,AG22}}(V_{S30}, \text{PSA}_{1100})$ is the V_{S30} scaling from the AG22 GMM; and a_1 is a coefficient that centers the model on the average site condition in the data set.

Most of the stations have only one recording, as shown in Figure 8. Only 52 of the 274 stations have more than three records per site. Thus, the common criterion of requiring at least three records per site used for analyses of site terms was relaxed,

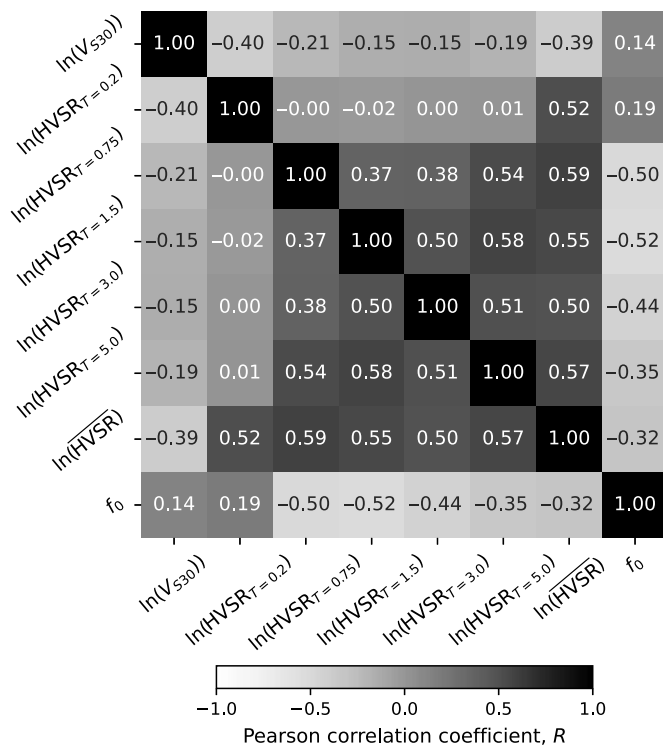


Figure 7. Correlogram between the site amplification predictors. The standard error of the correlation coefficient ranges between 0.7 and 0.11, except for sites with f_0 . As the sites with f_0 are a subset of our data set, the standard error for these correlation coefficients ranges between 0.11 and 0.14.

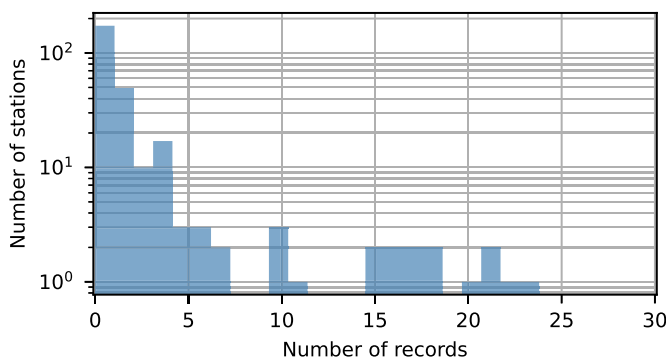


Figure 8. Histogram of the number of records per station at 0.2 s spectral period. The color version of this figure is available only in the electronic edition.

and all the stations in the data set were used. $\delta S2S'_s$ were estimated using a random-effect regression (Abrahamson and Youngs, 1992). The $\phi_{S2S'_s}$ and ϕ_{SS} from regression are shown in Figure 9. For periods less than 1 s, $\phi_{S2S'_s}$ is larger than ϕ_{SS} , indicating that most of the within-event variability for periods less than 1 s is due to systematic site amplification effects. At longer periods, ϕ_{SS} is larger than $\phi_{S2S'_s}$, indicating the within-event standard deviation at long periods is dominated by path and finite-fault effects and not an average site effect.

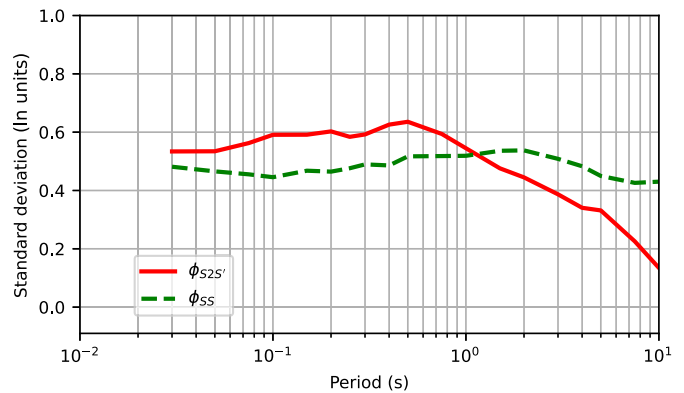


Figure 9. Decomposition of the within-event standard deviation for South America. The color version of this figure is available only in the electronic edition.

There is large epistemic uncertainty in the estimate of the $\delta S2S$ terms due to the small number of recordings per site. Based on data from sites with large numbers of recordings per site, the aleatory variability of the site amplification about the median amplification is in the range 0.26–0.32 natural log units (Stewart *et al.*, 2017). This range provides an estimate of the epistemic uncertainty in the $\delta S2S$ values if we have only one recording per site. This epistemic uncertainty is similar to the ϕ_{S2S} estimated for the South America data set, indicating that for only one recording per station, we cannot tell if the $\delta S2S$ at a particular site is significantly different from zero.

Our evaluation of the scaling of the $\delta S2S$ with V_{S30} or with HVSR takes advantage of the large number of sites. With the available data, we have a weak constraint on the $\delta S2S$ for a particular site, but by considering the scaling for the set of 274 sites, we have better constraints on the evaluation of the proxy-based V_{S30} scaling and the HVSR-based scaling. We have at least 25 sites for different V_{S30} ranges (Figure 5b). With only one recording per site, the standard error of the mean of the $\delta S2S$ for a V_{S30} range would be about $\frac{0.3}{\sqrt{25}} = 0.06$ ln units. This is small enough to have confidence in evaluating proxy-based V_{S30} scaling and the HVSR-based scaling for sites in South America.

Analysis of site parameters and site terms

The scaling between $\ln(V_{S30})$ and the site terms, $\delta S2S'_s$, for four spectral periods is shown in Figure 10. The solid circles indicate sites with measured V_{S30} , and the open circles indicate sites for which the V_{S30} was estimated by the topographic-slope proxy. The stations with V_{S30} estimated by proxy have a larger scatter than sites with measured V_{S30} .

The determination coefficient, R^2 , between $\ln(V_{S30})$ and $\delta S2S'_s$ in the period range 0.03–10 s is shown in Figure 11. For the full data set, the R^2 is below 0.2. We evaluated the scaling of $\ln(V_{S30})$ with the site terms of stations with measured V_{S30} (Fig. 11). The R^2 increases significantly for the subset with

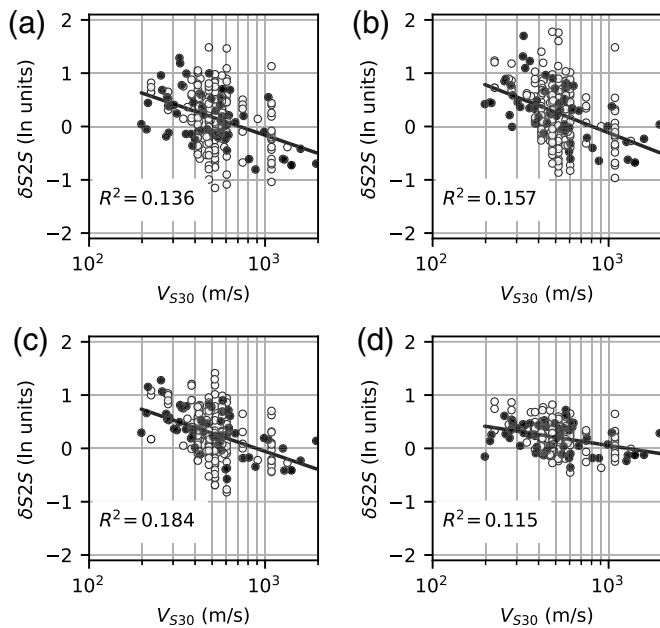


Figure 10. Scatter plots between the V_{S30} and the site terms, $\delta S2S'$, for four different spectral periods: (a) 0.2 s; (b) 0.5 s; (c) 1.0 s; and (d) 3.0 s. The black circles are stations with measured V_{S30} , and the white circles are stations with V_{S30} estimated via proxy.

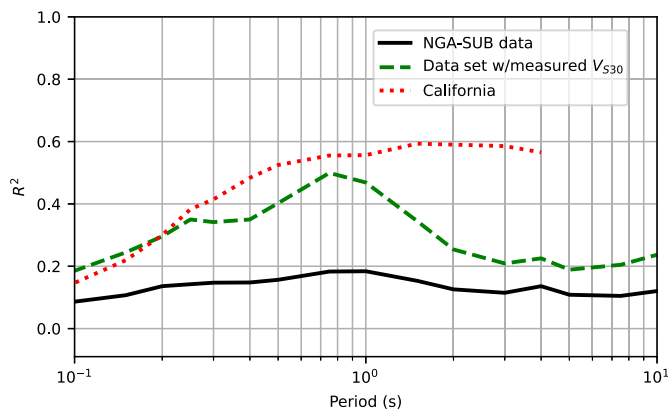


Figure 11. Period dependence of R^2 for the relation between $\ln(V_{S30})$ and the site terms, $\delta S2S'$ for the NGA-Sub data set and the subset of the NGA-Sub data set with measured V_{S30} values. The R^2 for California crustal events is shown for comparison. The color version of this figure is available only in the electronic edition.

measured V_{S30} values with a peak at 0.75 s. At periods greater than 1 s, the R^2 decays rapidly. For comparison, the R^2 for crustal events in California is also shown in Figure 11. At periods less than 0.75 s, the R^2 for the California data is similar to the R^2 for the South America data with measured V_{S30} , but at long periods, the R^2 for the California data does not show the strong decay seen in the South America data. This may be due to differences in the typical site conditions in the two regions:

the South America sites tend to be stiff sites, whereas the California sites include more soil sites with stronger long-period amplification from the shallow soil. For the South America data, the V_{S30} captures short-wavelength amplification as in California but correlates poorly with deeper geologic structures that controls the long-period amplification at stiff sites.

Because of the poor performance of proxy-based V_{S30} in explaining the variance of site terms, we used an alternative methodology to improve the predictive capabilities of site amplification based on the HVSr from microtremors. The period dependence of R^2 for each candidate site parameter is shown in Figure 12a for the 84 stations with HVSr data. We also computed the R^2 for a subset of 46 stations with clear f_0 (Fig. 12b). The R^2 for the V_{S30} scaling for these two subsets is smaller than the R^2 for the full data set shown in Figure 11. The lower R^2 values for the $\ln(V_{S30})$ scaling for sites with HVSr data are likely due to the smaller fraction of sites with HVSr data that have measured V_{S30} : only 12% of the sites with HVSr measurements have measured V_{S30} values compared to 30% for the full data set.

For these two subsets, all four of the HVSr-based site parameters perform better than V_{S30} for spectral periods above 0.2 s. For the South America data, the unnormalized HVSr has a larger R^2 than the normalized HVSr, which is in contrast to the results for California data that showed a larger R^2 for the normalized HVSr parameter (Pinilla-Ramos *et al.*, 2022).

The R^2 for the normalization constant is smaller for the subset of sites with a clear f_0 than for all sites with measured HVSr, indicating that the amplitude of the HVSr is more useful for estimating the site amplification if there is not a strong resonance. If there is a strong resonance, then the f_0 value without the amplitude information is a better predictor of the site amplification than the unnormalized HVSr for periods near the resonance period.

Modeling site terms using HVSr from microtremors

In this section, we evaluate the modeling of the site terms using the four microtremor HVSr-based site parameters: $\text{HVSr}(T)$, $\text{HVSr}^*(T)$, $\overline{\text{HVSr}}$, and f_0 . Several alternative combinations of the four HVSr parameters are considered. Because not all the sites have a clear f_0 , we compare the site scaling using the $\delta S2S'$ from two data sets: the 84 stations with HVSr measurements and the 46 stations with clear f_0 .

The general functional form assumed for the site term using all the HVSr parameters is given by

$$f_{\text{site}}(T) = c_0 + c_1 \ln(\text{HVSr}(T)) + c_2 \ln(\overline{\text{HVSr}}) + c_3 \ln(\text{HVSr}^*(T)) + c_4 f_0. \quad (6)$$

Previous studies have used $\ln(f_0)$ in the development of site terms based on f_0 ; however, for the South America data set, a linear f_0 term provided a better fit (larger R^2) than using $\ln(f_0)$.

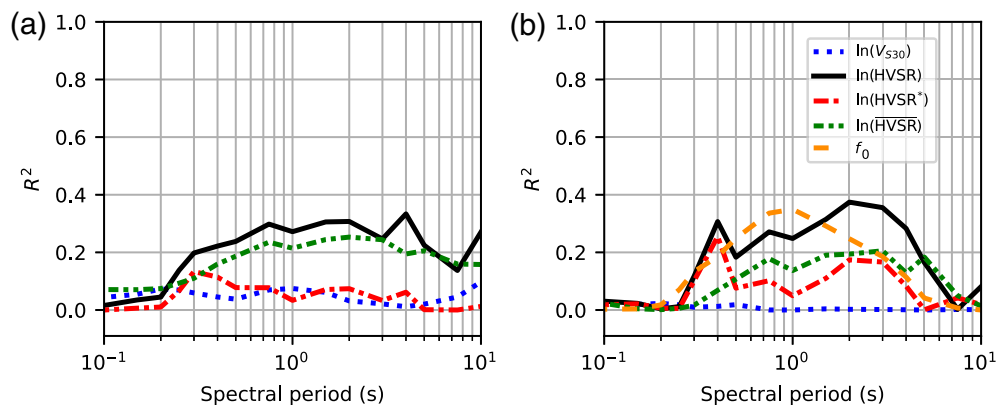


Figure 12. Period dependence of R^2 for different site parameters. (a) Based on the 84 stations with HVSR spectra. (b) Based on the 46 sites with a clear f_0 . Sites with measured V_{S30} and proxy-based V_{S30} values are included in these two subsets. The color version of this figure is available only in the electronic edition.

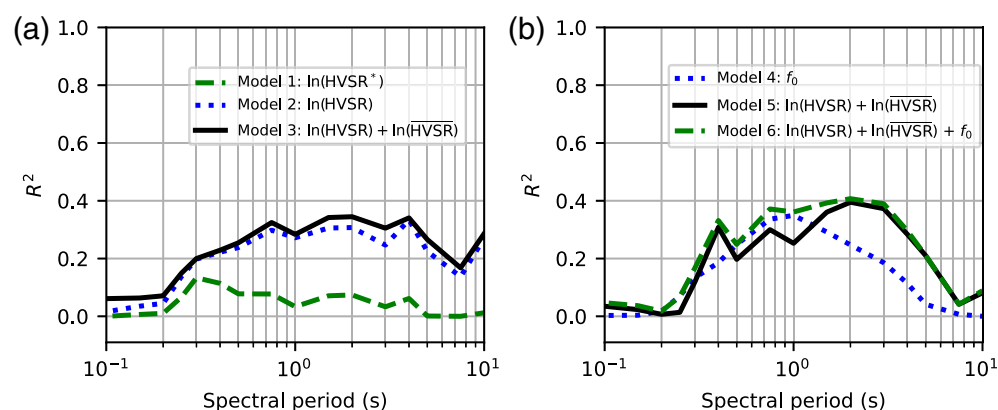


Figure 13. Comparison of several site-term models based on parameters from HVSR. (a) Models developed for the data set with available HVSR measurements. (b) Subset of sites with f_0 . The models are described in the text. The color version of this figure is available only in the electronic edition.

The alternative models considered are listed in Table 1. If a parameter is not used in a model, the coefficient for that parameter in equation (6) is set to zero.

Starting with the data set of all the stations with HVSR measurements, we evaluated the models using the normalized $\text{HVSR}^*(T)$ compared to using the unnormalized $\text{HVSR}(T)$.

Model 2, which is based on the unnormalized $\text{HVSR}(T)$, has a larger R^2 than model 1, which is based on the normalized $\text{HVSR}(T)$, as shown in Figure 13a. Including both the unnormalized $\text{HVSR}(T)$ and the normalization constant, $\overline{\text{HVSR}}$, in model 3, leads to a slightly larger R^2 in the 1–4 s period range. Using the unnormalized $\text{HVSR}(T)$ and the normalization constant, $\overline{\text{HVSR}}$, as independent parameters is different from using the normalized $\text{HVSR}^*(T)$ because the two terms can have different coefficients: c_1 and c_2 in equation (6).

To evaluate the predictive power of f_0 compared to the other site parameters, we use the subset of data with f_0 for models 4, 5, and 6. Model 4 includes only f_0 as a site parameter. Model 5 is the same as model 3 but applied to the subset of data with clear f_0 values. Model 6 includes three site parameters: $\text{HVSR}(T)$, $\overline{\text{HVSR}}$ and f_0 . The R^2 values for these models are compared in Figure 13b. For model 4, the R^2 reaches a peak near 1 s, and its performance is similar to model 5 for periods below 0.75 s. The performance of

model 6 at periods less than 0.75 s and above 1.5 s is similar to the $\text{HVSR}(T)$ -based model without f_0 (model 5), indicating that the inclusion of f_0 does not significantly improve the model performance. The main improvement from including f_0 in model 6 is near a period of 1 s, which is the period range with the strongest f_0 scaling (peak in R^2 for model 4).

TABLE 1

Set of Site Amplification Predictors Extracted from Horizontal-to-Vertical Spectral Ratio (HVSR) Used in Each Site-Term Model

Period (s)	Full HVSR Data Set			Stations with f_0		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$\text{HVSR}(T)$	—	X	X	—	X	X
$\text{HVSR}^*(T)$	X	—	—	—	—	—
$\overline{\text{HVSR}}$	—	—	X	—	X	X
f_0	—	—	—	X	—	X

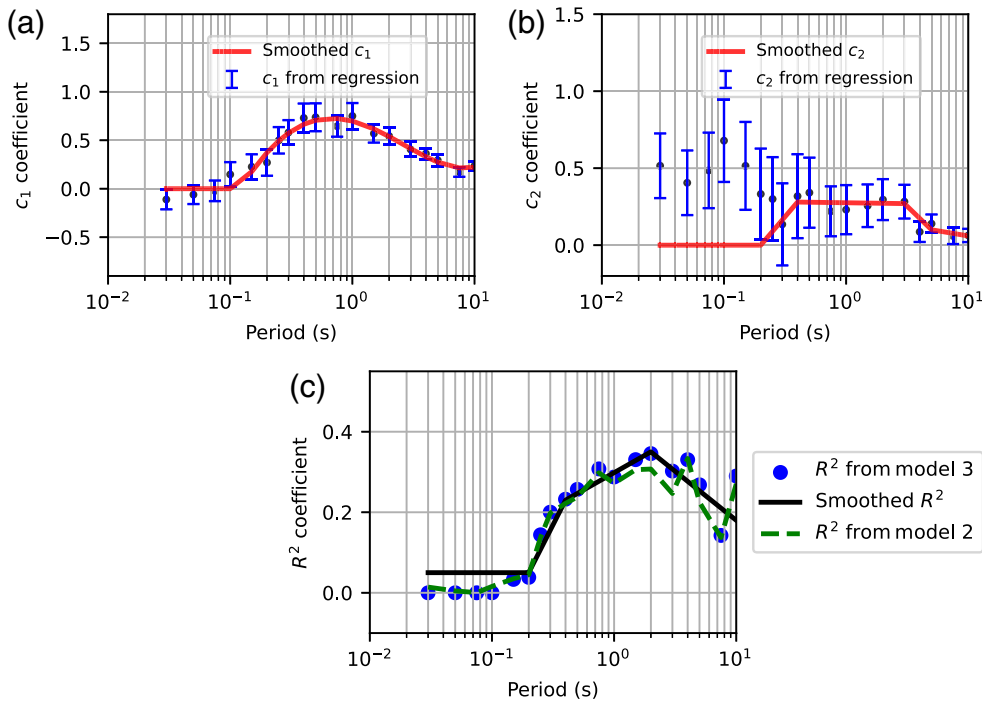


Figure 14. Coefficients obtained from the regression (black dots with error bars) and smooth coefficients (the line). (a) Coefficients for HVSR(T). (b) Coefficients for HVSR. The error bars represent the 16th to 84th percentile. (c) Determination coefficient, R^2 . The black solid and green dashed curves showed the improvement in the scaling by including HVSR in the model, compared to just adopting HVSR(T) as the only model parameter. The color version of this figure is available only in the electronic edition.

Models 3 and 5 use the same functional form but are based on different data sets: model 3 is based on the whole data set, whereas model 5 is based on the subset of sites with a clear f_0 . This functional form explains more of the long-period ($T > 4$ s) site terms for the full data set (model 3) than for the data set with clear f_0 (model 5). In the period range near 2 s, the subset of data from sites with a clear f_0 has a higher R^2 . The additional site amplification term based on f_0 (model 6) that applies only to sites with a clear f_0 could be included in the site term model. At this time, we limited our model to HVSR so that it can be applied to all sites.

A limitation of an f_0 -based site term is that the f_0 is not defined for all sites: some sites do not have a clear f_0 . In contrast, the HVSR amplitude is a continuous parameter that can be computed for any site with HVSR measurements. Therefore, we developed a site-term model based only on the HVSR amplitude and $\overline{\text{HVSR}}$, using the model 3 functional form:

$$\delta S2S'(T) = c_0 + c_1 \ln(\text{HVSR}(T)) + c_2 \ln(\overline{\text{HVSR}}). \quad (7)$$

The model coefficients c_1 and c_2 are partially correlated. To address this correlation, we applied a smoothing function to the period dependence of the c_1 coefficient. We then performed a second regression to estimate the c_2 coefficient with the smoothed c_1 held fixed. The c_2 was smoothed and the intercept

coefficient, c_0 , was estimated to center the model on the data after smoothing the other two coefficients. The coefficients obtained from the regression, their standard errors, and the smoothed R^2 are shown in Figure 14. The smoothed coefficients are listed in Table 2, and the model residuals are shown in Figure 15 for four spectral periods.

The performance of the HVSR-based site-term model is evaluated by comparing the standard deviation of the site terms with and without the HVSR-based site-term model, denoted $\phi_{S2S'}$ and ϕ_{S2S} , respectively. The results are shown in Figure 16. The reduction in ϕ_{S2S} reaches a maximum of 28% at a period of 2 s. The ϕ_{SS} is the same with and without the HVSR site term because it is computed after removing the sum of the site-term model and the site-specific $\delta S2S$ terms.

DISCUSSION

Model implementation in seismic hazard analysis

If site-specific HVSR(T) measurements are available, the proposed model can be incorporated in the AG22 GMM by first removing the V_{S30} scaling from the AG22 ergodic median and adding the site term based on the HVSR(T). The median for the modified GMM is given by

$$\begin{aligned} \overline{Y} = & [\overline{Y}_{\text{AG22}}(M, R_{\text{RUP}}, V_{S30}, \dots) - f_{\text{site}}(V_{S30}, \text{PSA}_{1100})] \\ & + (c_0 + c_1 \text{HVSR}(T) + c_2 \overline{\text{HVSR}}). \end{aligned} \quad (8)$$

With site-specific HVSR(T) data, the within-event aleatory variability for the modified AG22 GMM is given by

$$\phi_{\text{HVSR}} = \sqrt{\phi_{S2S}^2 + \phi_{SS}^2}. \quad (9)$$

The ϕ_{SS} and ϕ_{S2S} are listed in Table 2.

If a partially nonergodic approach is used with a nonergodic site term, the total standard deviation of the GMM, denoted σ_{SS} , is given by

$$\sigma_{SS} = \sqrt{\phi_{SS}^2 + \tau^2}. \quad (10)$$

The reduction of the aleatory variability term of the AG22 GMM for the partially nonergodic GMM is shown in Figure 17.

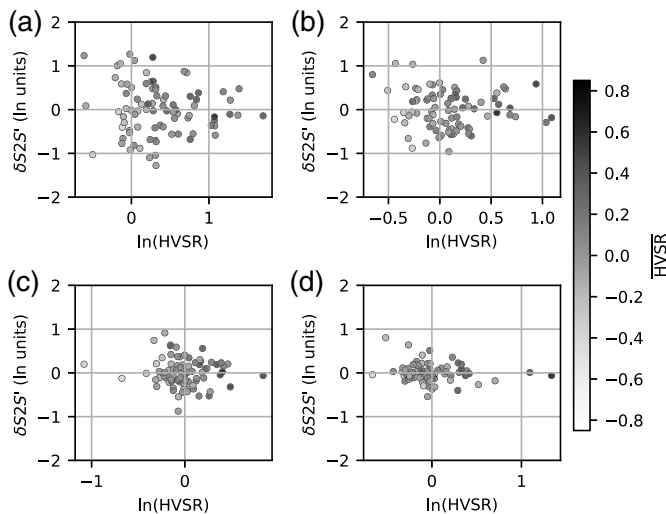


Figure 15. Residuals after fitting model 3 for four spectral periods: (a) at 0.2 s; (b) at 0.5 s; (c) at 1.0 s; and (d) at 3.0 s.

The σ in Figure 17 is the ergodic total standard deviation of the AG22 GMM, which is significantly larger than σ_{SS} at short periods, showing the value of having site-specific ground-motion data from either observed ground motions or from analytical modeling of the site amplification in addition to the $HVSR(T)$ for the site.

Comparison of performance of HVSR-based site terms in other regions of the world

The R^2 for HVSR-based and V_{S30} -based site terms for South America, California, and Taiwan are compared in Figure 18a.

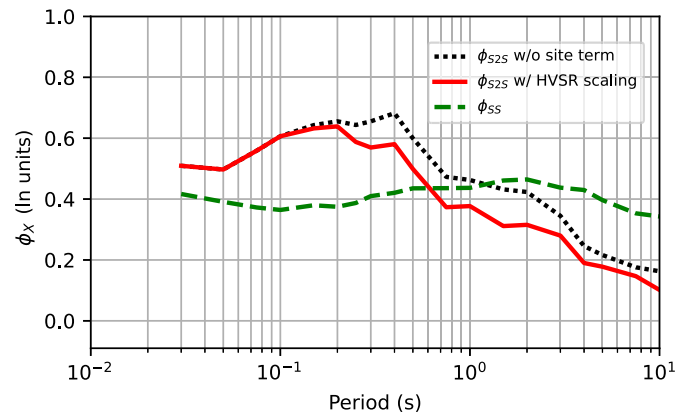


Figure 16. Comparison of the within-event standard components before and after fitting the site-term model based on $HVSR(T)$. The color version of this figure is available only in the electronic edition.

The R^2 values for the HVSR-based site terms for South America are similar to R^2 values for California for periods less than 1 s, but the R^2 values for South America decrease at periods above 1 s, unlike the R^2 values for California. The V_{S30} -based R^2 values for the three regions are also similar for periods up to 1 s, but for periods greater than 1 s, there is a sharp reduction in the R^2 for South America. Although the R^2 for the V_{S30} -based site terms in Taiwan is similar to California, the R^2 for the HVSR*-based site terms for Taiwan is much smaller than for California.

In California, the normalized HVSR* performs better than the unnormalized HVSR, as shown in Figure 19c; however, in South America, the unnormalized HVSR performs significantly better.

TABLE 2
Smoothed Coefficients for the Site-Term Model

Period (s)	c_0	c_1	c_2	$\phi_{S2S'}$	ϕ_{S2S}	ϕ_{SS}	τ
0.030	0.051	0.000	0.000	0.509	0.496	0.417	0.470
0.050	0.003	0.000	0.000	0.497	0.485	0.391	0.470
0.075	-0.028	0.000	0.000	0.559	0.545	0.372	0.470
0.100	-0.027	0.000	0.000	0.606	0.591	0.364	0.470
0.150	-0.045	0.176	0.000	0.643	0.627	0.380	0.470
0.200	-0.099	0.373	0.000	0.655	0.639	0.375	0.470
0.250	-0.159	0.494	0.090	0.643	0.608	0.387	0.470
0.300	-0.120	0.574	0.164	0.655	0.602	0.410	0.470
0.400	-0.017	0.664	0.280	0.682	0.599	0.421	0.470
0.500	0.050	0.706	0.279	0.600	0.521	0.435	0.470
0.750	0.073	0.725	0.277	0.473	0.402	0.436	0.470
1.000	0.144	0.698	0.275	0.463	0.388	0.437	0.470
1.500	0.130	0.617	0.273	0.432	0.354	0.461	0.470
2.000	0.127	0.538	0.272	0.423	0.341	0.465	0.470
3.000	0.090	0.414	0.270	0.345	0.287	0.438	0.470
4.000	0.096	0.330	0.174	0.246	0.209	0.430	0.470
5.000	0.084	0.274	0.100	0.216	0.187	0.397	0.470
7.500	0.080	0.215	0.077	0.176	0.156	0.353	0.470
10.000	0.089	0.221	0.060	0.163	0.147	0.343	0.470

$\phi_{S2S'}$ and ϕ_{S2S} are the standard deviation of $\delta S2S$ before and after implementing our site term model.

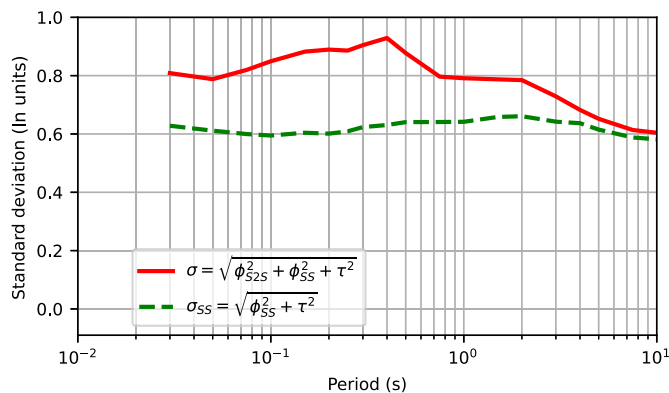


Figure 17. Comparison between the ergodic total standard deviation, σ , and the single-station sigma, σ_{SS} . The τ is from the AG22 GMM. The color version of this figure is available only in the electronic edition.

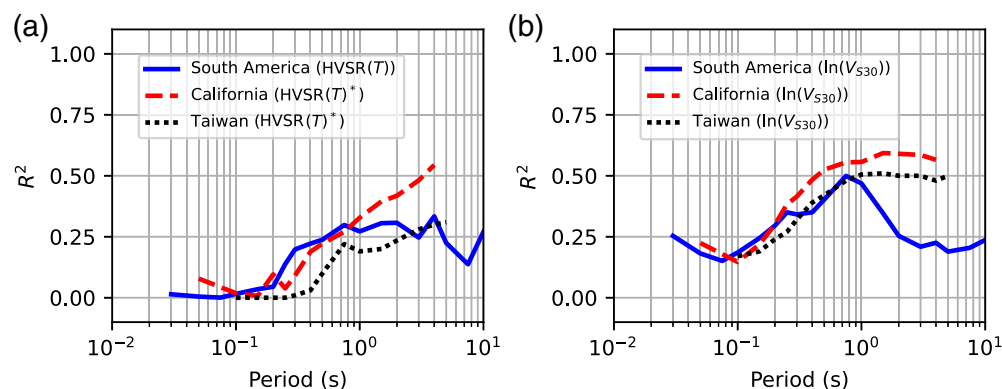


Figure 18. (a) R^2 values for South America using HVSR-based site terms and for Taiwan and California using HVSR*-based site terms. (b) Same as panel (a), but showing the R^2 values for the $\ln(V_{S30})$ -based site terms. For South America, the results are shown for the subset with measured values for V_{S30} . The color version of this figure is available only in the electronic edition.

To better understand the differences in the performance of HVSR(T) and HVSR*(T), we separated the California sites into two V_{S30} bins: stiffer sites with $V_{S30} > 550$ m/s, and softer sites with ($V_{S30} < 550$ m/s). The HVSR*-based R^2 values for these two groups of sites in California are shown in Figure 19. The R^2 is larger for HVSR* than for HVSR for soft sites, whereas the opposite trend is seen for stiff sites. The R^2 values for the Californian stiff sites are similar to the R^2 values for South America. This suggests that in South America HVSR(T) works better than HVSR*(T) because the geologic conditions lead to stiffer sites. All the HVSR data come from Chilean stations, and the Chilean sites tend to be governed by dense gravel and outcrop site conditions. While ground-motion data sets grow and HVSR data and reliable V_{S30} measurements become available, a V_{S30} dependence in the coefficient for HVSR's can be included in the site-term model to take into account this transition between the HVSR scaling of the site terms for soft and stiff sites.

As shown in Figure 12, sites with a clear f_0 had different R^2 values than sites without a clear f_0 . This suggests that part of the regional differences in the R^2 values may be due to differences in the strength of the site resonances between regions in addition to the different sampling of soft and stiff sites.

Subregionalization of South America

The subductive contact in the western margin of the South America plate has an extension of about 7500 km. The convergence rate and the convergence angle with respect to the tectonic contact changes along this 7500 km, inducing changes in the regional geology. In addition, the climate has significant changes between southern Chile and northern Colombia, from cold rainy forests to tropical forests, passing through one of the driest deserts in the world and Mediterranean climates. These changes result in differences in the rock weathering process and soil fabric, which can lead to strong regional differences in the correlation between the site amplification and the V_{S30} or HVSR. While data sets expand, nonergodic GMM methods can be used to evaluate the spatial variation of coefficients of the V_{S30} -based or HVSR-based site-term model (Lavrentiadis *et al.*, 2022).

CONCLUSIONS

The correlation of V_{S30} and site amplification in South America is much weaker than for many other regions of the world. The large fraction of sites with proxy-based V_{S30} in the South America data set is likely the cause of the poor performance

of V_{S30} as a metric for site amplification for the NGA-Sub database for South America. The correlation at short periods ($T < 1$ s) improves greatly by considering only sites with measured V_{S30} values; however, the measured V_{S30} still performs poorly for periods greater than 1 s. The poor performance of V_{S30} at long periods suggests that V_{S30} is not well correlated with the shear-wave velocity profile at depth for sites in South America as compared to other regions. The proxy-based V_{S30} values based on the topographic slope are a poor predictor of the site amplification for sites in western South America. A different approach for estimating the site conditions for sites without measured V_{S30} is needed. One alternative would be to use broad site categories based on surface geology.

A key limitation of the NGA-Sub data set is the small number of recordings per site; however, evaluating the site terms scaling for the full set of 274 sites helps reduce the effect of most sites having only one recording. While the data sets in

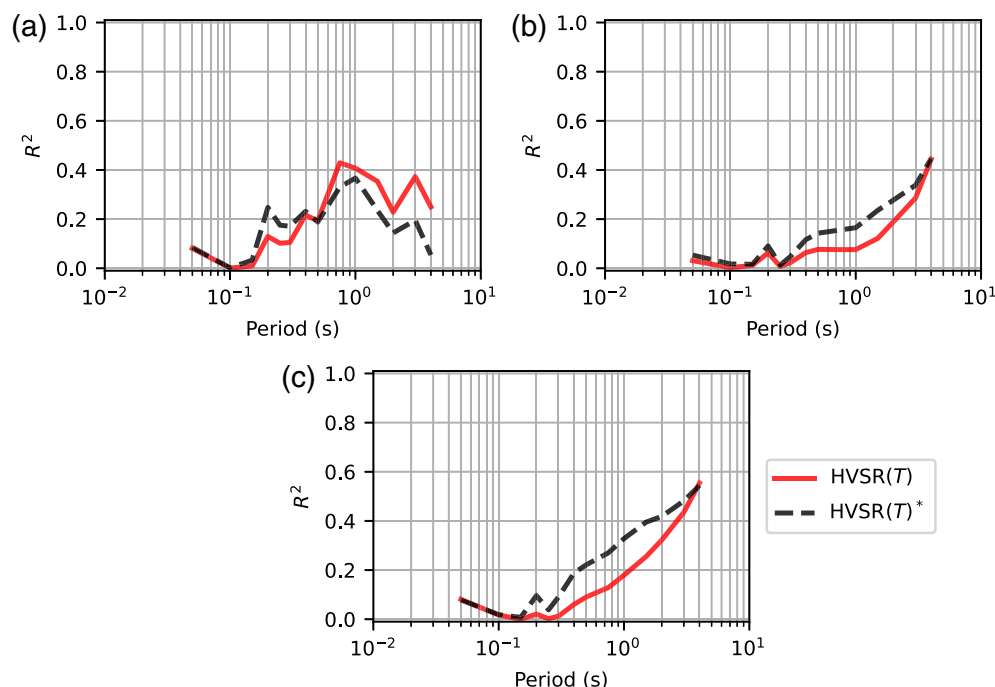


Figure 19. Comparison of the site-term scaling using HVSR* and HVSR for California, adopting the data set from [Pinilla-Ramos et al. \(2022\)](#). (a) Sites with V_{S30} larger than 550 m/s (58 stations). (b) Sites with V_{S30} smaller than 550 m/s (138 stations). (c) All data (196 stations). The color version of this figure is available only in the electronic edition.

South America grow and there are more recordings per site, the conclusions of this study can be checked.

In South America, the HVSR-based site amplification models perform significantly better than V_{S30} -based models at long periods ($T > 1$ s). An HVSR-based site term can explain about 30% of the site term variance at periods of 1 to 4 s compared to the V_{S30} -based site term that explains less than 10% of the long-period variance for the data set that includes the proxy-based V_{S30} values. The HVSR-based site term is becoming a promising approach to constrain site terms in this region, but it explains less of the variance than V_{S30} -based models for California data.

The f_0 site metric by itself can explain up to 37% of the variance of the site terms at periods near 1 s, but adding f_0 as an additional parameter in the site-term model that already includes the HVSR amplitude does not significantly improve the model performance. In addition, only about half of the stations exhibit a f_0 , which limits the applicability of a general site-term model that includes f_0 . In contrast, all the stations have an HVSR(T) spectrum. To incorporate the HVSR scaling into the site term for the subduction GMM and have the model be applicable to all sites, our proposed site-term model is based on the HVSR(T) amplitude and does not include f_0 .

There are several outstanding questions about the causes of the regional differences in the site scaling. Which geologic features between these three regions are the sources of these similarities and discrepancies at different periods? Why does the site-term scaling become weaker above 1 s in South America

compared to California for both parameters: V_{S30} and HVSR amplitude? Why is the scaling of the site terms with V_{S30} similar between California and Taiwan, but the HVSR amplitude performs more poorly in Taiwan? The causes of the differences in the performance of the V_{S30} -based and HVSR-based site terms in different regions is a topic for future research.

DATA AND RESOURCES

The analysis of this study was performed with the data from [Mazzoni et al. \(2021\)](#). In the supplemental material, we provide the horizontal-to-vertical spectral ratio (HVSR) data for the 83 stations considered in this study in Xarray format. The geographic coordinates, V_{S30} values, and within-event residuals from the Next Generation Attenuation-Subduction (NGA-

Sub) data set are included in the two CSV files of the supplement.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

Chih-Hsung (Karen) Sung provided the analysis of the R^2 between HVSR* and the site terms without V_{S30} scaling for Taiwan. Irene Liou provided helpful reviews of the article.

REFERENCES

- Abrahamson, N., and Z. Gulerce (2022). Summary of the Abrahamson and Gulerce NGA-Sub ground-motion model for subduction earthquakes, *Earthq. Spectra* **38**, no. 4, 2638–2681.
- Abrahamson, N., and R. Youngs (1992). A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am.* **82**, no. 1, 505–510.
- Abrahamson, N., W. Silva, and R. Kamai (2014). Summary of the ASK14 ground motion relation for active crustal regions, *Earthq. Spectra* **30**, no. 3, 1025–1055.
- Akcar, S., and Z. Çağnan (2010). A local ground-motion predictive model for Turkey, and its comparison with other regional and global ground-motion models, *Bull. Seismol. Soc. Am.* **100**, no. 6, 2978–2995.
- Al-Atik, L., N. Abrahamson, J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). The variability of ground-motion prediction models and its components, *Seismol. Res. Lett.* **81**, no. 5, 794–801.
- Anderson, J., and J. Brune (1999). Probabilistic Seismic Hazard Analysis without the ergodic assumption, *Seismol. Res. Lett.* **70**, no. 1, 19–28.

- Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology, *Seismol. Res. Lett.* **81**, no. 3, 530–533.
- Bindi, D., F. Pacor, R. Puglia, M. Massa, G. Ameri, and R. Paolucci (2011). Ground motion prediction equations derived from the Italian strong motion database, *Bull. Earthq. Eng.* **9**, 1899–1920.
- Boore, D., E. Thompson, and H. Cadet (2011). Regional correlations of VS30 and velocities averaged over depths less than and greater than 30 meters, *Bull. Seismol. Soc. Am.* **101**, no. 6, 3046–3059.
- Bozorgnia, Y., and J. Stewart (2020). *Data Resources for NGA-Subduction Project (Rept. 2020/02)*, Pacific Earthquake Engineering Research (PEER) Center, Berkeley, California.
- Bozorgnia, Y., N. Abrahamson, L. Al-Atik, T. Ancheta, G. Atkinson, J. Baker, A. Baltay, D. Boore, K. Campbell, B. Chiou, *et al.* (2014). NGA-West2 research project, doi: [10.1193/072113EQS209M](https://doi.org/10.1193/072113EQS209M).
- Bragato, P. L., and D. Slejko (2005). Empirical ground-motion attenuation relations for the eastern Alps in the magnitude range 2.5–6.3, *Bull. Seismol. Soc. Am.* **95**, no. 1, 252–276.
- Contreras, V., F. Ruz, S. Ahdi, and J. Stewart (2018). VS profile database and proxy-based model for VS30 prediction in Chile for NGA-subduction, *Proc. of the 11th National Conf. in Earthquake Engineering, Earthquake Engineering Research Institute*, Los Angeles, California, 25–29 June 2018.
- Danciu, L., and G. Tselentis (2007). Engineering ground-motion parameters attenuation relationships for Greece, *Bull. Seismol. Soc. Am.* **97**, no. 1B, 162–183.
- Kamai, R., N. Abrahamson, and W. Silva (2016). VS30 in the NGA GMPEs: Regional differences and suggested practice, *Earthq. Spectra* **32**, no. 4, 2083–2108.
- Kayen, R., B. Carlin, S. Corbet, C. Pinilla, A. Ng, E. Gorbis, and C. Truong (2014). *Seismic Velocity Site Characterization of Thirty-One Chilean Seismometer Stations by Spectral Analysis of Surface Wave Dispersion Rept. 2014/05*, Pacific Earthquake Engineering Research (PEER) Center, Berkeley, California.
- Lavrentiadis, G., N. Abrahamson, K. Nicolas, Y. Bozorgnia, C. Goulet, A. Babič, J. Macedo, M. Dolšek, N. Gregor, A. Kottke, *et al.* (2022). Overview and introduction to development of non-ergodic earthquake ground-motion models, *Bull. Earthq. Eng.* **21**, 5121–5150, doi: [10.1007/s10518-022-01485-x](https://doi.org/10.1007/s10518-022-01485-x).
- Leyton, F., A. Leopold, G. Hurtado, C. Pastén, S. Ruiz, G. Montalva, and E. Saéz (2018). Geophysical characterization of the Chilean seismological stations: First results, *Seismol. Res. Lett.* **89**, no. 2a, 519–525.
- Mazzoni, S., K. Kishida, S. Ahdi, T. Ancheta, V. Contreras, R. Darragh, N. Kuehn, D. Kwak, A. Kwok, B. Chiou, *et al.* (2021). NGA-Sub Flatfile, doi: [10.34948/N3BC7J](https://doi.org/10.34948/N3BC7J).
- Pinilla-Ramos, C., N. Abrahamson, and R. Kayen (2022). Estimation of site terms in ground-motion models for California using horizontal-to-vertical spectral ratios from microtremor, *Bull. Seismol. Soc. Am.* **112**, no. 6, 3016–3036.
- SERNAGEOMIN (2003). Mapa geológico de Chile, Publicación geológica digital, No. 4 (in Spanish).
- SESAME (2004). Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing, and interpretation, available at https://sesame.geopsy.org/Papers/HV_User_Guidelines.pdf (last accessed June 2024).
- Stewart, J. P., K. Afshari, and C. A. Goulet (2017). Non-ergodic site response in Seismic Hazard Analysis, *Earthq. Spectra* **33**, no. 4, 1385–1414.

Manuscript received 1 February 2024

Published online 3 July 2024