Title
HVSR database and multi-measurement consistency for California sites

Authors
Gospe, Tatiana
Wang, Pengfei
Zimmaro, Paolo
et al.

Publication Date
2022-12-22

Peer reviewed

HVSR DATABASE AND MULTI-MEASUREMENT CONSISTENCY FOR CALIFORNIA SITES

T. Gospe[(1)], P. Wang [(1)], P. Zimmaro[(1,2)] , J.P. Stewart[(1)]

[(1)] *Department of Civil & Environmental Engineering, UCLA*
[(2)] *University of Calabria, Italy*

Abstract

Frequency-dependent horizontal-to-vertical spectral ratios (HVSR) can provide information on site resonant frequencies, which are potentially useful for predicting site amplification. We adapt a relational database developed to archive and disseminate V_S data to include HVSR and investigate the consistency of HVSR derived from different measurements of ambient noise (temporary instruments, permanent instruments) and earthquake recordings. The database as a whole consists of 2,797 sites in California. HVSR consistency is analyzed using subsets of sites with multiple data sources; noise and seismic data are consistent for 60% of sites, whereas different noise measurements have about 75% consistency.

Keywords: horizontal-to-vertical spectral ratios, resonant frequencies, site response, relational database

1. Introduction

Seismic site response is influenced by several factors, including: resonance, nonlinearity, amplification due to impedance contrasts, and amplification related to wave propagation in sedimentary basins. Ground-motion models predict site response conditioned on relatively simple site parameters such as the time-averaged shear wave velocity (V_S) to 30 m depth (V_{S30}) and the depth to 1 km/s or 2.5 km/s V_S ($z_{1.0}$ or $z_{2.5}$) (Bozorgnia et al., 2014). These models are referred to as ergodic (Anderson and Brune, 1999) even if the site parameters are measured on site. The underlying models are ergodic because they are derived from large global or regional databases, and as such are not site-specific.

Any particular site would be expected to produce site amplification that departs from the ergodic estimate for a variety of reasons related to location-specific geologic conditions. A site amplification model that accounts for the effects of these features on site amplification is non-ergodic (e.g., Stewart et al., 2017). One common feature of non-ergodic site response is resonance at one (fundamental site frequency, f_0) or more site frequencies (Di Alessandro et al., 2012; Bonilla et al., 2002; Bonilla et al., 1997), which produce peaks that are smoothed out in ergodic models. While not currently used in NGA models nor in general practice, horizontal-to-vertical Fourier amplitude spectral ratio (HVSR) vs. frequency plots have the potential to add this site-specific attribute to predictions of ergodic site response at low cost, relative to non-ergodic procedures. While V_{S30} provides a reasonable, first-order estimate of site response over a wide frequency range (Seyhan and Stewart 2014), f_0 can be effective at describing site

amplification for frequencies proximate to f_0, but it has limited utility elsewhere. Hence, the two parameters serve different purposes and we postulate that they can be most effectively utilized together (Cadet et al., 2012; Ghofrani et al., 2013).

Current HVSR-based site amplification models, whether using HVSR parameters solely (e.g., Zhao and Xu 2013; Hassani and Atkinson 2016), or in combination with V_{S30} (e.g., Cadet et al. 2012; Ghofrani et al. 2013; Kwak et al. 2017; Hassani and Atkinson 2018a, 2018b; Hashash et al. 2020), are derived using HVSR computed from the same earthquake ground motion data that is being predicted by the model. This model development practice is inconsistent with how the models would be used in forward applications, which will typically be for sites without earthquake recordings. We posit that for HVSR to gain traction in California practice, several technical issues need to be addressed. Among these are the following:

1. Practical best practices for collecting and analyzing HVSR data need to be developed and agreed upon by the informed technical community.
2. A database of HVSR data, assembled to the extent possible in a manner consistent with best practices, should be provided and made publically available.
3. Procedures for identifying when HVSR peaks are present and should be used in model development are needed, as well as procedures for characterizing those peaks (i.e., frequency, amplitude, width).
4. The reliability of HVSR peaks as established from a particular noise-based measurement is needed, under the assumption that the measurement is made by a credible analyst. The issue in this case is the repeatability of HVSR when measured from noise at different times or with different equipment.
5. The consistency of HVSR peaks as established from earthquake data and noise is needed. Noise-based measurements will dominate practical forward applications, but they are intended to predict earthquake shaking attributes. As a result, consistency between HVSR from these two data sources is desirable.
6. Development of HVSR-based site amplification models conditioned on interpretations of HVSR data (i.e., identification of peaks, peak parameters) in combination with V_{S30} and perhaps sediment depth.

The aforementioned models derived from ground motion-based HVSR in effect assume that earthquake- and noise-based HVSR are perfectly consistent (Issue 5) and that noise-based HVSR measurement are fully repeatable (Issue 4).

This paper presents work on the first five issues described above. We extend a V_S profile database (PDB), an early version of which is described by Ahdi et al. (2018), to incorporate HVSR data. Gospe et al. (2020) present a schema for the HVSR components of the database, which shows information that is stored and the results that can be readily extracted for ground motion studies. That paper also explains the data processing procedures and the procedures used to compute HVSR from the data. We describe here the data acquisition process and external (to the database) routines that can be used to evaluate the presence of peaks and identify HVSR-related parameters used for site response studies. The 4th and 5th issues above are also taken up in a preliminary manner using a subset of the full dataset for which noise signals are available from two sources and earthquake recordings are available.

2. Data Sources

2.1 Instrument Types

The database is structured to allow entry of HVSR data from three sources: (1) microtremor array measurements (MAM) obtained from temporary deployments of three-component seismometers specifically targeting noise measurement (Yong et al., 2013); (2) three-component instruments installed temporarily or in permanent housings to record ground motions, but which can also be used to record non-seismic natural ground vibrations noise) -- often these instruments continuously stream data that can be captured; and (3) seismic strong motions (Hassani et al., 2019). Most of the data in NGA databases is from strong motion accelerographs, some of which currently operate with continuous streaming and others of which are triggered. Moreover, modern deployments often feature strong motion accelerographs and co-located relatively sensitive seismometers. Among sites with accelerometers, we have collected HVSR data from sites with co-located continuously-streamed seismometers and we are currently in the process of evaluating the potential for doing this for sites having only accelerometers. For the development of HVSR-based site amplification models, sources 1 and 2 are preferred because these match the data type that would generally be used in forward applications.

Source 1 obtains data from velocity transducers such as Trillium sensors, which are broadband seismometers, whereas Sources 2 and 3 may utilize seismometers or accelerometers. Figure 1 demonstrates the bandwidth and gain for different sensors. Different colors correspond to different sensors, and the dotted vertical line indicates the threshold for the sensors' frequency range. The 40T1, L28, L22, L4C, S13, HS10 and the STS2, 3T, ESP, 40T30, TR240, TR120, TR40 are short period and broadband sensors, respectively (Figure 1). The sensors with the largest bandwidth and highest gain are ideal for our analysis because these sensors provide the best signal resolution. Source 2 may come from velocity transducers or 24-bit accelerometers, and the sensor response with respect to period and signal amplitude is illustrated in Figure 2. In Figure 2, broadband seismometers such as the STS1 capture low earth noise, and accelerometers capture earthquakes. In our study 24-bit accelerometers are likely required so that microtremor signals can be captured.

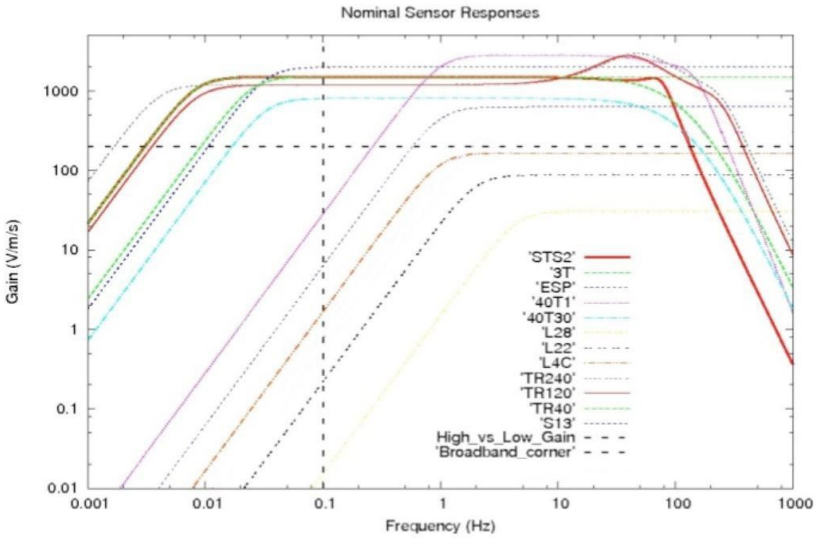

Figure 1. Different sensor responses and the cutoff between broadband and short period sensors as well as low versus high gain sensors. (after IRIS PASSCAL, 2020)

A potential challenge with the use of accelerometers, as might be used with Sources 2-3, is their ability to capture meaningful signals over the frequency range of interest. As shown in Figure 2, the motions from "low earth noise" fall below the range from accelerometers; if true, this suggests that accelerometers cannot record meaningful noise signals for HVSR analysis. Anecdotal evidence, shown in Figure 3, in which HVSR from co-located seismometers and accelerometers are compared, demonstrates how HVSR from accelerometers may not capture low-frequency peaks (in this case below 3 Hz).

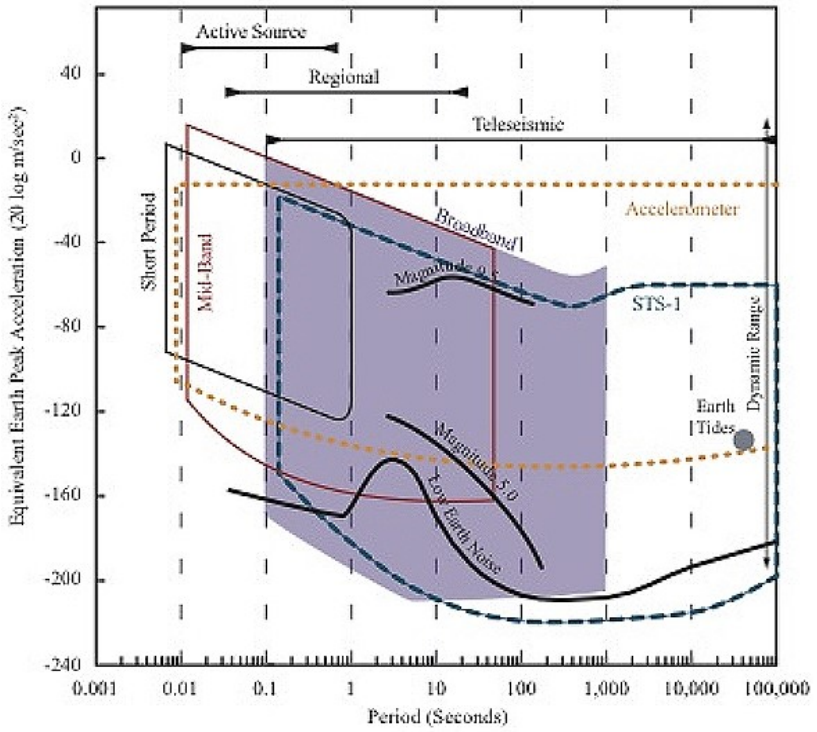

Figure 2. Period and signal amplitudes with respect to sensor response.

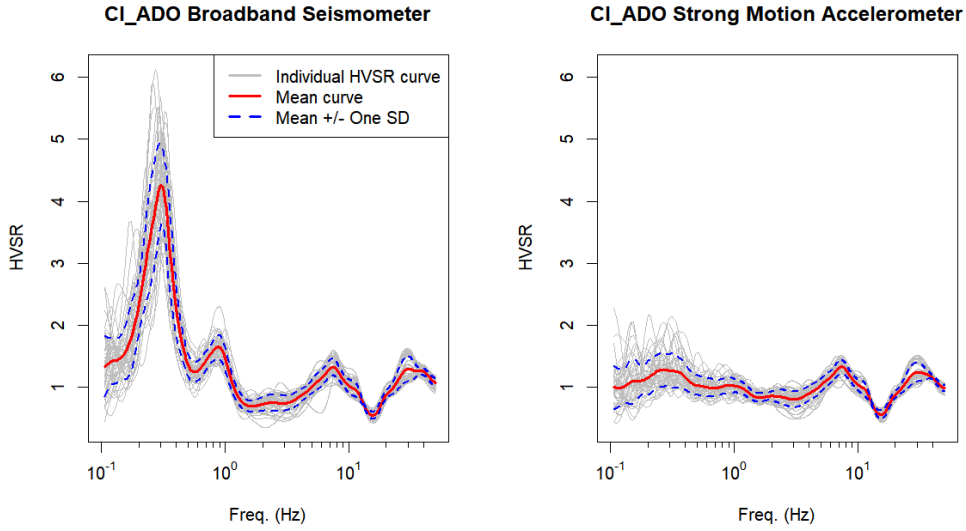

Figure 3. Comparison of HVSR between broadband seismometer and strong motion accelerometer

2.2 HVSR Site Inventory

While in California around 1,700 V_S profiles are publicly available via the PDB (Ahdi et al., 2018), no HVSR site data was available from a public repository in California prior to the present effort. We have assembled a database for HVSR data, which is an extension of the PDB. Because of its preferred utility for site response model development, we have emphasized Source 1 and 2 data in populating the database.

The largest inventory of Source 1 HVSR data at strong motion stations is Yong et al. (2013). The study (aka: American Recovery and Reinvestment Act funded project; hereafter as ARRA project) presents data from 191 strong-motion stations, the majority of which are located in California (187 stations), with an additional four stations in the central and eastern United States. The ARRA data was provided as time-domain signals, which was processed in the manner described in Section 3 of this paper. Yong et al. (2013) provide 589 HVSR results for the 191 sites, due to multiple measurements at most sites. Another major data source is Geometrics, which shared HVSR from 638 sites. This data was provided as mean HVSR-frequency curves, which has been digitized and added to the database. Additional Source 1 contributions included in the database include:

1. 33 sites in the Sacramento-San Joaquin Delta (T. Buckreis, personal communication, 2020).
2. 40 ground motion accelerograph sites maintained by the California Strong Motion Instrumentation Program (CSMIP), part of the California Geological Survey (CGS). Reports are from GEOVision (GEOVision, 2016), Petralogix (Petralogix, 2017), and GEOVision (GEOVision, 2018).
3. 24 sites, some of which are ground motion stations, investigated as part of non-ergodic ground motion investigations by ENGEO (D. Teague, personal communication, 2020).

Time series data from the Delta sites was processed as in Section 3 below. For the CSMIP and ENGEO sites, we obtained mean HVSR-frequency plots, which were added to the database following digitization.

For Source 2, we queried three data centers: Incorporated Research Institutions for Seismology (IRIS), Southern California Earthquake Data center (SCEDC), and the Northern California Earthquake Data Center (NCEDC) (IRIS, 2020; SCEDC, 2013; NCEDC, 2014). We sampled continuously streamed data for 404 sites instrumented with high-gain seismometers with sampling rates between 80-250 Hz. The time series from these data were processed using procedures in Section 3.

Altogether, the database currently contains HVSR data for 1330 sites, locations of which are shown in Figure 4. Many of these sites, including all of the ARRA sites, have HVSR from both Source 1 and Source 2, which causes the number of HVSR entries (1728) to exceed the number of sites (1330). Of the 1330 sites with HVSR, 668 are located in the immediate vicinity of strong motion stations.

Using the data currently incorporated into the PDB, Figure 5 shows the relative number of V_S profiles and HVSR sites in California. Whereas various techniques have been used to collect profile data since the 1960s, the collection of HVSR data is much more recent. The sudden jump in microtremor data is from the present project, using the above sources.

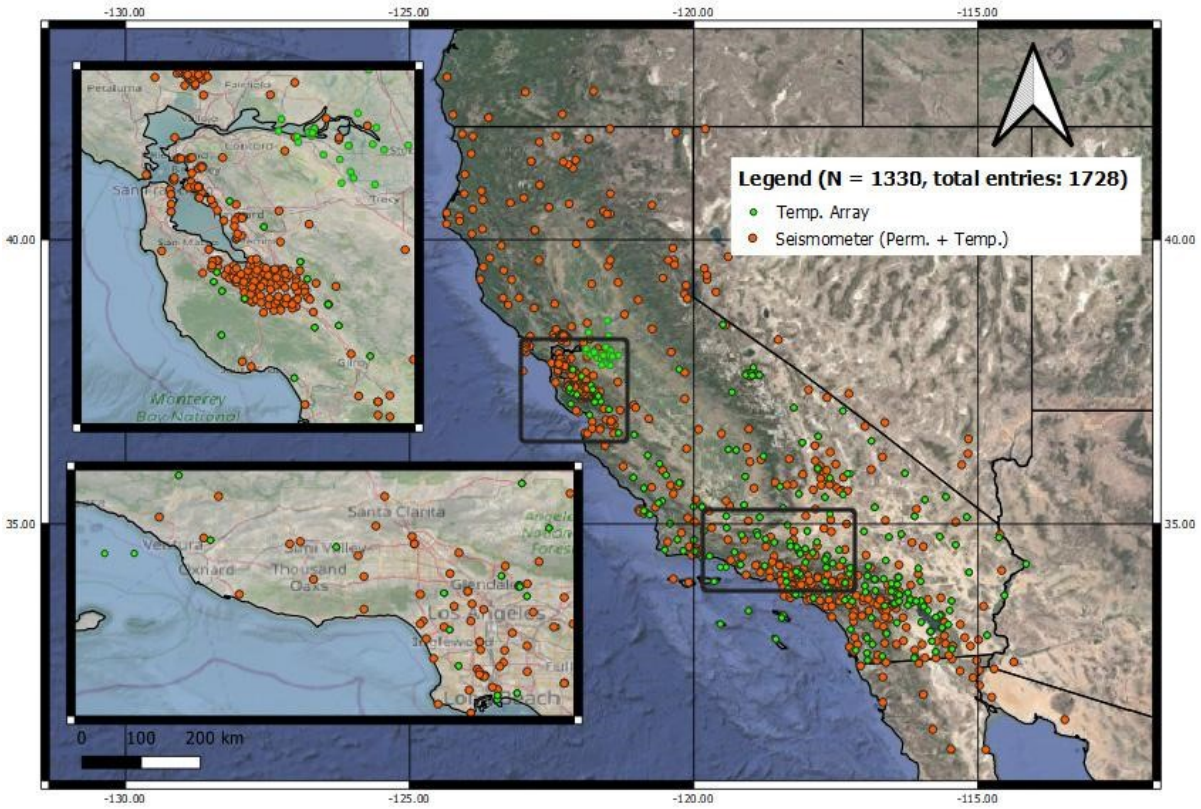

Figure 4. Locations of sites in PDB with HVSR from either temporary deployments (MAM) or continuously streaming ground motion sensors (seismometers).

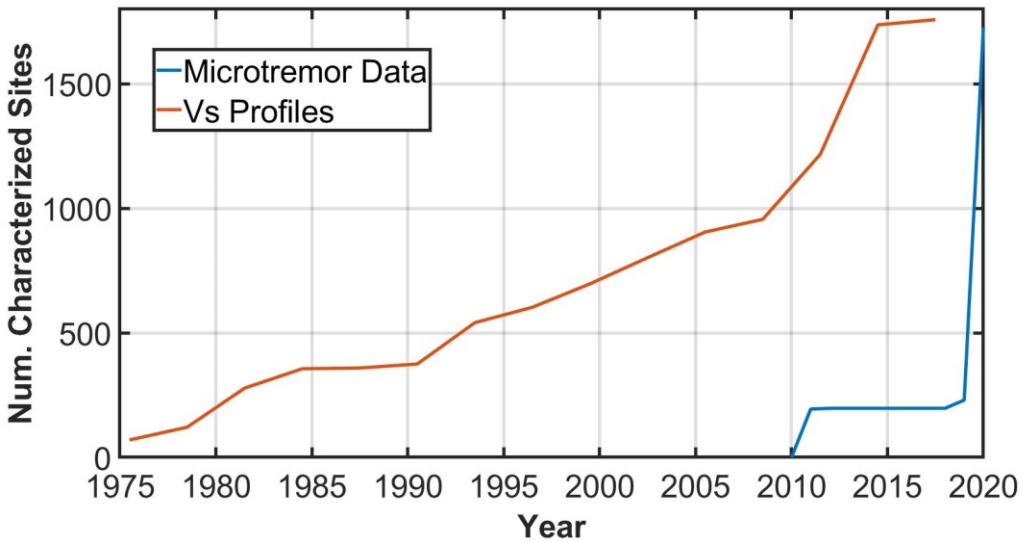

Figure 5. Cumulative distribution of V_S profiles and HVSR data in California versus time.

3. Data Interpretation Tools

The HVSR database provides plots of median-component (RotD50, per Boore 2010) or geometric mean HVSR between time windows and tables showing azimuthal variations but does not provide specific parameters derived from these results, such as might be used as site parameters to supplement V_{S30}. To facilitate such applications, the HVSR data archived in the relational database can be accessed via online Jupyter Notebook tools (example output in

Figure 6) or R-scripts. These tools interact with the data to interpret the data. The interpreted parameters include (1) identification of features as peaks; (2) plots of azimuthal variations of HVSR; and (3) for each peak in the median-component HVSR, fitting of a pulse function to evaluate peak frequency, peak amplitude, and width of peak. Jupyter notebooks are currently available for (2) and R-scripts are available for (1) and (3) (Jupyter notebooks for these tasks are in preparation as of this writing). We envision that such post-processing tools will be used to analyze the data in the cloud without the need to download data locally.

Figure 6 shows an example RotD50 HVSR for the CI.GR2 site (Griffiths Park Observatory) in Los Angeles, California. Site CI.GR2 is located near the nose of a ridge in the Santa Monica Mountains; azimuths from approximately 0-45 deg align approximately with the ridgeline axis, whereas azimuths of 90-180 deg are oriented down-slope for different portions of the ridge nose. The strongest 1 Hz resonance is between azimuths ≈ 110-170 deg, which roughly aligns with the down-slope directions. In these down-slope directions, we expect topographic amplification effects to be strongest (Di Giulio et al., 2009).

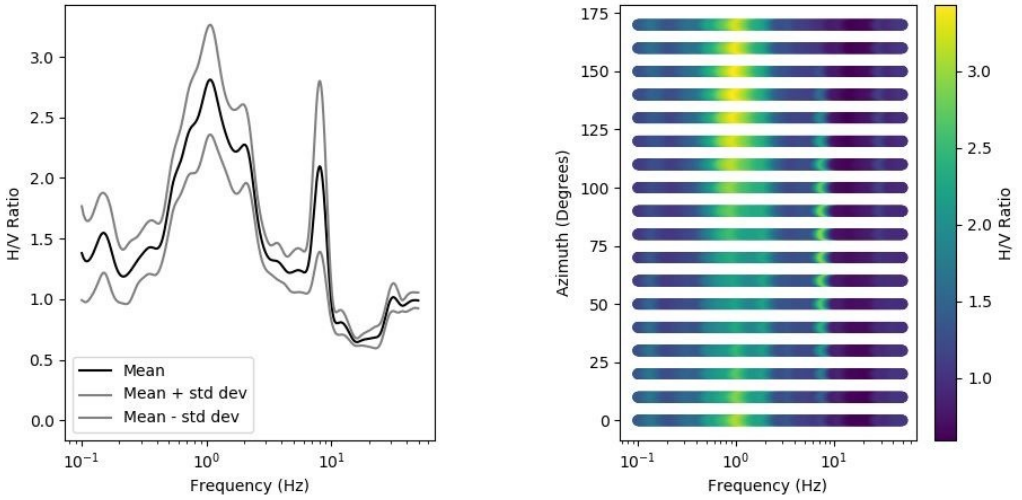

Figure 6. A site near the Griffith Park Observatory in Los Angeles (CI.GR2). Left: frequency versus HVSR from a microtremor recording; right: azimuthal variation of the same recording.

3.1 Peak Identification

HVSR plots can generally be classified as containing no peaks, one peak, or multiple peaks. If there are multiple peaks, we identify the first two peaks (i.e., the two peaks at the lowest frequencies). A peak generally indicates the site has strong impedance contrast(s) near one or more modal frequencies (e.g., Tuan et al., 2011) whereas multiple peaks may indicate multiple impedance contrasts at different depths. When there is no peak present in an HVSR, this suggests the site is either underlain with a sediment-filled depth profile that lacks a significant impedance contrast or it is a rock site with nearly depth-invariant near-surface velocities.

The mean HVSR curve is used for peak identification. SESAME guidelines (SESAME, 2004) provide a procedure for the identification of peaks that first considers three criteria that assess the reliability of the HVSR curve and then considers six conditions intended to establish the presence of a clear HVSR peak. The first two criteria for the reliability of HVSR curves

constrain the minimum required number of sub-windows and duration; these requirements are accounted for in the query and processing procedures described by Gospe et al. (2020). Hence, the additional procedures used to identify peaks are the third reliability criterion and the six conditions, which are listed in Table 1.

Table 1. Reliability criterion and conditions for peak identification from SESAME (2004)

Parameters	CESAME
Reliability 3: $f_{\text{peak}} > 0.5\,\text{Hz}, f \in [0.5f_{\text{peak}}, 2f_{\text{peak}}]$	$\sigma_A(f) < 2$
Reliability 3: $f_{\text{peak}} < 0.5\,\text{Hz}, f \in [0.5f_{\text{peak}}, 2f_{\text{peak}}]$	$\sigma_A(f) < 3$
Clear 1: $f \in [0.25f_{\text{peak}}, f_{\text{peak}}]$	$A_{H/V}(f) < 0.5A_{\text{peak}}$
Clear 2: $f \in [f_{\text{peak}}, 4f_{\text{peak}}]$	$A_{H/V}(f) < 0.5A_{\text{peak}}$
Clear 3:	$A_{\text{peak}} >= 2$
Clear 4: peak of SD curve $f_{\text{peak}}[A_{H/V}(f) - \sigma_A(f)]$	within $[f_{\text{peak}}/1.05, 1.05f_{\text{peak}}]$
Clear 4: peak of SD curve $f_{\text{peak}}[A_{H/V}(f) + \sigma_A(f)]$	within $[f_{\text{peak}}/1.05, 1.05f_{\text{peak}}]$
Clear 5: $f_{\text{peak}} < 0.2\,\text{Hz}$	$\sigma_f < 0.25f_{\text{peak}}$
Clear 5: $f_{\text{peak}} \in [0.2, 0.5)\,\text{Hz}$	$\sigma_f < 0.2f_{\text{peak}}$
Clear 5: $f_{\text{peak}} \in [0.5, 1.0)\,\text{Hz}$	$\sigma_f < 0.15f_{\text{peak}}$
Clear 5: $f_{\text{peak}} \in [1.0, 2.0]\,\text{Hz}$	$\sigma_f < 0.1f_{\text{peak}}$
Clear 5: $f_{\text{peak}} > 2.0\,\text{Hz}$	$\sigma_f < 0.05f_{\text{peak}}$
Clear 6: $f_{\text{peak}} < 0.2\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 3$
Clear 6: $f_{\text{peak}} \in [0.2, 0.5)\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 2.5$
Clear 6: $f_{\text{peak}} \in [0.5, 1.0)\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 2$
Clear 6: $f_{\text{peak}} \in [1.0, 2.0]\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 1.78$
Clear 6: $f_{\text{peak}} > 2.0\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 1.58$

In Table 1, f_{peak} is the peak frequency of interest (there could be multiple f_{peak} values in a single curve); f is the independent frequency; $A_{H/V}(f)$ is the amplitude of the HVSR mean curve at frequency f; A_{peak} is the amplitude at f_{peak}; $\sigma_A(f)$ is the standard deviation of $A_{H/V}(f)$ at f; $\sigma_A(f_{\text{peak}})$ is the standard deviation of $A_{H/V}(f)$ at f_{peak}; and σ_f is the standard deviation of f_{peak}. In Table 1, the rows labelled Reliability 3, Clear 5, and Clear 6 are f_{peak}-dependent. The greater f_{peak} is, the more stringent are the standards for establishing a peak as reliable and clear.

The six conditions consider factors such as the amplitude of the peak relative to ordinates at neighboring frequencies and the width of the peak. In the case of the CI.GR2 site, the conditions are all satisfied except for #5, which is not satisfied (the peak is too wide).

Examination of similar results from many sites suggest that the criteria in SESAME (2004) are too conservative. Alternative criteria are developed that are more effective at identifying the presence of peaks at California sites (Wang 2020). These criteria were established based on visual inspections of HVSR to identify sites with peaks, and for the subset of those sites that fail SESAME criteria, identification of the SESAME criteria that are not satisfied. The new recommended criteria are summarized in Table 2 which excludes the Clear 5 condition and weakens other conditions.

Table 2. Suggested new reliability criterion and peak identification conditions, modified from SESAME (2004) by Wang (2020).

Parameters	New Criteria
Reliability 3: $f_{\text{peak}} > 0.5\,\text{Hz}, f \in [0.5f_{\text{peak}}, 2f_{\text{peak}}]$	$\sigma_A(f) < 2$
Reliability 3: $f_{\text{peak}} < 0.5\,\text{Hz}, f \in [0.5f_{\text{peak}}, 2f_{\text{peak}}]$	$\sigma_A(f) < 3$
Clear 1: $f \in [0.25f_{\text{peak}}, f_{\text{peak}}]$	$A_{H/V}(f) < 0.6A_{\text{peak}}$
Clear 2: $f \in [f_{\text{peak}}, 4f_{\text{peak}}]$	$A_{H/V}(f) < 0.6A_{\text{peak}}$
Clear 3:	$A_{\text{peak}} >= 1.6$
Clear 4: peak of SD curve $f_{\text{peak}}[A_{H/V}(f) - \sigma_A(f)]$	within $[f_{\text{peak}}/1.15, 1.15f_{\text{peak}}]$
Clear 4: peak of SD curve $f_{\text{peak}}[A_{H/V}(f) + \sigma_A(f)]$	within $[f_{\text{peak}}/1.12, 1.12f_{\text{peak}}]$
Clear 5: $f_{\text{peak}} < 0.2\,\text{Hz}$	-
Clear 5: $f_{\text{peak}} \in [0.2, 0.5)\,\text{Hz}$	-
Clear 5: $f_{\text{peak}} \in [0.5, 1.0)\,\text{Hz}$	-
Clear 5: $f_{\text{peak}} \in [1.0, 2.0]\,\text{Hz}$	-
Clear 5: $f_{\text{peak}} > 2.0\,\text{Hz}$	-
Clear 6: $f_{\text{peak}} < 0.2\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 3$
Clear 6: $f_{\text{peak}} \in [0.2, 0.5)\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 2.5$
Clear 6: $f_{\text{peak}} \in [0.5, 1.0)\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 2$
Clear 6: $f_{\text{peak}} \in [1.0, 2.0]\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 1.78$
Clear 6: $f_{\text{peak}} > 2.0\,\text{Hz}$	$\sigma_A(f_{\text{peak}}) < 1.58$

An R script implements these criteria and determines if an HVSR curve contains a peak. The R script allows the user to select the conditions to be satisfied for assessing the presence of a peak, and notifies the user of which conditions the a particular peak satisfies.

3.2 Peak Fitting

For mean HVSR plots with a peak, we fit a Gaussian pulse function adapted from Hassani and Atkinson (2016) as follows (Wang 2020):

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

where f_{pi} is the fitted peak frequency, $c_{1,i}$ is the peak amplitude relative to $c_{0,i}$, w_i is peak width, $c_{0,i}$ is a frequency-independent constant, i is the order of peak, and f is frequency in Hz. The fit is performed using nonlinear regression in R with the *Optim* function, which minimizes the sum of squared errors. Figure 7 shows results for the CI.GR2 site, which contains a peak of amplitude 2.8 at frequency 1.2 Hz.

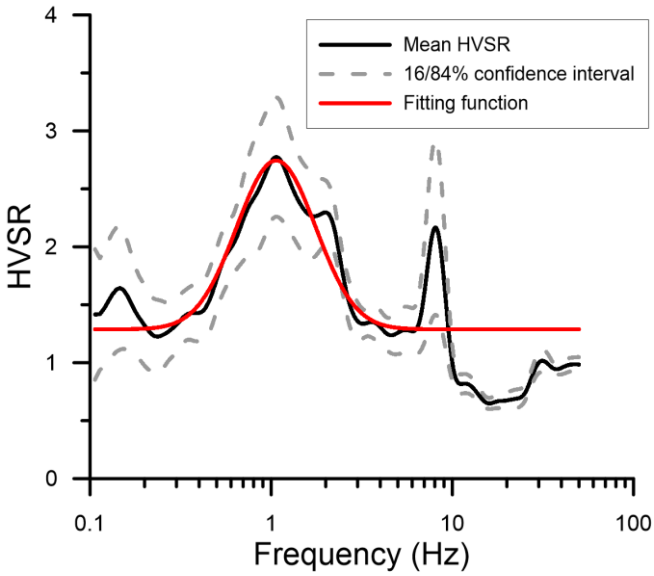

Figure 7. RotD50 HVSR for CI.GR2 site with Gaussian fit to the peak using Eq. (1).

4. HVSR Comparisons Between Data Sources

In the HVSR database, we have processed and stored three types of data, microtremor array measurements (MAM), microtremor noise queried from permanently installed seismometers continuously streamed data (CSD), and recorded seismic ground motions. As described in the Introduction, the reliability and consistency of HVSR are important questions related to the eventual development of practice-oriented HVSR models. Here we perform a preliminary investigation of these questions using a dataset consisting of 102 sites with both MAM and CSD HVSR, and a related dataset of 138 sites with both noise and seismic HVSR. In the following subsections, we investigate differences in noise-vs-seismic HVSR and noise-based HVSR. The comparisons are made in terms of presence of peaks, fitted peak frequencies, and fitted peak amplitudes.

4.1 Comparison Between Earthquake- and Noise-Based HVSR

As described in Section 3.1, HVSR mean curves can be classified into two broad categories, clear peaks and no peaks. Using the criteria in Section 3.1, we have identified the presence of peaks for a group of 138 sites with HVSR from common instruments that have recorded earthquake motions and CSD. The 138 sites can be divided into four groups: (1) both data sources produce peaks, (2) both data sources produce no peak, (3) earthquake ground motion HVSR has a peak but CSD HVSR does not, and (4) CSD HVSR has a peak but earthquake ground motion HVSR does not. The breakdown of sites into these four groups is presented in Table 3. Figure 8 shows examples of "P-P", "N-N", "N-P", and "P-N" sites.

Table 3. The comparison of peaks presence from HVSR computed using strong motion and CSD

Eqk: CSD	Pk.: Pk.	No Pk. : No Pk.	No Pk. : Pk.	Pk.: No Pk.
Count	39	45	35	19
Percent	~28%	~33%	~25%	~14%

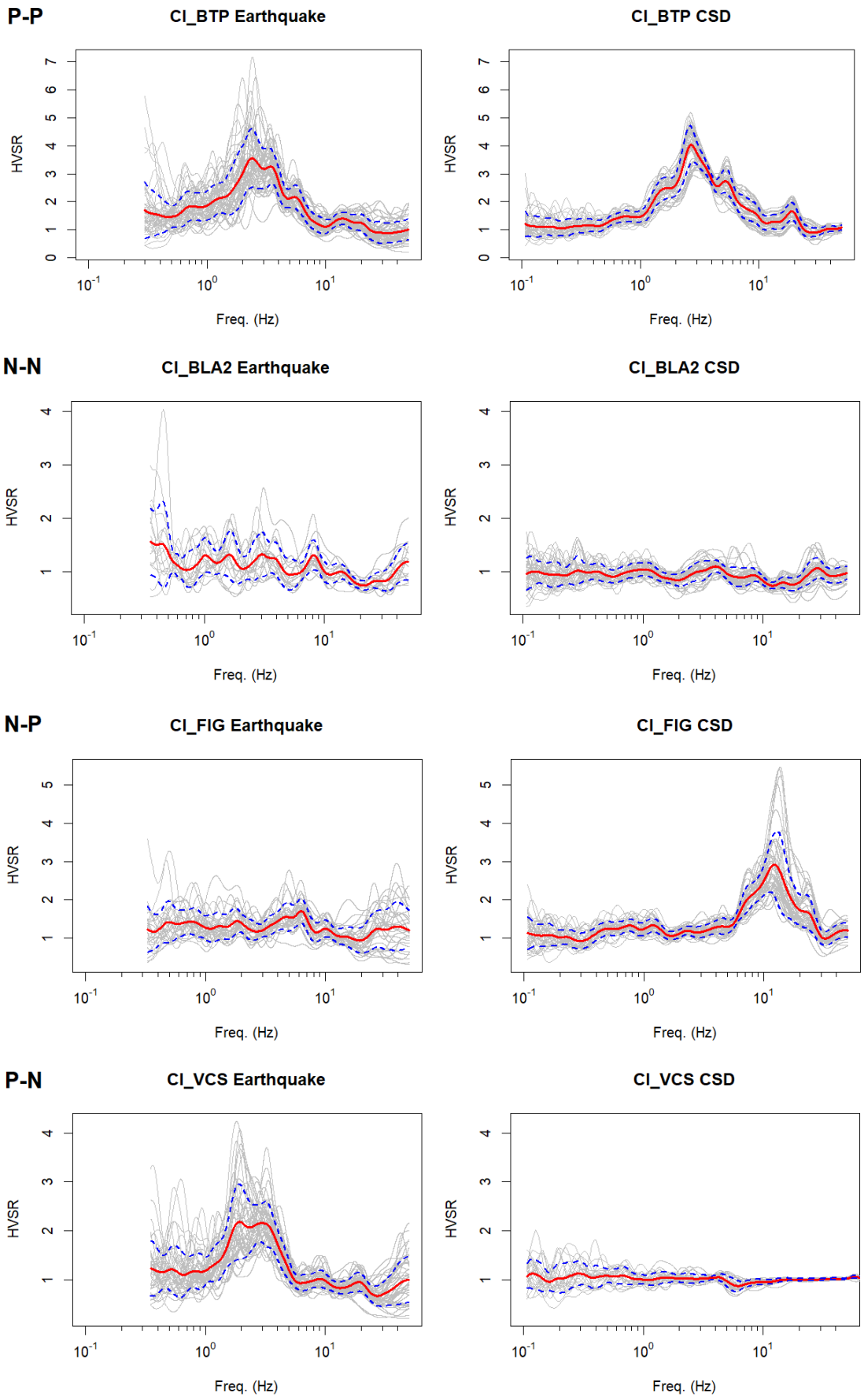

Figure 8. The examples of HVSR for "P-P", "N-N", "N-P", and "P-N" sites.

Table 3 shows that 61% of sites produce consistent results from the noise and earthquake HVSR ("P-P" and "N-N" groups) . Among the 74 sites with peaks from noise-based HVSR, approximately half have peaks in seismic HVSR. This suggests the potential for a significant rate of "false positives" (peaks identified from noise that are not present in ground motions). Among the 64 sites without peaks from noise-based HVSR, approximately ⅔ also lack peaks in seismic HVSR. This suggests a relatively low rate of false negatives. If these rates of false positives and false negatives persist in the larger database that will be used for model development, it will add uncertainty to HVSR-based models.

An additional important question is: if HVSR from both data sources have peaks, then how do the fitted coefficients from the two sources compare? To investigate this question, we compare fitted parameters for the 39 "P-P" sites in Figure 9. The figure shows that most points are along the 45-degree line (15 sites have f_p misfits < 20%), however, 9 sites have misfits that exceed a factor of four. Overall, the peak frequencies are moderately correlated (correlation coefficient, ρ = 0.65). The plot of a_p indicates a weaker correlation. There are more points below the 45-degree line, which indicates that peak amplitudes from earthquake HVSR are generally slightly larger than those from noise HVSR. This finding is consistent with strong motion versus noise comparisons found in soft sites in Mexico (Lermo and Chávez-García, 1994), sites in Iceland (Field et al., 1995), Greece (Atakan et al., 1997), the Garner Valley array in California (Lachet et al., 1996), southern Italy (Theodulidis et al., 1996), and various sites across Europe (Mucciarelli et al., 2003), the Caribbean, and Tehran (Haghshenas et al., 2008).

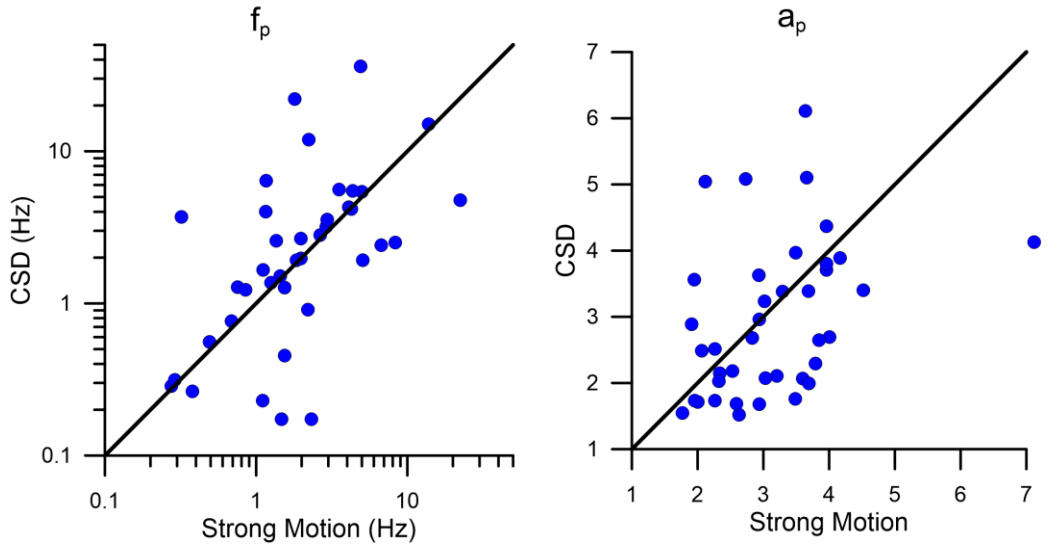

Figure 9. Comparison of peak fitted parameters f_p and a_p from earthquake and noise (CSD) data

4.2 Comparison Between Microtremor- and Continuously Steamed Noise HVSR

Similar to Section 4.1, we have identified the presence of peaks for a group of 102 sites with HVSR derived from ambient noise as recorded by MAMs and CSD. The instruments that made these recordings are not co-located, because the MAM sensors could not always be positioned directly adjacent to the strong motion station (Yong et al. 2013). The statistics of peaks and no peaks are presented in Table 4. Figure 10 illustrates examples of "P-P", "N-N", "N-P", and "P-N" sites.

Table 4. The comparison of peaks presence from HVSR computed using MAM and CSD

Eqk: CSD	Pk.: Pk.	No Pk. : No Pk.	No Pk. : Pk.	Pk.: No Pk.
Count	52	24	8	18
Percent	~51%	~24%	~8%	~17%

Table 4 shows that 75% of sites produce consistent peak identification results from the two noise-based HVSR ("P-P" and "N-N" groups). We have no reason to suspect one noise measurement is preferred to another, so these data reflect the reliability of HVSR when only a single measurement is made (there is a ¾ chance that a second measurement would produce a similar outcome regarding the presence of a peak).

We compare the fitted coefficients from the two sources for the 52 "P-P" sites in Figure 11. Of the 52 sites, 80% have f_p values within 20% of each other, and only 20% have misfits > a factor of four. The correlation coefficient is 0.87. Inspections of sites that are located off of the 1:1 line (Figure 12) show that the peaks in these cases are relatively weak, falling only marginally within the peak category. As this work progresses and the peak identification procedures are refined, some of these sites might be re-classified as no-peak sites. The plot of peak amplitudes (a_p) in Figure 11 indicates a weaker correlation (ρ=0.64) than the f_p results.

4.3 Discussion

As explained in the *Introduction*, the purpose for measuring and compiling HVSR data is to use it for the derivation of site parameters that can be used in ground motion models, as an augment to V_{S30}. In a typical forward application (i.e., use of a model to predict ground motions that have not yet occurred), an engineer will measure HVSR at the site of interest, decide if a peak is present, and if so, identify peak parameters. The comparisons in Section 4.2 shows that had the engineer made the noise measurement in a slightly different manner, and perhaps at a different time, the likelihood of obtaining a significantly different outcome is small but not negligible. Studies of this type, comparing results from multiple noise-based measurements, are relatively rare in the literature, so we are unable to compare to previous findings.

The results in Section 4.1 show that if a peak is identified, there is only about a 50% chance that a peak will also be present in seismic HVSR data. This high rate of false-positives will decrease, but not eliminate, the effectiveness of models conditioned on HVSR peak parameters. On the other hand, if no-peak is identified, there is a strong likelihood that the seismic HVSR also lacks peaks. The consistency of seismic and noise-based HVSR peaks has been studied previously, with most investigators finding consistent results (Lermo and Chávez-García, 1994; Field et al., 1995; Atakan et al., 1997; Lachet et al., 1996; Theodulidis et al., 1996; Mucciarelli et al., 2003; Haghshenas et al., 2008; and Hassani et al. 2019) and a few finding some inconsistent results (Satoh et al. 2001). Comparisons of HVSR from the two vibration sources might well vary depending on site geology, so further investigation of this issue for site conditions in California is needed.

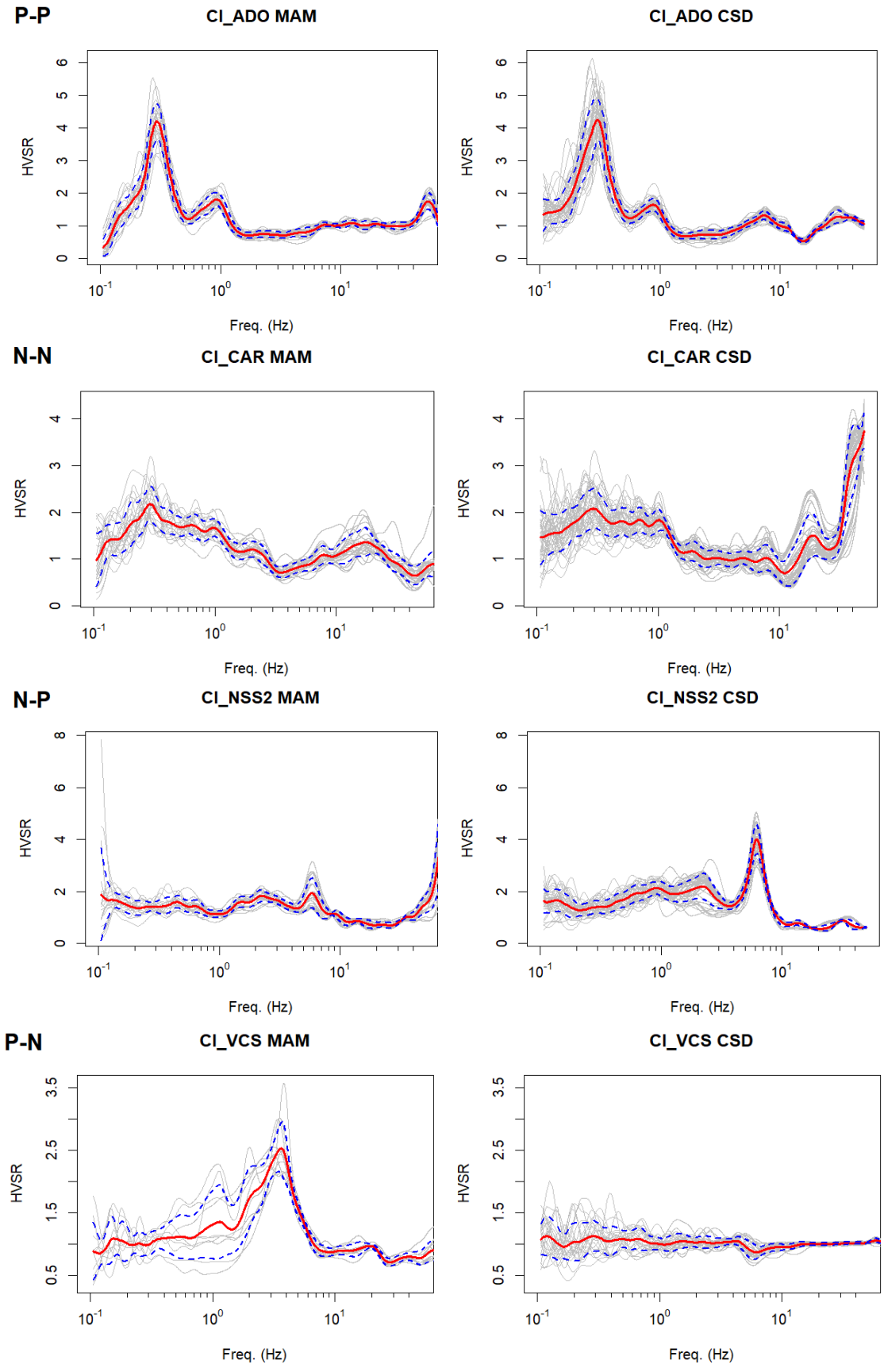

Figure 10. The examples of HVSR for "P-P", "N-N", "N-P", and "P-N" site for MAM versus CSD.

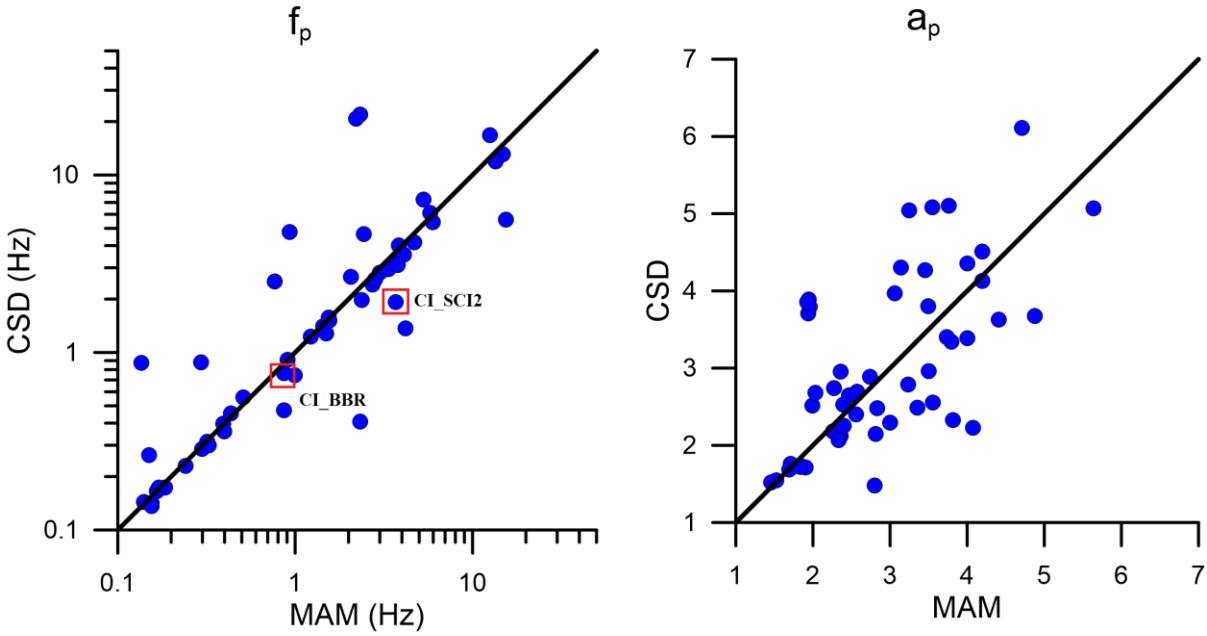

Figure 11. Comparison of f_p (left) and a_p (right) of peak fitted parameters between CSD and MAM.

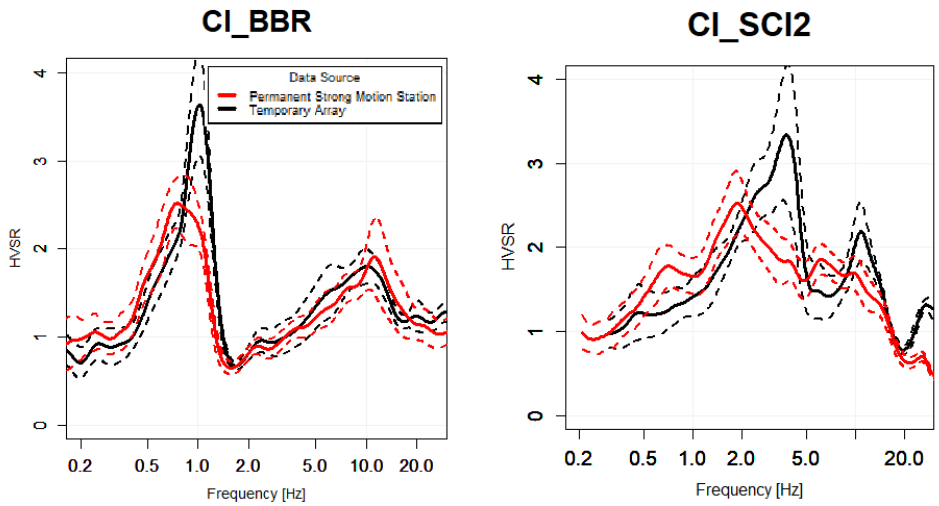

Figure 12. HVSR for permanent strong motion stations and temporary arrays.

5. Conclusions

Because HVSR-based parameters are not used currently in ground motion prediction applications, a number of steps are required to support eventual model development and utilization in practice. This study represents a step in that direction. We have created an open-source relational database of HVSR and associated processing parameters and incorporated this information into an existing community *Vs* Profile Database (PDB) in the United States. Users can utilize and analyze the processed records through interactive Jupyter Notebook tools that evaluate azimuthal dependence, identify the presence of peaks in an HVSR, and fit peaks using Eq. (1).

To demonstrate the value of the compiled data, we compare HVSR attributes for seismic and noise based data, and for two different noise measurements. The different noise measurements are more consistent with each other than the noise-to-seismic comparison. Of the California sites considered, about 30-40% do not have peaks. Accordingly, it will be important for eventual HVSR-based models to be able to accommodate this common result of HVSR testing.

Acknowledgments

Funding for this study is provided by California Strong Motion Instrumentation Program, California Geological Survey, Agreement 1016-985 and 1018-569. We gratefully acknowledge this support. We also thank Antony Martin from GEOVision for sharing his expertise on processing the HVSR data. Site data was provided for addition to the database by Alan Yong (USGS), Koichi Hayashi (Geometrics), David Teague (Engeo), and Tristan Buckreis (UCLA); this data sharing and cooperation are deeply appreciated. We thank Silvia Castellaro, Seth Carpenter, Brady Cox, Hiroshi Kawase, Shinichi Matsushima, Robert Nigbor, Stefano Parolai, Marco Pilz, Lisa Schleicher, Jamison Steidl, Alan Yong, and Wang Zhenming for their input on this research. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Ahdi, S.K., O. Ilhan, S. Sadiq, Y. Bozorgnia, Y.M.A. Hashash, D. Park, A. Yong, and J.P. Stewart, 2018. Development of a United States Community Shear Wave Velocity Profile Database, *5th Conference on Geotechnical Earthquake Engineering and Soil Dynamics (GEESD-V)*, June 10–13, 2018, Austin, Texas.

Anderson, J. G. and J. Brune, 1999. Probabilistic seismic hazard analysis without the ergodic assumption." *Seism. Res. Lett.*, **70**, 19-28.

Atakan, K., B. Brandsdottir, P. Halldorsson, and G.O. Fridleifsson, 1997. Site response as a function of near-surface geology in the South Iceland seismic zone. *Natural Hazards*, *15*(2-3), 139-164.

Bonilla, L. F., J. H. Steidl, G. T. Lindley, A. G. Tumarkin, and R. J. Archuleta, 1997. Site amplification in the San Fernando Valley, California: Variability of site-effect estimation using the S-wave, coda, and H/V methods, *Bull. Seismol. Soc. Am.* **87,** 710-730.

Bonilla, L.F., J. H. Steidl, J.-C. Gariel, and R.J. Archuleta, 2002. Borehole response studies at the Garner Valley downhole array, Southern California, *Bull. Seismol. Soc. Am.* **92,** 3165-3179.

Boore, D. M., 2010. Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bull. Seismol. Soc. Am.* **100,** 1830-1835.

Bozorgnia, Y., N. A. Abrahamson, L. .A. Atik, T. D. Ancheta, G. M. Atkinson, J. W. Baker, A. Baltay, D. M. Boore, K. W. Campbell, B. S.-J. Chiou, R. Darragh, S. Day, J. Donahue, R. W. Graves, N. Gregor, T. Hanks, I. M. Idriss, R. Kamai, T. Kishida, A. Kottke, S. A. Mahin, S. Rezaeian, B. Rowshandel, E. Seyhan, S. Shahi, T. Shantz, W. Silva, P. Spudich, J. P.

Stewart, J. Watson-Lamprey, K. Wooddell, and R. Youngs, 2014. NGA-West 2 research project*, Earthq. Spectra*, **30**, 973-987.

Cadet, H., P.-Y. Bard, A.-M. Duval, and E. Bertrand, 2012. Site effect assessment using KiK-net data: Part 2—Site amplification prediction equation based on f0 and Vsz, *Bull. Earthq. Eng.* **10**, 451–489.

Di Alessandro, C., L. F. Bonilla, D. M. Boore, A. Rovelli, and O. Scotti, 2012. Predominant-period site classification for response spectra prediction equations in Italy, *Bull. Seismol. Soc. Am.* **102**, 680–695.

Di Giulio G., F. Cara, A. Rovelli, G. Lombardo, and R. Rigano, 2009. Evidences for strong directional resonances in intensely deformed zones of the Pernicana fault, Mount Etna, Italy. *J. Geophy. Res.*, **114**, B10308.

Field, E. H., A.C. Clement, K.H. Jacob, V. Aharonian, S.E. Hough, P.A. Friberg, H.A. Abramian, 1995. Earthquake site-response study in Giumri (formerly Leninakan), Armenia, using ambient noise observations. *Bull. Seismol. Soc. Am.*, **85**(1), 349-353.

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.

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.

Ghofrani, H., G.M. Atkinson, and K. Goda, 2013. Implications of the 2011 M 9.0 Tohoku Japan earthquake for the treatment of site effects in large earthquakes, *Bull. Earthq. Eng.* **11,** 171-203.

Gospe, T., P. Zimmaro, P. Wang, T. Buckreis, S.K. Ahdi, A.K. Yong, S.J. Brandenberg, J.P. Stewart, 2020. Supplementing shear wave velocity profile database with microtremor-based H/V spectral ratios, *Proc. 17th World Conf. on Earthquake Engineering*, Sendai, Japan, September 14-18 2020. Paper 003010.

Haghshenas, E., P.Y. Bard, N. Theodulidis, and Sesame WP04 Team, 2008. Empirical evaluation of microtremor H/V spectral ratio. *Bull. Earthq. Eng.*, **6**(1), 75-108.

Hashash, Y.M.A., O. Ilhan, B. Hassani, G.M. Atkinson, 2020. Significance of site natural period effects for linear site amplification in central and eastern North America: Empirical and simulation-based models, *Earthq. Spectra*, **36**, 87-110.

Hassani, B. and G.M. Atkinson, 2016. Site-effects model for central and eastern North America based on peak frequency, *Bull. Seismol. Soc. Am.* **106**, 653–664.

Hassani, B., and G.M. Atkinson, 2018a. Site-effects model for central and eastern North America based on peak frequency and average shear-wave velocity, *Bull. Seismol. Soc. Am.* **107**, 338-350.

Hassani, B., and G.M. Atkinson, 2018b. Application of a site-effects model based on peak frequency and average shear-wave velocity to California, *Bull. Seismol. Soc. Am.* **108**, 351-357.

Hassani, B., Yong, A., Atkinson, G. M., Feng, T., and Meng, L. (2019). Comparison of site dominant frequency from earthquake and microseismic data in California. *Bull. Seismol. Soc. Am.*, **109**(3), 1034-1040.

Incorporated Research Institutions for Seismology (IRIS), IRIS PASSCAL, https://www.passcal.nmt.edu/content/instrumentation/sensors/sensor-comparison-chart, last accessed <10/12/2020>

Kwak, D.Y., J.P. Stewart, S.U.J. Mandokhail, and D. Park, 2017. Supplementing V_{S30} with H/V spectral ratios for predicting site effects. *Bull. Seismol. Soc. Am.* **107**, 2028-2042.

Lachet, C., D. Hatzfeld, P.Y. Bard, N. Theodulidis, C. Papaioannou, and A. Savvaidis, 1996. Site effects and microzonation in the city of Thessaloniki (Greece) comparison of different approaches. *Bull. Seismol. Soc. Am.*, **86**, 1692-1703.

Lermo, J., and F.J. Chávez-García, 1994. Are microtremors useful in site response evaluation?. *Bull. Seismol. Soc. Am.*, **84**, 1350-1364.

Mucciarelli, M., M. R. Gallipoli, and M. Arcieri, 2003. The stability of the horizontal-to-vertical spectral ratio of triggered noise and earthquake recordings. *Bull. Seismol. Soc. Am.* **93**(3), 1407-1412.

NCEDC, 2014, Northern California Earthquake Data Center. UC Berkeley Seismological Laboratory. Dataset. doi:10.7932/NCEDC.

Petralogix, 2017. VS30 Site Characterization Report, Los Angeles, Orange, Ventura, San Bernardino, and Riverside Counties. Report 2017-00006.

Satoh, T., H. Kawase, and S. Matsushima, 2001. Differences between site characteristics obtained from microtremors, S-waves, P-waves, and codas, *Bull. Seismol. Soc. Am*. **91**, 313–334.

SCEDC, 2013. Southern California Earthquake Data Center. Caltech. Dataset. doi:10.7909/C3WD3xH1

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, accessed September 2012, at http://sesame-fp5.obs.ujf-grenoble.fr/

Seyhan, E., and Stewart, J. P. 2014. Semi-Empirical Nonlinear Site Amplification from NGA-West2 Data and Simulations, *Earthq, Spectra*, **30**, 1241-1256.

Stewart, J. P., K. Afshari, and C. A. Goulet, 2017. Non-ergodic site response in seismic hazard analysis, *Earthq, Spectra*, **33**, 1385-1414.

Theodulidis, N., P.Y. Bard, R. Archuleta, and M. Bouchon, 1996. Horizontal-to-vertical spectral ratio and geological conditions: The case of Garner Valley downhole array in southern California. *Bull. Seismol. Soc. Am*., **86**(2), 306-319.

Tuan, T. T., F. Scherbaum, and P. G. Malischewsky, 2011. On the relationship of peaks and troughs of the ellipticity (H/V) of Rayleigh waves and the transmission response of single layer over half-space models, *Geophys. J. Int*. **184**, 793–800.

Wang, P., 2020. Predictability and Repeatability of Non-Ergodic Site Response for Diverse Geological Conditions. Ph.D. thesis. UC Los Angeles.

Yong, A., A. Martin, K.H. Stokoe, and J. Diehl, 2013, ARRA-funded VS30 measurements using multi-technique approach at strong-motion stations in California and central-eastern United States: U.S. Geological Survey Open-File Report 2013–1102, 59 p. and data files, http://pubs.usgs.gov/of/2013/1102/.

Zhao, J.X., and H. Xu, 2013. A comparison of V_{S30} and site period as site-effect parameters in response spectral ground-motion prediction equations, *Bull. Seismol. Soc. Am.* **103**, 1–18.