

USGS Award No. G13AP00070

SITE RESPONSE MAPPING WITH ONE LESS PROXY:
COLLABORATIVE RESEARCH WITH SDSU AND THE USGS

FINAL TECHNICAL REPORT

SUBMITTED: November 2014

Eric M. Thompson
Geological Sciences
San Diego State University
5500 Campanile Dr.
San Diego, CA 92182-1020
ethompson@mail.sdsu.edu
Tel: 858-848-0095

David J. Wald
U.S. Geological Survey
Golden, CO

AWARD PERIOD: May 2013 to April 2014

Abstract

The goal of this project is to process and analyze site-specific empirical amplification factors (EAFs). EAFs are the ratio of the recorded motions to a common reference rock motion in order to best capture the actual site amplifications. EAFs are computed at seismic stations within California that have sufficient earthquake recordings to determine robust amplification factors. As a representation of observed amplification, EAFs have a number of useful applications for understanding and mapping site response as well as analyzing its uncertainty. Site response is more generally estimated through empirical correlations with the average shear-wave velocity to 30 m (V_{S30}). For mapping purposes, a second proxy is then needed to estimate V_{S30} , such as surface geology, topographic slope, or terrain. One promising strategy for improving the accuracy of site response maps is to map the amplification directly, rather than rely on mapped values of V_{S30} as an intermediate step. In contrast to previous studies that have focused on compiling databases of V_{S30} measurements, we focus on compiling a database of EAFs. Site-specific EAFs at stations that have recorded numerous ground motions provide more accurate estimates of site response than approximations based on V_{S30} and each EAF can be directly compared with amplification factors inferred from V_{S30} . We describe what conditions these simple approximation site factors are valid and show where they break down significantly. In addition to basic understanding of the nature of amplification at California's seismic stations, EAFs may be important in site response mapping applications. For example, the use of EAFs for correcting recordings to rock conditions will improve the underlying rock reference layer of ShakeMap, which is currently estimated from recorded ground motions that are adjusted to rock conditions with V_{S30} -based corrections. Similar factors are subsequently used to modify the reference rock layer to account for site response (V_{S30} correction factors, where V_{S30} is estimated from slope or geology). Thus, the proposed research will improve both of these steps.

Introduction

The primary goal of this project is to develop new protocols/algorithms for mapping site response. For many purposes, such as ground motion prediction equations (GMPEs), building codes, and earthquake hazard mapping, site response is generally estimated through empirical correlations with the average shear-wave velocity to 30 m depth (V_{S30}). Though widely employed, V_{S30} is often described as being very limited in representing observed amplifications (e.g., Castellaro et al., 2008; Cadet et al., 2010; Régnier et al., 2014). Herein we analyze the uncertainty of V_{S30} as a predictor for site response for a large number of sites in California. Moreover, for mapping purposes a second proxy is needed just to estimate V_{S30} , such as surface geology (Wills and Clahan, 2006) or topographic gradient (Wald and Allen, 2007), or a combination of proxies (Thompson et al., 2014). One of the major challenges of this approach is the limited number of V_{S30} measurements and the limited distribution of the V_{S30} measurements across geologic units. This creates a potentially unnecessary “weak link” in statistical models of site response that can be eliminated by creating a model of site response amplification directly from geospatial variables (geology, topography, terrain). For example, in the Next Generation of Attenuation (NGA) project database, only 33% of the stations have

measured V_{S30} values, and most of these are in site class E or D (Chiou et al., 2008). The percentage of stations with measured V_{S30} values increased to 43% for the NGA-West 2 project (Seyhan et al., 2014). Many stations have recorded years of ground motion data that can be exploited to measure *actual* site amplifications. Thus, the total number of observations available to develop the geospatial model of site response can be significantly increased if we focus on empirical amplification factors (EAFs).

EAFs are computed as the ratio of the recorded motions to a common reference rock motion. In this project we define the rock motion with existing GMPEs (and an appropriate reference V_{S30}). Thus, the EAFs include all of the variability in the recorded motion that is not accounted for by the GMPE, such as site, basin, and topographic effects.

A key limitation of V_{S30} for mapping purposes is that very few V_{S30} measurements are collected in hard rock. An extreme example of this is the V_{S30} database for the Central and Eastern U.S. (CEUS); the Pacific Engineering and Analysis V_{S30} database contains only two measurements with V_{S30} greater than 1.5 km/sec (site class A) for the CEUS, though we expect that much of the region consists of site class A (Silva et al., 2011). There are good reasons for this sampling bias from the perspective of those who typically collect the data: more construction and geotechnical investigations take place in soil environments, drilling is more difficult in hard rock, access can be more difficult in rugged terrain, and little to no site response is expected at these hard rock sites. However, the lack of velocity measurements in hard rock presents a challenge when attempting to differentiate between rock and sediment response in hazard maps. While station locations also suffer from a similar bias, the bias is less severe than for the V_{S30} measurements because there are many broadband stations that have purposefully been sited on rock. These sampling limitations can be partly mitigated by working directly with EAFs rather than V_{S30} .

Tinsley and Fumal (1985) provided an influential early effort at mapping site response. They presented an index of site amplification that is primarily based on soil type, age, and the average V_S range of the geologic unit. Other efforts have built upon this method, and generally focus on correlations of V_{S30} with some other variable that is easily measured at the scale and resolution of interest. This includes correlations with surficial geology (Wills and Silva, 1998; Romero and Rix, 2001; Wills and Clahan, 2006), topographic slope (Wald and Allen, 2007; Allen and Wald, 2009), and geomorphologic terrain inferred from satellite imagery (Yong et al., 2012). Eliminating the link of V_{S30} between the geospatial proxy (geology, slope, terrain, or some combination) and site amplification is a promising alternative approach that we investigate in this report.

In contrast to previous studies that have focused on compiling databases of V_{S30} measurements, we will focus on compiling a database of EAFs. The benefits discussed above pertain to developing a site response model for mapping purposes. Thus the site response must be a function of geospatial parameters. However, an additional benefit of focusing on EAFs (i.e., the repeatable site effects inferred from ground motions) is to achieve the best possible estimate of the site response at the strong motion stations and to characterize where and why some V_{S30} approximations are sufficient. Site-specific EAFs at stations that have recorded numerous ground motions will provide a more accurate estimate of site response than approximations based on V_{S30} or correlations of EAFs with

geospatial parameters. This latter point is particularly useful for ShakeMap when correcting the recorded motions to a reference rock motion, which is used to interpolate between stations (Worden et al., 2010). This will improve the underlying rock reference layer of ShakeMap, which is currently estimated from recorded ground motions with V_{S30} -based site corrections. Similar V_{S30} -based factors are subsequently used to modify the reference rock layer to account for site response (where V_{S30} is estimated from slope or geology). Thus, the proposed research will not only improve our basic knowledge of the nature of site amplification at California's seismic stations and provide a reality check on the use of V_{S30} -based amplification factors, it will better constrain these two important steps in constructing ShakeMaps.

Data and Processing

The data that we analyze in this project was compiled and processed for the NGA-West 2 project (Ancheta et al., 2014). The database (or flatfile) includes recordings from around the globe, but we only use the records from California for this project. This database contains extensive information on the earthquake source, the conditions at the recording station, and the different distance measures that are required to evaluate GMPEs.

One important attribute of the ground motion data reported in the NGA-West 2 flatfile is the lowest usable frequency (f_{\min}), which is a function of the high-pass corner frequency used in processing the record. In this project, we only use records for response spectra at periods less than $1/f_{\min}$, which means that the number of records available to compute the EAFs decreases as period increases.

Next, we remove records that do not meet some general requirements. The Geomatrix classification is also reported in the flatfile. We use the first letter of the Geomatrix classification to remove records that may exhibit soil-structure-interaction. Additionally, the flatfile indicates if a record exhibited a late S-wave trigger so we remove records where this was noted. We only use records where the Joyner-Boore distance (R_{JB}) is less than 400 km and records where no concerns have been observed in the spectral quality (see Ancheta et al., 2014, for discussion of spectral quality flags).

We define the EAFs relative to the Boore et al. (2014) reference rock motion (i.e., $V_{S30} = 760$ m/sec). Thus, we can only use records where the event terms can be computed. Therefore, we only use records from events with four or more records that meet the previously described criteria. Since the number of records available for a given earthquake will vary by period, whether or not an earthquake can be used will be a function of period as well. A map of the earthquake epicenters, the number available records, and the data distribution in magnitude and distance space (as a function period) that remain after these criteria have been applied is summarized in Figure 1.

V_{S30} -Based Amplifications

Site response in GMPEs has traditionally been modeled as a function of V_{S30} and parameters to approximate basin depth, such as the depth at which the shear-wave velocity profile exceeds 1.0 km/sec (z_1) which has been added in recent years. Following the approach of Boore et al. (2014), site amplification is broken up into three terms as follows:

$$F_S = \ln(F_{lin}) + \ln(F_{nl}) + F_{\delta z_1}, \quad (1)$$

where F_{lin} is for the linear site effects, F_{nl} is for the nonlinear site effects, and $F_{\delta z_1}$ is for the effects of basin depth. Throughout most of this report, we will focus on F_{lin} because it is directly comparable to the EAFs. F_{lin} is a relatively simple function of V_{S30} (see Boore et al., 2014 for more details).

Computing the Empirical Amplification Functions

The first step for computing the EAFs is to compute the event terms. This is the average residual for each event, which will vary with period and must be subtracted from the data in order to isolate the site effects. We refer to the recorded ground motion intensity (peak acceleration, peak velocity, or response spectra) as y , which is a function of oscillator period T , but we leave this out of the equations for brevity. In order to compute the event terms, we must adjust the recorded data to a consistent reference site condition. Since we will use the Boore et al. (2014) GMPE, we adopt their definition of reference rock for this purpose, which is a V_{S30} of 760 m/sec. We refer to the $V_{S30}=760$ -adjusted recorded intensities as $y_r = y/F_S$. This leads to two definitions of residuals:

$$R = \ln(y) - \ln(\hat{y}), \quad (2)$$

and

$$R_r = \ln(y_r) - \ln(\hat{y}), \quad (3)$$

where \hat{y} is the intensity value predicted by the Boore et al. (2014) GMPE for $V_{S30} = 760$ m/sec. The event terms are then computed by separating the total residuals defined in equation 3 into inter-event and intra-event residuals using the mixed-effects framework:

$$(R_r)_{i,j} = c + \eta_i + \varepsilon_{i,j}, \quad (4)$$

where i is the earthquake event index, j is the record index, c is the mean residual, η_i is the mean residual for the i th event (i.e., event term, inter-event residual, or between-event residual), and $\varepsilon_{i,j}$ is the intra-event (i.e., within event) residual. We estimate the residuals in equation 3 using the mixed effects regression code in R (Pinheiro and Bates, 2000). The pertinent residual for defining the EAFs is to subtract the event term from the unadjusted residuals (equation 2):

$$(R_{ec})_{i,j} = R_{i,j} - \eta_i, \quad (5)$$

where the “ec” subscript indicates that this is the “event corrected” residual.

With multiple recordings at a site, the EAFs can be estimated as the repeatable component of $(R_{ec})_{i,j}$ at that site (e.g., Joyner and Boore, 1993; Lin et al., 2011; Rodriguez-Marek et al., 2011). Because these factors are estimated from recorded motions, they are not limited by the simplicity of the V_{S30} proxy or the assumptions in a 1D site response analysis. Thus, the EAFs are able to capture any of the repeatable site effects, including velocity structure that is deeper than 30 m and deviations from 1D

behavior such as horizontally propagating surface waves (Graves, 1993) and seismic scattering (Thompson et al., 2009).

EAFs, however, are still subject to epistemic uncertainty because they are estimated from the recorded motions. This epistemic uncertainty decreases as the number of recordings at each station increases. Therefore we only include stations in our analysis for which we can estimate the EAFs with at least five records. There are 1,536 stations with recordings that contribute to the data after the previously described screening criteria were applied (summarize in Figure 1). But there are only 483 stations with at least five recordings for which the Boore et al. (2014) PGA is less than 0.1 g. To summarize this smaller database, Figure 2 reproduces Figure 1 with only the data from these stations. Additionally, Figure 3 is a map of the station locations. Note that we have not checked the distributions of epicentral distance or backazimuth for the events at each station. Therefore, it is possible that some stations contain a bias to particular source-to-site geometries and the uncertainties may be underestimated in such cases. To summarize: these stations contain at least five recordings that fulfill the following criteria:

1. The Boore et al. (2014) PGA for $V_{S30} = 760 \text{ m/sec} < 0.1 \text{ g}$
2. Each recording must be from an earthquake with at least five responding stations that can be used to estimate the event term (η_i)
3. Each of the recordings in #2 must fulfill the initial screening criteria discussed previous (e.g., f_{min} , $R_{JB} < 400 \text{ km}$, etc.)

The final step is to compute the median and 95% confidence interval (CI) for the EAFs at each station. We refer to the mean $(R_{ec})_{i,j}$ at a station as $\overline{R_{ec}}$, and thus the median amplification is $\exp(\overline{R_{ec}})$. Given the large number of sites, these plots are provided in Appendix 1 and are sorted alphabetically by station name. The plots in Appendix 1 also compare the EAFs to the linear component of the V_{S30} -based amplifications (F_{lin}), where the V_{S30} is given by the correlations with surface geology by Wills and Clahan (2006). The V_{S30} value is given in the lower-left corner of each plot. Perusing Appendix 1, it is easy to find examples of stations where the EAFs are in close agreement with F_{lin} , such as Mirage and Desert View Tower; but there are also many stations where the EAFs give substantially different results than F_{lin} , such as McLaughlin Mine and Rock Hill.

Empirical Amplification Functions by Geologic Unit

The EAF median and 95% CI for each available station is given in Appendix 1, which are valuable for correcting the recorded motions to a consistent reference rock condition. However, for mapping purposes, we need spatial proxies, such as surface geology, to predict the EAFs at any arbitrary location. Therefore, it is useful to look at the median and 95% CI of the EAFs within the different surficial geologic units that Wills and Clahan (2006) used to map V_{S30} . Additionally, it will be of interest to compare the median EAFs for each geologic unit with the F_{lin} computed for the median V_{S30} reported by Wills and Clahan (2006). This comparison, provided in Figure 4, will provide an assessment of how consistent the EAF method developed in this report are with the approach of computing F_{lin} from mapped V_{S30} values. The geology-based F_{lin} and the EAFs by geologic unit are surprisingly similar, given the differences in how they are

computed. After observing how different the EAFs and station-specific F_{lin} functions can be in Appendix 1, it is reassuring to see the consistency in Figure 4. However, this also indicates that the EAF approach to mapping site response will not substantially reduce the uncertainty with respect to F_{lin} computed from geology-based V_{S30} values in site response maps for locations where the inferred EAF relies solely on surface geology.

The largest differences between the geology-based F_{lin} and the geologic median EAFs are in the Tv unit (Tertiary volcanics) where the V_{S30} reported by Wills and Clahan (2006) for this unit is 609 m/sec; this V_{S30} value results in negligible F_{lin} amplifications. In contrast, the EAFs indicate that there is substantial amplification in this unit (approximately a factor of three across most periods). The V_{S30} of this unit is relatively uncertain since it is based on only 3 measurements (Wills and Clahan, 2006.). The EAFs in this unit are based on 155 recordings from six stations. This summary information for each geologic unit is provided in Table 1. Note that Qal- fine includes only one station, which is concerning. However, the EAFs agree relatively well with the geology F_{lin} amplifications. Also note that there are some units for which there are many more V_S profiles than recording stations (e.g., Qoa) and the opposite is true for other units (e.g., xtaline).

An important issue is that two of the geologic units do not contain any stations in the processed database. This includes Qi and Qal- thin West LA. Given the general consistency between the two approaches, it appears that it would be a reasonable approach to use the geology-based F_{lin} amplifications in these units alongside the EAFs in other units.

Correlations with Topographic Slope

Following the hybrid strategy of mapping V_{S30} by Thompson et al. (2014), we now look for trends of the EAFs with topographic slope. For EAFs, this is a little more complicated than for V_{S30} because the trends will vary with spectral period. Thus, we need to investigate the trends at each period of interest. For brevity, we only show the results for PGA (Figure 5), PGV (Figure 6), $T=0.3$ sec (Figure 7), $T=1.0$ sec (Figure 8), and $T=3.0$ sec (Figure 9) in this report. These figures plot the logarithmic residuals defined as

$$R_{geo} = R_{ec} - R_{ec}^{geo}, \quad (6)$$

where R_{ec}^{geo} is the mean R_{ec} within the geologic unit, computed as:

$$R_{ec}^{geo} = \frac{1}{n_k} \sum_{i=1}^{n_k} (R_{ec})_k, \quad (7)$$

where k is the geologic unit index, and n_k is the number of records in geologic unit k . Since R_{ec} is in logarithmic units, $\exp(R_{ec})$ is the median amplification for each unit. In Figures 5 through 9, we are looking for trends about the median amplification for each geologic unit that correlate with the topographic slope. We use those trends with slope only when the regression yields a slope coefficient that is significantly different than zero at the 95% significance level (p-value less than 0.05). When this trend with slope is

significant, we plot the regression line and its 95% confidence interval in Figures 5 through 9.

Some of the trends identified in Figures 5 through 9 appear spurious, for example, those within *xtaline*, *KJf*, and *Tv*. Thus, we do not use those trends to construct our maps. Additionally, we decided to combine the two different Qal- deep units. Table 2 summarizes whether or not the trends with slope are significant at the 95% confidence level. It is easy to see that Qal- thin and Qal- deep are significant at all periods, while Qal- coarse is only significant at longer periods and PGV, and Qoa is only significant at $T=3.0$ sec. It makes sense that the trends with slope would be stronger at longer periods because this is where site response tends to have the greatest effect. This is illustrated by the value of the c coefficients reported by Boore et al. (2014) in the equations for computing F_{lin} . Thus, even though V_{S30} may be correlated with slope in some of these units, the correlation with the EAF may not be present due to the weaker correlation between V_{S30} and site response at shorter periods.

We now define A_{hyb} as the hybrid geology-slope EAF model that employs the period-specific median EAF for each geologic unit (i.e., $\exp(R_{ec}^{geo})$) illustrated in Figure 4) and also the trends with slope summarized in Table 2.

Spatial Correlation Structure

Next, we inspect the spatial correlation structure of the A_{hyb} model residuals, defined as

$$R_{hyb} = R_{ec} - \ln(A_{hyb}). \quad (8)$$

While it would be ideal to allow the spatial correlation structure to vary with geologic unit in the same manner as we allow the trends with slope to vary with geologic unit, we do not have sufficient data to constrain the correlation parameters for each unit. Thus, we bin the geologic units and inspect the correlation structure within the different bins. We break up the observations into three categories: Holocene sediments, Pleistocene to Tertiary sediments, and rock (Tertiary and older). The sample and model semivariograms are given for all the data and these three subsets for PGA, PGV, $T=0.3$ sec, $T=1.0$ sec, and $T=3.0$ sec in Figures 10-14. From these plots, we can see that the Holocene units are spatially correlated at all periods, while the rock units are spatially correlated at all periods except for $T=3.0$ sec, and the Pleistocene units are not correlated at any period. It is worth noting that there are 174 Holocene stations, 95 Pleistocene stations, and 214 rock stations. Thus, the lack of correlation could be a result of the relatively few measurements in the Pleistocene units. We refer to the estimate of the EAFs that kriges the residuals with the models summarized in Figures 10-14 along with the hybrid geology-slope trend model as A_{lin} , and indicate the spectral period in the superscript (e.g., the PGA A_{lin} is A_{lin}^{PGA} and the $T=1.0$ sec A_{lin} is $A_{lin}^{T=1.0}$).

Nonlinearity

In the previous sections, we have developed the components that are necessary to bypass V_{S30} and map EAFs directly, following the regression kriging approach that Thompson et al. (2014) employed for mapping V_{S30} . However, the EAFs only represent the linear

component F_{lin} of the site amplification F_S (equation 1). Existing models of F_{nl} (e.g., as described by Boore et al., 2014) are a function of V_{S30} and PGA_r (PGA on rock). There are at least three options for handling the nonlinearity within the EAF mapping context:

1. Use maps of V_{S30} to compute F_{nl}
2. Compute V_{S30} from correlations with EAFs
3. Develop new F_{nl} functions from the available database

There are two benefits to using approach #1: 1) the F_{nl} relationships based on V_{S30} have received the most attention and are therefore the well-vetted by the geotechnical community, and 2) no new relationships need to be developed based on the our current database, which is more limited in the high strain range and does not include simulated records (such as those by Kamai et al., 2014). The primary problem with approach #1 is that of consistency between F_{nl} and F_{lin} . If one were to use the EAFs for F_{lin} at a site for which these two functions strongly disagree (e.g., the China Lake site in Appendix 1) then the nonlinearity factors would be inconsistent with the linear amplification factors. At China Lake, the mapped V_{S30} value 280 m/sec but the EAFs indicate that the site deamplifies the motion at all but the longest periods. Thus, it would be unconservative to apply F_{nl} functions based on a V_{S30} of 280 m/sec, which would serve to additionally deamplify the motion as PGA_r increases.

The primary benefits of approach #2 are that it directly addresses the concerns about consistency in approach #1 and it is also a relatively simple model to develop and implement. This approach relies the correlations between V_{S30} and the EAFs, so Figure 16 gives the scatter plot and correlation coefficients of V_{S30} with EAFs for the stations with measured V_{S30} values in the NGA-West 2 database. Note that we do not show the correlation with the PGA EAFs because it exhibited the lowest correlation coefficient (-0.50). Overall, these correlations are relatively large, and the best is for PGV. Stepwise linear regression led to the following model for computing V_{S30}

$$\ln(V_{S30}) = \beta_0 + \beta_1 \ln(A_{lin}^{PGV}) + \beta_2 \ln(A_{lin}^{T=3.0}) + \beta_3 \ln(A_{lin}^{PGV}) \times \ln(A_{lin}^{T=3.0}), \quad (9)$$

where the coefficients and their summary statistics are reported in Table 3. As Table 3 indicates, the interaction term is warranted and we illustrate the effect of this term in Figure 17, which plots V_{S30} as a function of A_{lin}^{PGV} and $A_{lin}^{T=3.0}$. Without the interaction term, the lines of constant V_{S30} would be straight lines. The interaction term allows the V_{S30} to be low if only one of the EAFs is high. The main concern with this approach is that the V_{S30} predicted through equation (9) may not be as accurate as maps of V_{S30} such as that provided by Wills and Clahan (2006) or Thompson et al. (2014).

For approach #3, the challenge is to build a new function that does not rely on an estimate of V_{S30} . Instead, we could substitute one or more of the A_{lin} values for V_{S30} . We can keep the same F_{nl} function as Boore et al. (2014)

$$\ln(F_{nl}) = f_2 \ln\left(\frac{PGA_r + 0.1}{0.1}\right). \quad (10)$$

We then estimate the coefficient f_2 , which controls the degree of nonlinearity as V_{S30} varies, by fitting a linear regression to the data for a range of A_{lin}^{PGV} (analogous to a V_{S30} bin). We followed this approach with the available dataset but we were not able to

resolve trends that are consistent with our understanding of nonlinearity. Thus, we cannot provide a model based on this approach at this time.

Effects of Basin Depth Parameters

For mapping purposes, the basin depth parameters are rarely known with sufficient precision or spatial coverage to be useful. However, the EAF database that we have developed for this project is useful for assessing the impact of the basin depth parameters.

Many of the sites in Appendix 1 for which the EAFs do not match the F_{lin} amplifications are characterized by large amplifications at short periods ($T < 0.5$ sec) and moderate to no amplification at mid-to-long periods. Examples include:

- Angel Island,
- Barrett,
- Ben Lomond Mountain,
- Capra Ranch,
- Hamilton Field,
- Mill Valley MUNI Golf Course,
- Mount Johnson,
- Park Hill,
- San Diego Road Dept.,
- Tamalpais Peak C, and others.

The interesting thing about these sites is that the amplification shape cannot be matched by the F_{lin} function for any V_{S30} value. In contrast, other sites that have a peak in the amplification at mid-to-long periods ($T > 0.5$ sec) can easily be modeled by F_{lin} , and so the inaccuracy of F_{lin} can be attributed to uncertainties in the mapped V_{S30} value. Examples of this type of site include:

- Big Mountain,
- La Fresa,
- Lone Juniper Ranch, and others.

The former type of site (amplifications are high at short periods) are typically stiff sites, and the high V_{S30} underpredicts the amplifications.

We analyze the Barrett site as an example because it has recently been profiled (Yong et al., 2013). Presumably V_{S30} is unable to model the site response here because the layer of sediment is shallow and thus V_{S30} can not distinguish between it and rock sites that do not include a shallow layer of slower materials. Thus, there is potential for the basin-depth parameters to improve the fit of F_{lin} . Figure 18 (left panel) compares F_5 to the EAF at this site for various value of z_1 to illustrate the sensitivity of the amplifications to this parameter. The value of z_1 based on the profile reported by Yong et al. (2013) is 25 m. This shows that the depth parameter does improve the fit modestly at mid-to-long periods ($T > 0.65$ sec) by lowering the amplification. However, it does not increase the amplification at shorter periods (this is repressed by the functional form). Given that this site is located in the mountains and near a dam, there is the potential for topographic and/or soil-structure-interaction effects. So it is interesting to check if the full resonance one-dimensional plane S-wave (SH1D) calculation with the full profile

measured at the site is able to capture the EAF at this site. This comparison is given in Figure 18 (right panel). Note that the SH1D amplification is for Fourier spectra, and so it is not exactly correct to compare to the EAF amplifications, which are from spectral accelerations. However, we can still see that the SH1D model does a good job of capturing the EAFs for most periods (about 0.02-10 sec). Thus, topographic and soil-structure-interaction cannot explain the misfit of F_S at this site.

We now look in more detail at site Saticoy as an example of a site with large long period amplifications that are not well modeled by F_{lin} . First, it is important to check F_{lin} for the measured V_{S30} value of 249 m/sec (again from Yong et al., 2013), which is given in Figure 19. With the default z_1 (i.e., “ z_1 off”), this V_{S30} value improves the fit at long periods (though it is still underpredicting the amplifications), but worsens the fit at short period (where it is overpredicting the amplifications). Based on the shape of the EAF at this site, it is evident that it is a deep soil site, and as such, the profile reported by Yong et al. (2013) does not reach z_1 . The depth to the bottom layer is 68 m and the V_S of the bottom layer is 584 m/sec. However, the NGA-West 2 flatfile has compiled z_1 values from the Southern California Earthquake Center (SCEC) community 3D velocity models. This includes the Harvard model, termed CVM-H (Tape et al., 2009), and the SCEC model, termed CVM-S (Magistrale et al., 2000). At this location, CVM-H gives $z_1=940$ m and CVM-s gives $z_1=740$ m. We add the F_S curves to Figure 19 for these two values of z_1 along with $z_1=70$ m. The reasons for including $z_1=70$ m are: 1) to visualize the sensitivity of the parameter, and 2) this value is just below the maximum depth of the profile, and so it might be a reasonable assumption in the absence of either SCEC CVM estimates. For $z_1=70$ m, the fit of the F_S curve is worsened by lowering the amplifications at longer periods. However, both CVM z_1 values improve the fit, and the deeper z_1 from CVM-H fits the data better. Again, due to the formulation of z_1 it cannot affect the shorter periods.

Uncertainty of V_{S30} -based Amplifications

This dataset presents us with the opportunity to assess the uncertainty in V_{S30} -based amplifications. First, it is prudent to compare the standard deviations with the database as processed for this report with the standard deviations reported by Boore et al. (2014). Within the mixed-effects framework of equation 4, the total standard deviation (σ) is a function of τ , defined as the standard deviation of η_i and referred to as the inter-event variability, and ϕ , defined as the standard deviation of $\varepsilon_{i,j}$ and referred to as the between-event variability:

$$\sigma = \sqrt{\tau^2 + \phi^2} . \quad (11)$$

Figure 20 compares the values for these standard deviations that we compute with those reported by Boore et al. (2014). It is clear that the standard deviations from our data are larger, particularly for shorter periods, and this is likely because we have been less restrictive in our processing of the data. For example, we only require five events within 400 km of the source whereas Boore et al. (2014) require four events within 80 km. While the more strict processing clearly leads to smaller standard deviations, it comes at the cost of removing many events from the database that allow us to estimate the EAFs at many more sites than if we used the more strict data criteria.

The component of the total standard deviation that is pertinent to site response is ϕ . The model of ϕ provided by Boore et al. (2014) includes variations with magnitude, distance, and V_{S30} . We are also interested in how ϕ varies with how the V_{S30} is estimated, either through correlations with geology, slope, or direct measurements. To address this we subset our data to only those stations with measured V_{S30} values. Of the 483 sites for which we have computed the EAFs, 163 of those sites have measured V_{S30} values. We then estimate ϕ in a number of different ways:

- We define ϕ_X as the estimate of ϕ for which no site term has been applied. Thus, it is computed as the sample standard deviation of $(R_{ec})_{i,j}$.
- We define ϕ_G as the estimate of ϕ for which F_{lin} is computed from Wills and Clahan (2006) V_{S30} (V_{S30}^{geo}). Thus it is computed as the sample standard deviation $(R_{ec})_{i,j}/F_{lin}(V_{S30}^{geo})$.
- We define ϕ_M as the estimate of ϕ for which F_{lin} is computed from the measured V_{S30} (V_{S30}^{mea}). Thus it is computed as the sample standard deviation $(R_{ec})_{i,j}/F_{lin}(V_{S30}^{mea})$.
- We define ϕ_E as the estimate of ϕ for which site response is computed from the EAF. Thus it is computed as the sample standard deviation $(R_{ec})_{i,j} - \overline{R_{ec}}$.

These four different standard deviation estimates are compared in Figure 21. As expected, $\phi_X > \phi_G > \phi_M > \phi_E$. The largest reduction is between ϕ_X and ϕ_G , and it is interesting to note that the ϕ_M is not that much smaller than ϕ_G . ϕ_E is less than ϕ_M , but this difference becomes negligible at periods of about 2 sec or longer. The larger impact at shorter periods is consistent with our earlier observation that the largest discrepancies between the ETFs and F_{lin} are for stiff sites where the short period amplifications cannot be reproduced with the existing F_{lin} functions.

Conclusions

The primary goal of this research is to develop and assess a method for mapping site response directly from EAFs. The method we have followed maps the EAFs at each period separately, following the regression kriging method of Thompson et al. (2014) where the trend EAF at a given period may be a function of geology or both geology and slope. The correlation structure of this trend varies with period and also with geologic unit. Comparisons of the linear site response amplifications predicted by this method and the amplifications from currently the employed methods (based on representative V_{S30} values by geologic unit) show that the two different methods give relatively consistent results. However, at a specific station, there are many sites where the two methods disagree significantly and it is at these stations where the EAF approach will provide the greatest improvement in accuracy. Nonlinearity can be handled by applying nonlinear amplification factors based on maps of V_{S30} , but we prefer the use of an EAF-consistent V_{S30} that is computed from the amplification for PGV and the 3 sec spectral acceleration.

The database of EAFs compiled for the purpose of mapping has also provided the opportunity to analyze the basin depth parameters. A detailed analysis of a single site demonstrates that the basin depth parameters can substantially improve the fit of the site response model for a deep soil site when paired with depth parameters from the SCEC-

CVM. However, the depth parameters are unable to the accuracy of the amplifications for stiff shallow soil sites where the amplifications at short periods are underpredicted by V_{S30} -based amplification models.

References

- Allen, T.I., and Wald, D.J. (2009). On the Use of High-Resolution Topographic Data as a Proxy for Seismic Site Conditions (Vs30). *Bull. Seism. Soc. Am.* 99, 935-943.
- Ancheta, T.D., R.B. Darragh, J.P. Stewart, E. Seyhan, W.J. Silva, B.S.J. Chiou, K.E. Wooddell, R.W. Graves, A.R. Kottke, D.M. Boore, T. Kishida, and J.L. Donahue (2014). NGA-West2 database, *Earthquake Spectra* 30, 989-1005.
- Boore, D.M., J.P. Stewart, E. Seyhan, and G.M. Atkinson (2014). NGA-West 2 equations for predicting PGA, PGV, and 5%-damped PSA for shallow crustal earthquakes, *Earthquake Spectra* 30, 1057-1085.
- Castellaro, S., F. Mulargia, and P. L. Rossi (2008). VS30: Proxy for seismic amplification?, *Seism. Res. Lett.* 79, 540–543.
- Chiou, B., Darragh, R., Gregor, N., and Silva, W. (2008). NGA Project Strong-Motion Database, *Earthquake Spectra* 24, 23-44.
- Graves, R.W. (1993). Modeling 3-dimensional site response effects in the Marina District basin, San Francisco, California, *Bull. Seism. Soc. Am.* 83, 1042-1063.
- Joyner, W.B., and Boore, D.M. (1993). Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am.* 83, 469-487.
- Kamai, R., Abrahamson, N. A., and Silva, W. J. (2014). Nonlinear horizontal site amplification for constraining the NGA-West2 GMPEs, *Earthquake Spectra* 30, 1223–1240.
- Lin, P.-S., Chiou, B., Abrahamson, N., Walling, M., Lee, C.-T., and Cheng, C.-T. (2011). Repeatable source, site, and path effects on the standard deviation for ground-motion prediction, *Bull. Seismol. Soc. Am.* 101, 2281-2295.
- Magistrale, H., S. Day, R. W. Clayton, and R. Graves (2000). The SCEC Southern California Reference Three-Dimensional Seismic Velocity Model Version 2, *Bull. Seismol. Soc. Am.* 90, S65–S76
- Pinheiro, J. C., and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*, Springer-Verlag, New York, 528 pp.
- Régnier, J., L. F. Bonilla, E. Bertrand, and J.-F. Semblat (2014). Influence of the VS Profiles beyond 30 m Depth on Linear Site Effects: Assessment from the KiK-net Data, *Bull. Seismol. Soc. Am.* 104, 2337-2348.
- Rodriguez-Marek, A., Montalva, G.A., Cotton, F., and Bonilla, F. (2011). Analysis of Single-Station Standard Deviation Using the KiK-net Data, *Bull. Seismol. Soc. Am.* 101, 1242–1258.
- Romero, S., and Rix, G.J. (2001). Regional Variations in Near-Surface Shear Wave Velocity in the Greater Memphis Area. *Engineering Geology* 62, 137-158.
- Seyhan, E., J. P. Stewart, T. D. Ancheta, R. B. Darragh, and R. W. Graves (2014). NGA-West 2 site database, *Earthquake Spectra* 30, 1007-1024.
- Silva, W., Thompson, E.M., Magistrale, H., and Wills, C. (2011). Development of a Vs(30m) map for central and eastern North America (CENA): median estimates and uncertainties. Report to PEER, 60 pp.
- Tape, C., Q. Liu, A. Maggi, and J. Tromp (2009). Adjoint Tomography of the Southern California Crust, *Science* 325, 988–992.
- Tinsley, J.C., and Fumal, T.E. (1985). In: Ziony, J.I. (Ed.), Mapping quaternary sedimentary deposits for areal variations in shaking response: Evaluating

- Earthquake Hazards in the Los Angeles Region-An Earth Science Perspective, 101-125.
- Thompson, E.M., Baise, L.G., Kayen, R.E., and Guzina, B.B. (2009). Impediments to predicting site response: Seismic property estimation and modeling simplifications, *Bull. Seism. Soc. Am.* 99, 2927-2949.
- Thompson, E. M., D. J. Wald, and C. B. Worden (2014). A Vs30 map for California with geologic and topographic constraints *Bull. Seism. Soc. Am.* 104, 2313-2321.
- Wald, D.J., and Allen, T.I., (2007). Topographic Slope as a Proxy for Seismic Site Conditions and Amplification, *Bull. Seism. Soc. Am.* 97 (5), 1379–1395.
- Wills, C.J. and Clahan, K.B. (2006). Developing a Map of Geologically Defined Site-Condition Categories for California. *Bull. Seism. Soc. Am.* 96, 1483-1501.
- Wills, C.J., and Silva, W. (1998). Shear-wave velocity characteristics of geologic units in California, *Earthquake Spectra* 14, 533-556.
- Worden, C.B., Wald, D.J., Allen, T.I., Lin, K., Garcia, D., and Cua, G. (2010). A revised ground-motion and intensity interpolation scheme for ShakeMap, *Bull. Seism. Soc. Am.* 100, 3083-3096.
- Yong, A., Hough, S.E., Iwahashi, J., and Braverman, A. (2012). A Terrain-Based Site-Conditions Map of California with Implications for the Contiguous United States. *Bull. Seism. Soc. Am.* 102, 114-128.
- Yong, A., Martin, A., Stokoe, K., and Diehl, J. (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, 60 p.

Tables

Table 1. Summary of V_{S30} and EAF data by geology unit.

Unit	V_{S30}		EAFs	
	Mean (m/sec)	No. Profiles	No. Stations	No. Recordings
af/Qi	217	44	5	96
KJf	782	32	33	935
Kss	566	6	10	288
Qal- coarse	354	18	43	1023
Qal- deep (incl LA)	280	225	77	1264
Qal- deep- Imperial V.	209	53	9	174
Qal- fine	236	13	1	11
Qal- thin	349	65	39	889
Qoa	387	132	73	1521
Qs	302	15	8	144
QT	455	18	14	313
serpentine	653	6	5	199
Tsh	390	55	39	990
Tss	515	24	26	793
Tv	609	3	6	155
xtaline	748	28	95	3535

Table 2. Solid dot indicates that the regression slope coefficient is significant and is used in the final model; cross indicates that the coefficient is significant but we decided not to include the trend in the final model.

Unit	PGA	PGV	T=0.3 sec	T=1.0 sec	T=3.0 sec
af/Qi					
KJf		×			
Kss					
Qal- coarse		•		•	•
Qal- deep (incl LA)	•	•	•	•	•
Qal- deep- Imperial V.					
Qal- fine					
Qal- thin	•	•	•	•	•
Qoa					•
Qs					
QT					
serpentine					
Tsh					
Tss					
Tv					×
xtaline	×				

Table 3. Regression coefficients for equation 7.

Coefficient	Estimate	Standard Error	P-value
β_0	6.5425	0.0302	$< 2 \times 10^{-16}$
β_1	-0.5888	0.0764	2.8×10^{-12}
β_2	-0.5027	0.0589	3.0×10^{-14}
β_3	0.3269	0.0617	4.9×10^{-7}

Figures

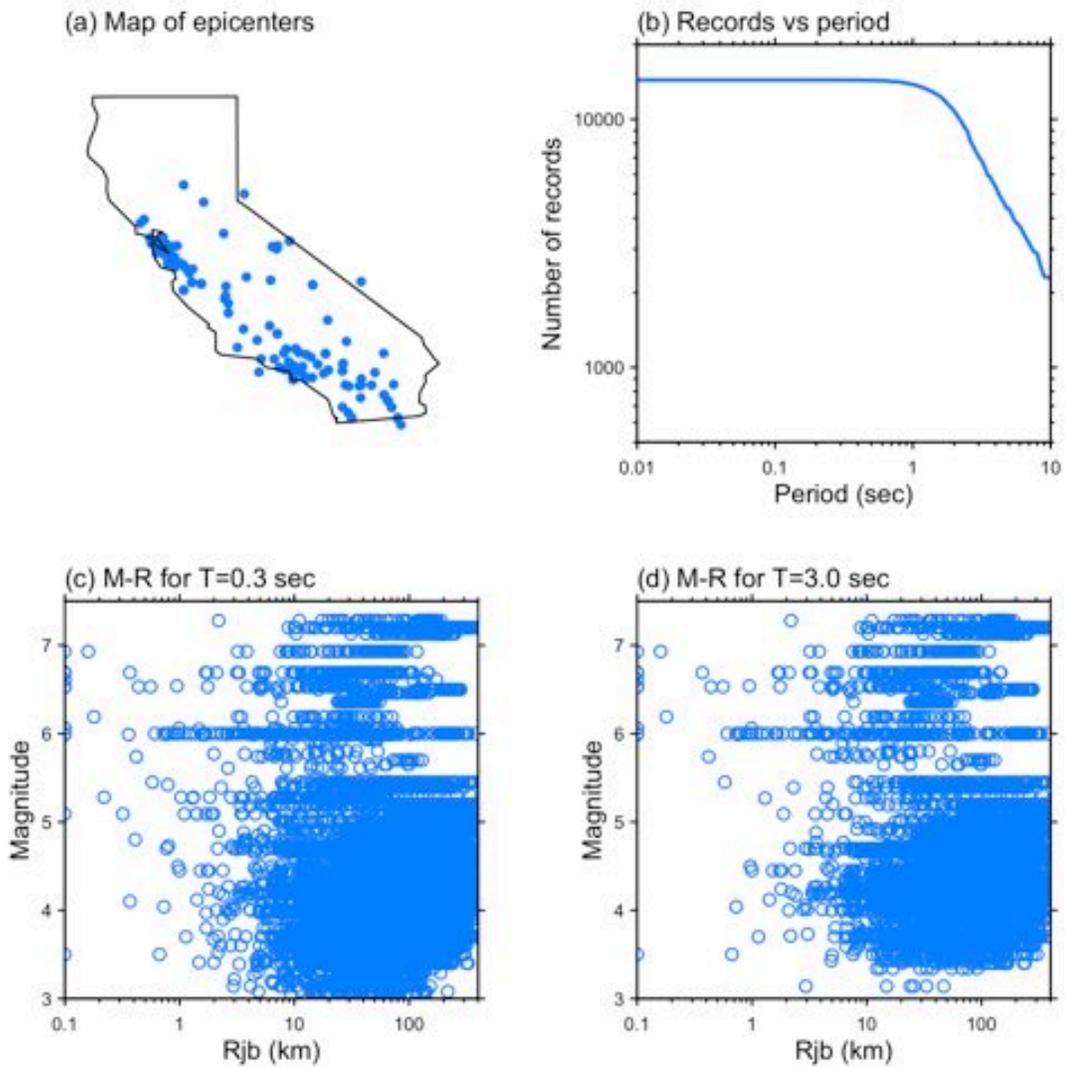


Figure 1. Summary of the data included in our analysis after the initial data screening has been applied to the NGA-West 2 database. (a) Map of the earthquake epicenters, (b) the number of records as a function of period, (c) the magnitude-distance (M-R) distribution of the records available at 0.3 sec, and (d) the M-R distribution at 3 sec. Note that points with $R_{JB} < 0.1$ are displayed at $R_{JB} = 0.1$ km.

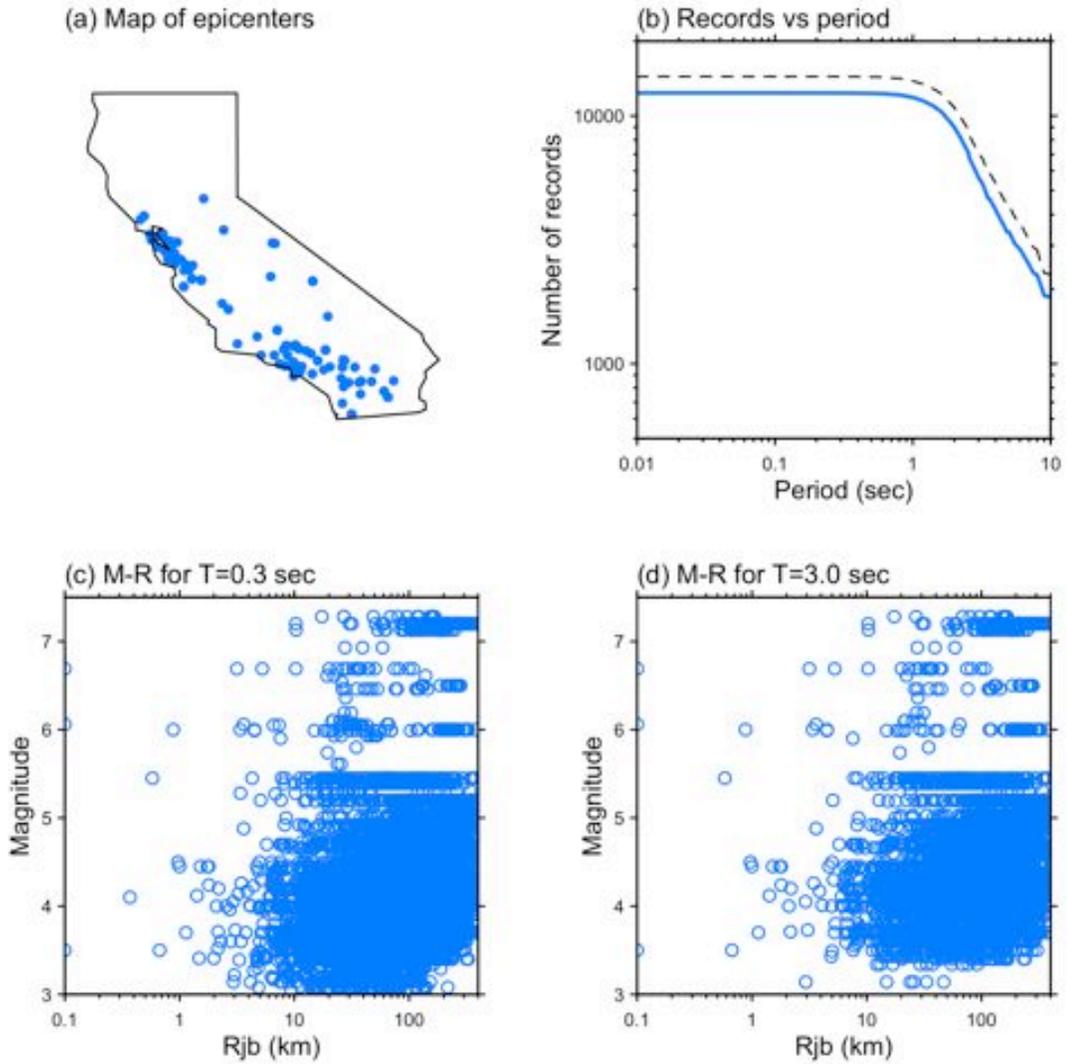


Figure 2. Same as Figure 1 except that the data has additionally been screened to only include stations with five or more recordings. For reference, the number of records from Figure 1 is given as the dashed line in (b).

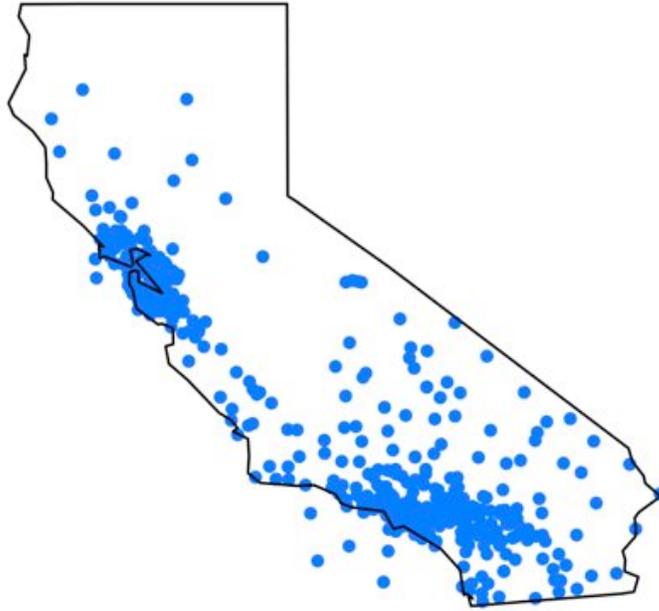


Figure 3. Stations where at least five linear recordings are available from earthquakes for which at least five recordings are available to compute the event terms, after the initial data screening criteria described in this report has been applied.

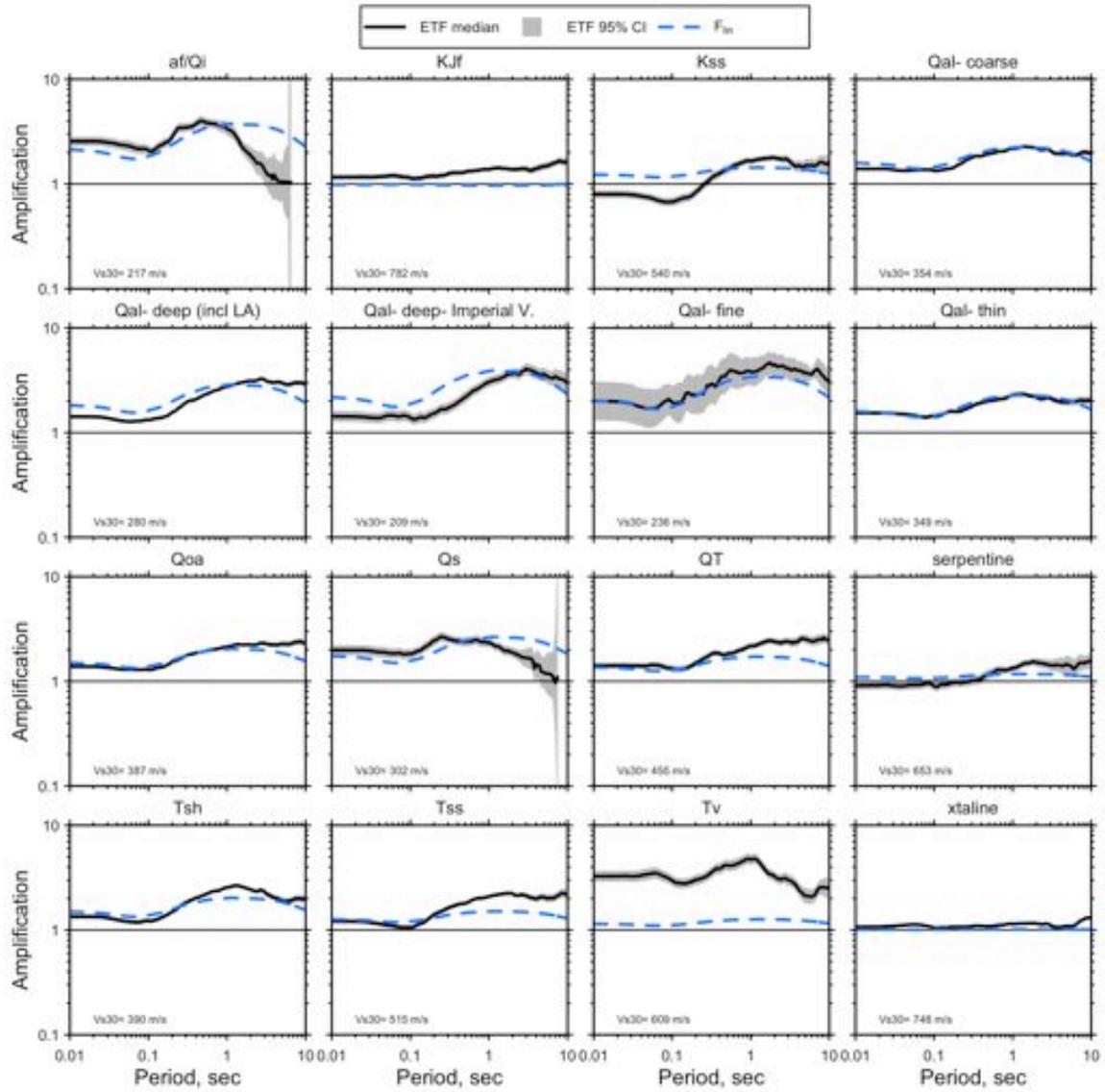


Figure 4. Comparison of site response amplifications by geologic unit. The median and 95% CI are compared to the F_{lin} computed for the median V_{S30} reported by Wills and Clahan (2006).

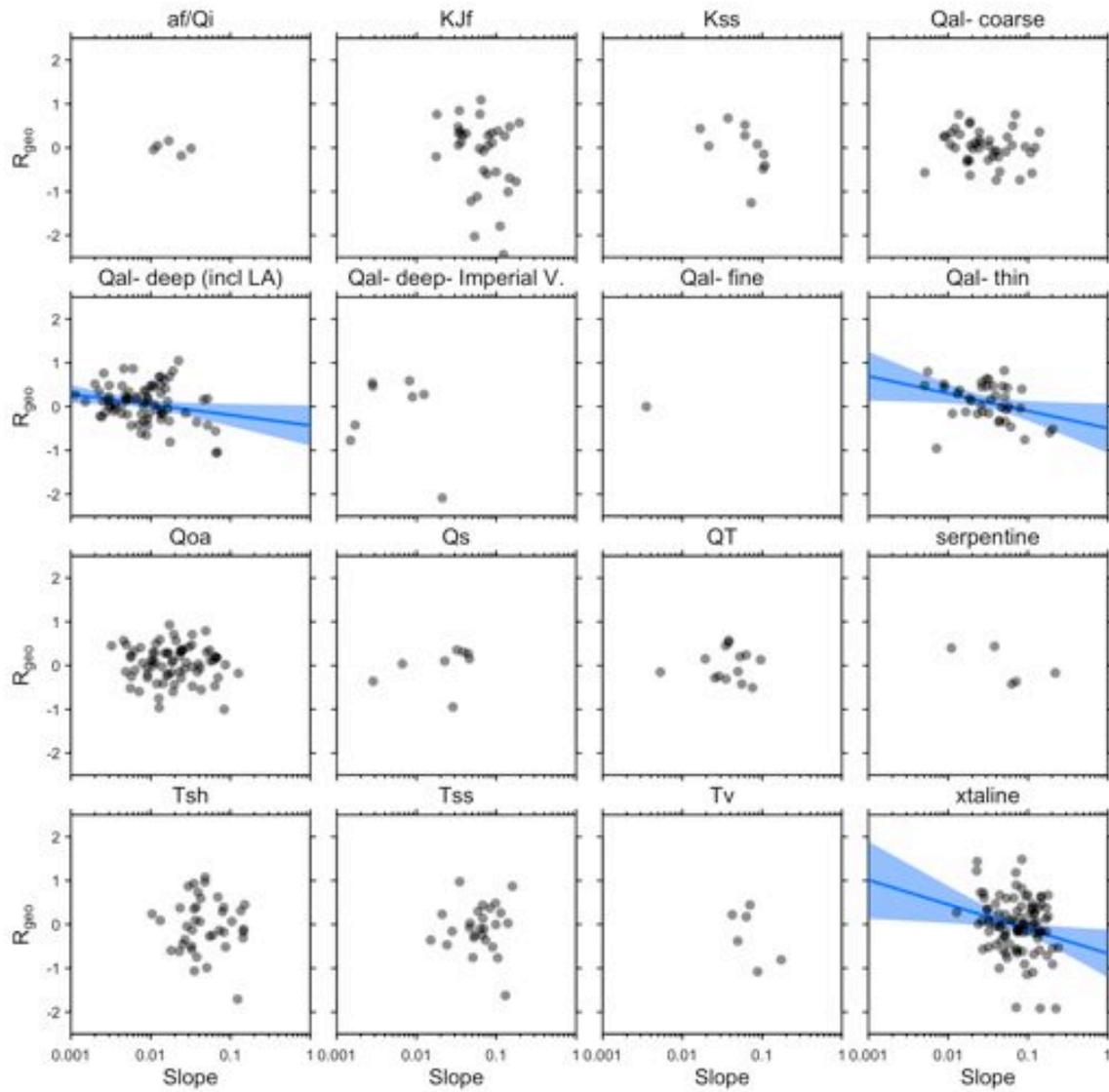


Figure 5. Plots of PGA R_{geo} vs slope for the different geologic units. Regression curves and their 95% confidence intervals are shown for units where the slope coefficient is significantly different than zero at the 95% significance level.

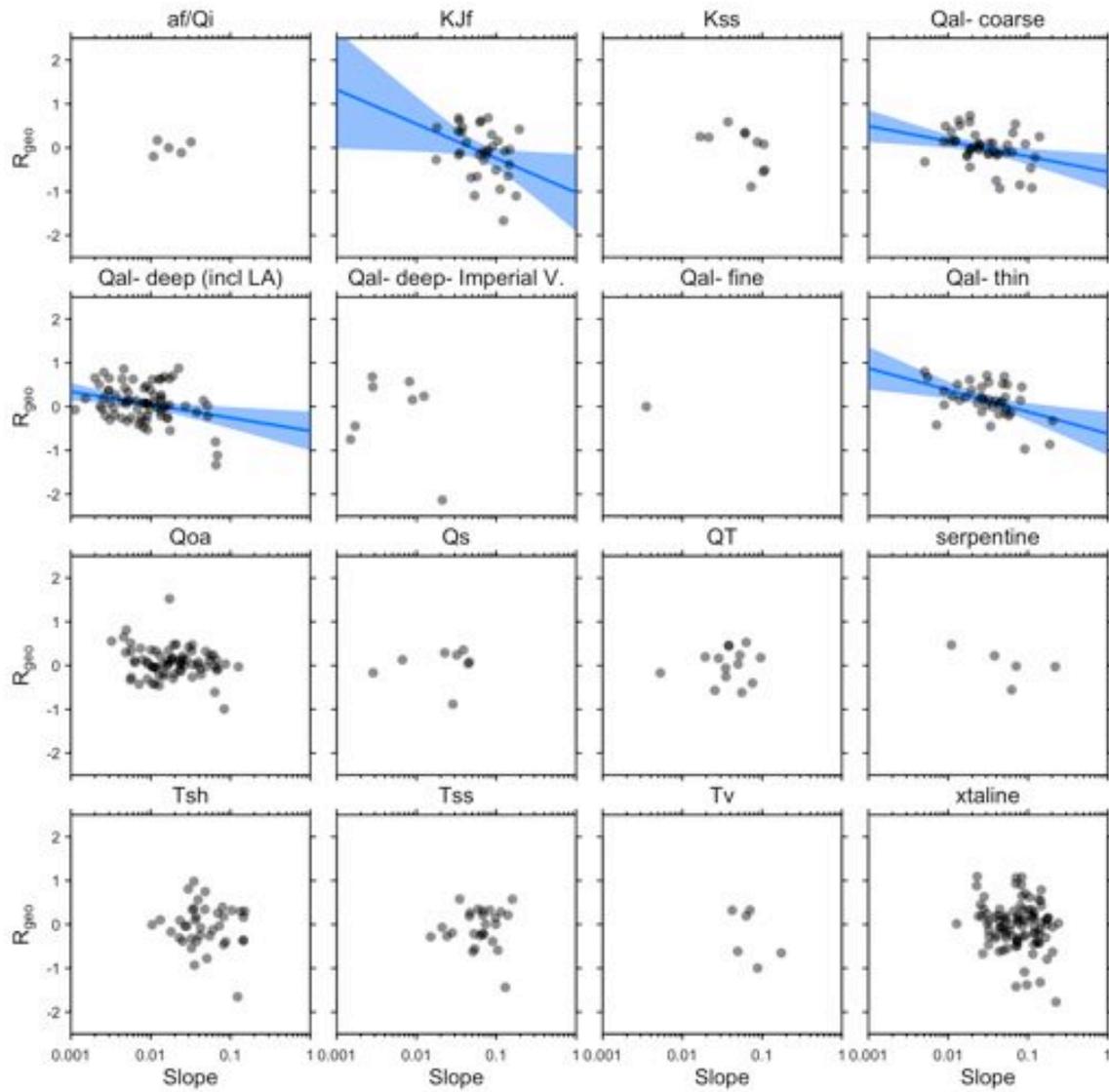


Figure 6. Plots of PGV R_{geo} vs slope for the different geologic units. Regression curves and their 95% confidence intervals are shown for units where the slope coefficient is significantly different than zero at the 95% significance level.

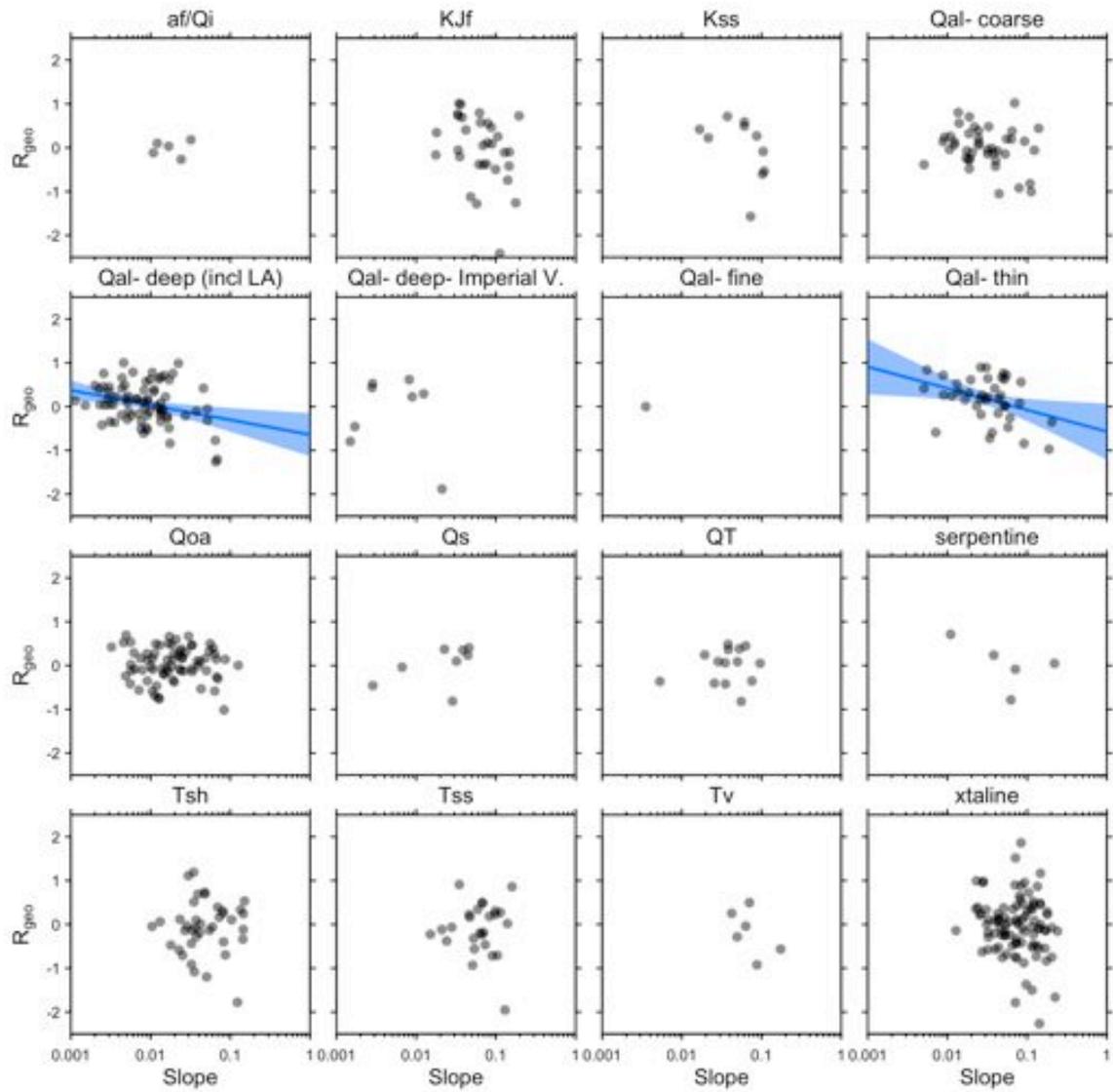


Figure 7. Plots of $T=0.3$ sec R_{geo} vs slope for the different geologic units. Regression curves and their 95% confidence intervals are shown for units where the slope coefficient is significantly different than zero at the 95% significance level.

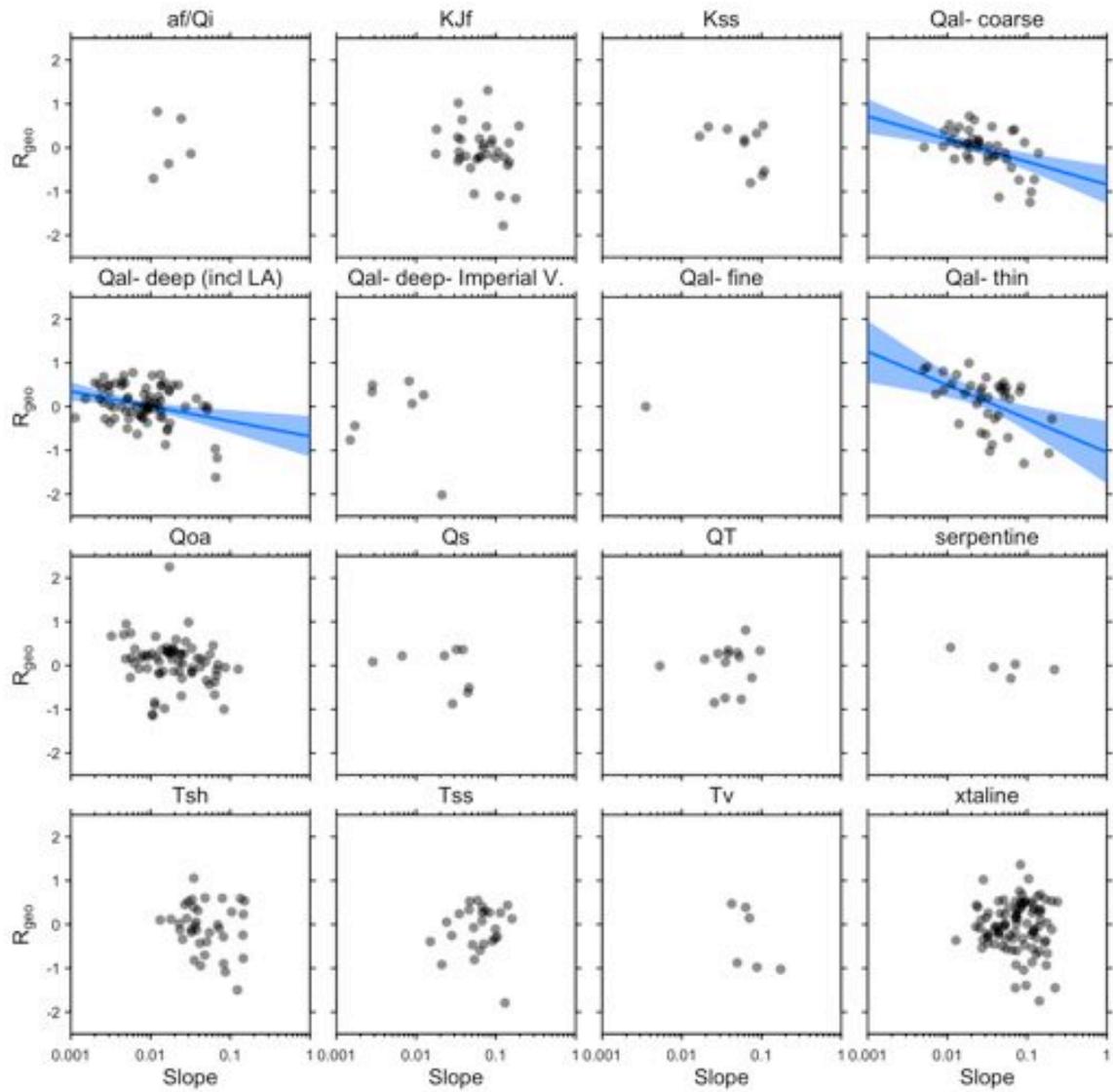


Figure 8. Plots of $T=1.0$ sec R_{geo} vs slope for the different geologic units. Regression curves and their 95% confidence intervals are shown for units where the slope coefficient is significantly different than zero at the 95% significance level.

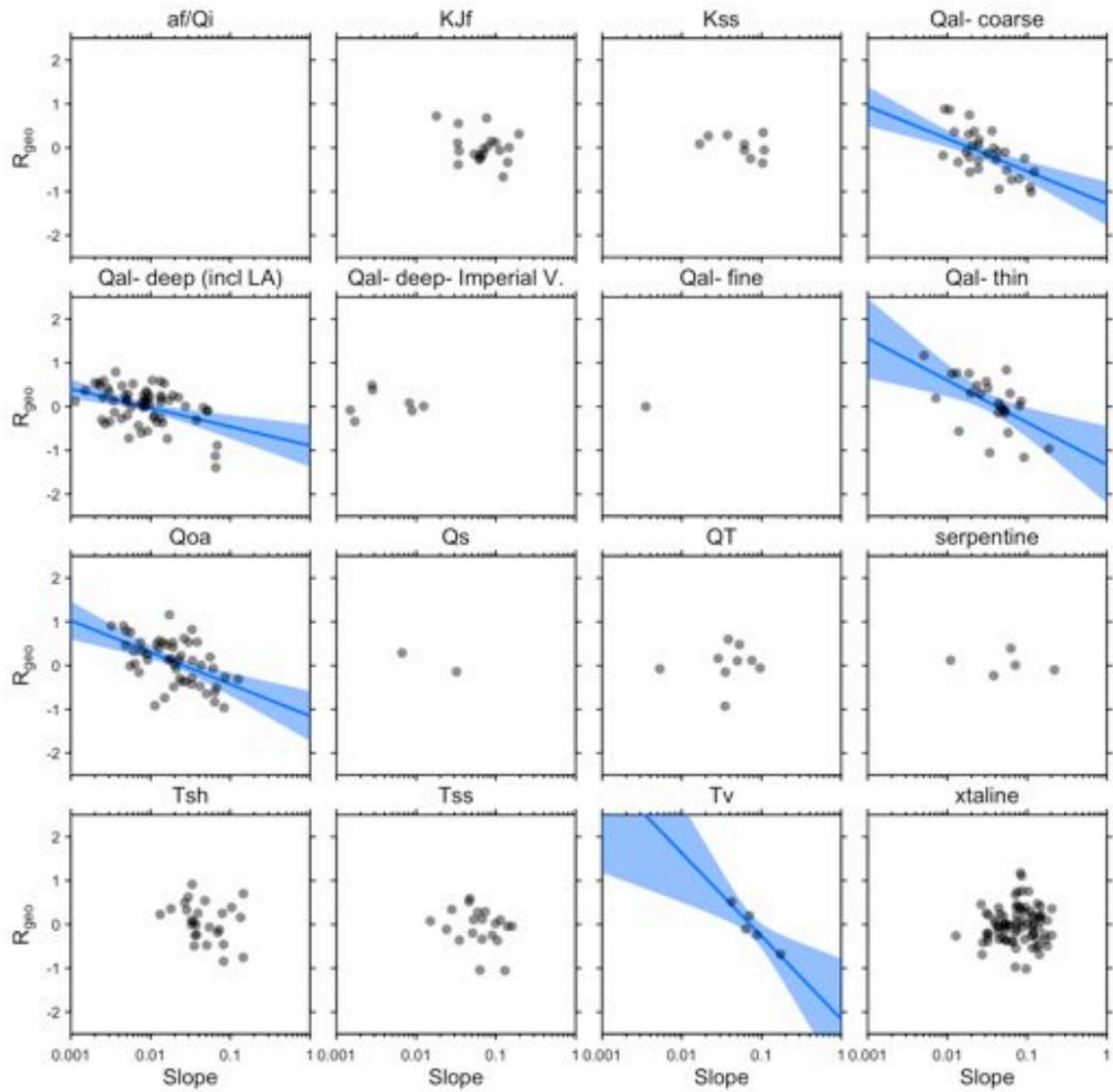


Figure 9. Plots of $T=3.0$ sec R_{geo} vs slope for the different geologic units. Regression curves and their 95% confidence intervals are shown for units where the slope coefficient is significantly different than zero at the 95% significance level.

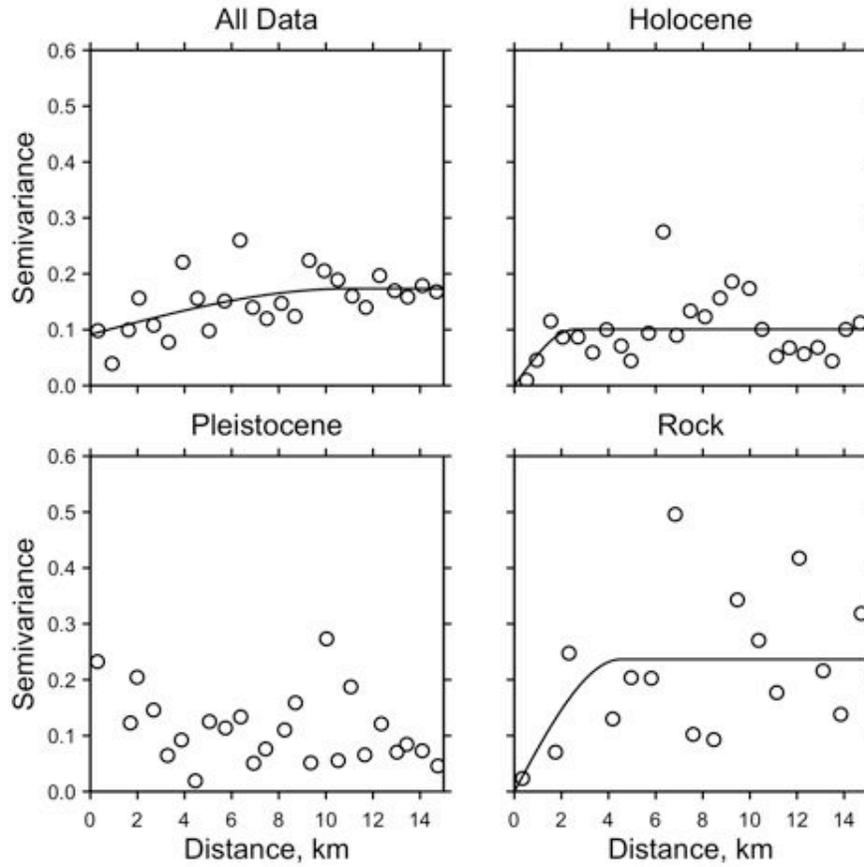


Figure 10. Empirical semivariograms of PGA R_{hyb} for all of the stations, and then broken down into three subsets described in the text. The model semivariogram is not displayed if the optimization algorithm failed to converge on a solution.

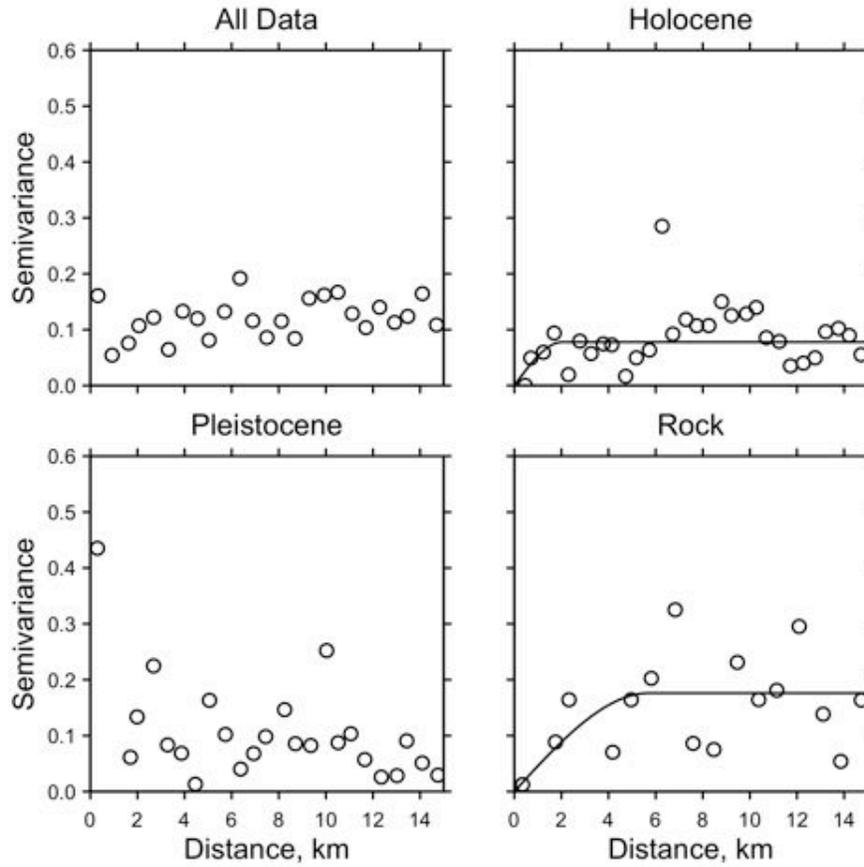


Figure 11. Empirical semivariograms of PGV R_{hyb} for all of the stations, and then broken down into three subsets described in the text. The model semivariogram is not displayed if the optimization algorithm failed to converge on a solution.

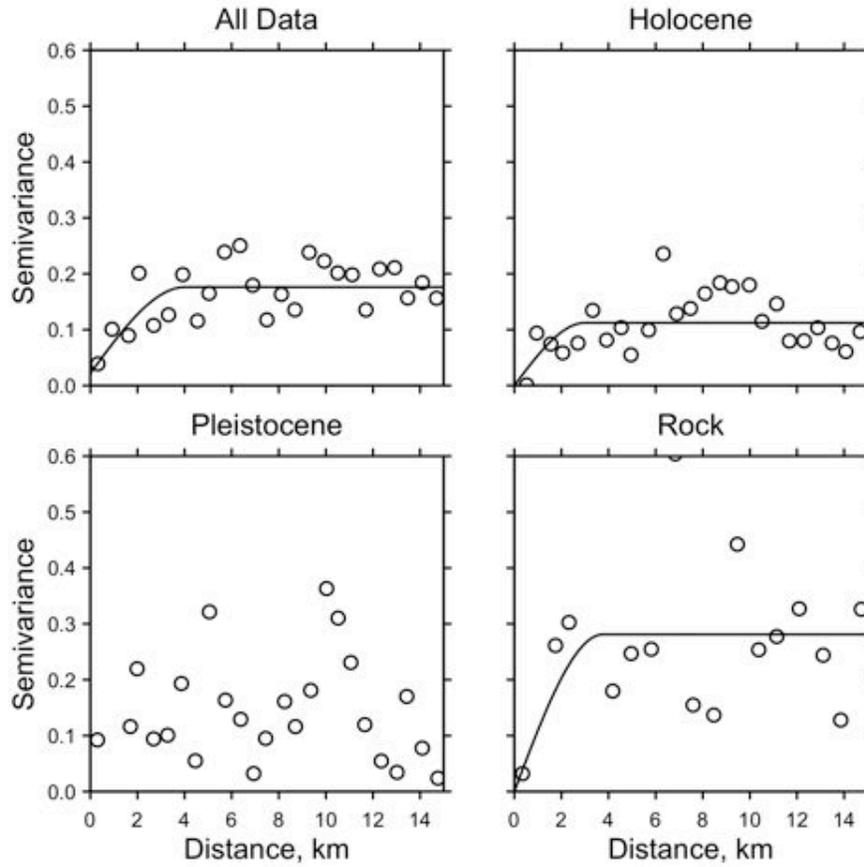


Figure 12. Empirical semivariograms of $T=0.3$ sec R_{hyb} for all of the stations, and then broken down into three subsets described in the text. The model semivariogram is not displayed if the optimization algorithm failed to converge on a solution.

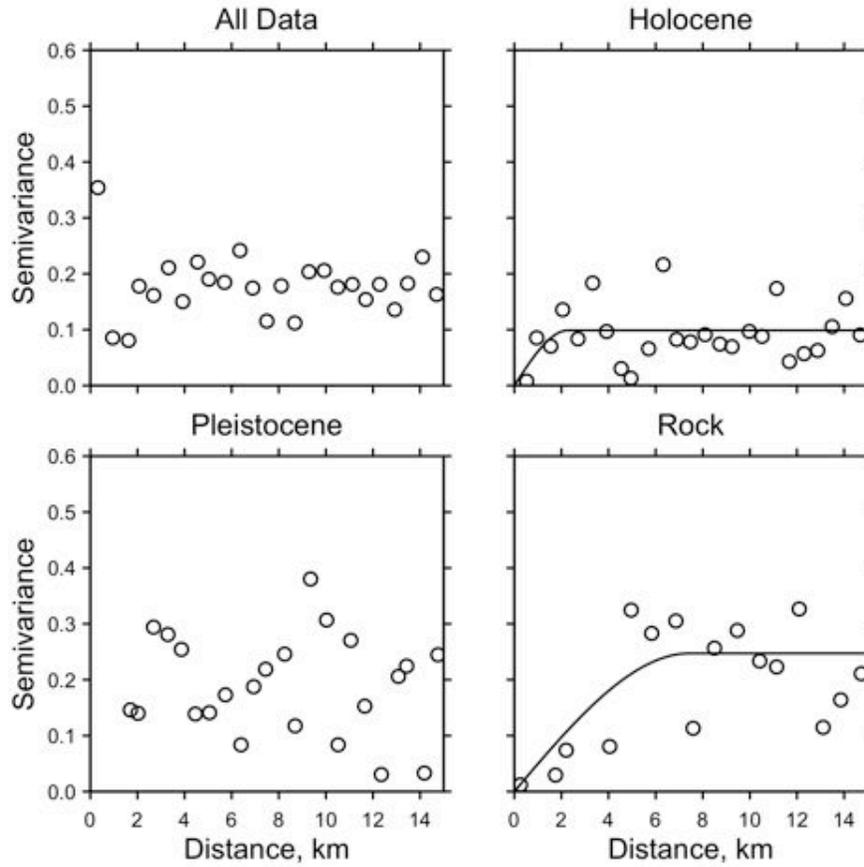


Figure 14. Empirical semivariograms of $T=1.0$ sec R_{hyb} for all of the stations, and then broken down into three subsets described in the text. The model semivariogram is not displayed if the optimization algorithm failed to converge on a solution.

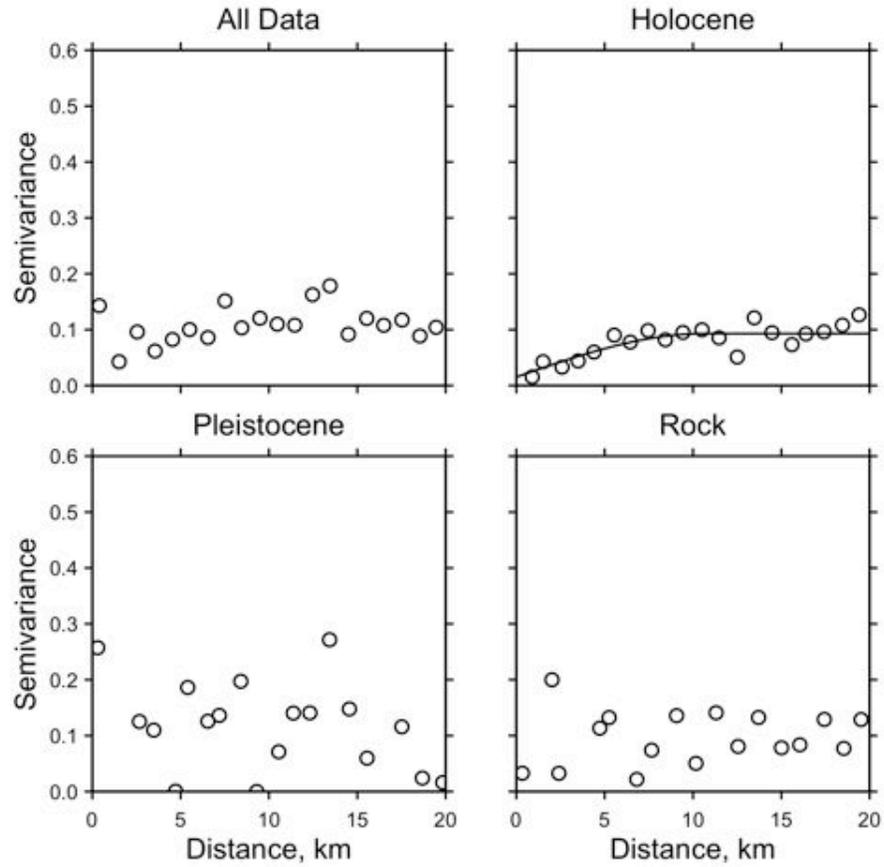


Figure 15. Empirical semivariograms of $T=3.0$ sec R_{hyb} for all of the stations, and then broken down into three subsets described in the text. The model semivariogram is not displayed if the optimization algorithm failed to converge on a solution.

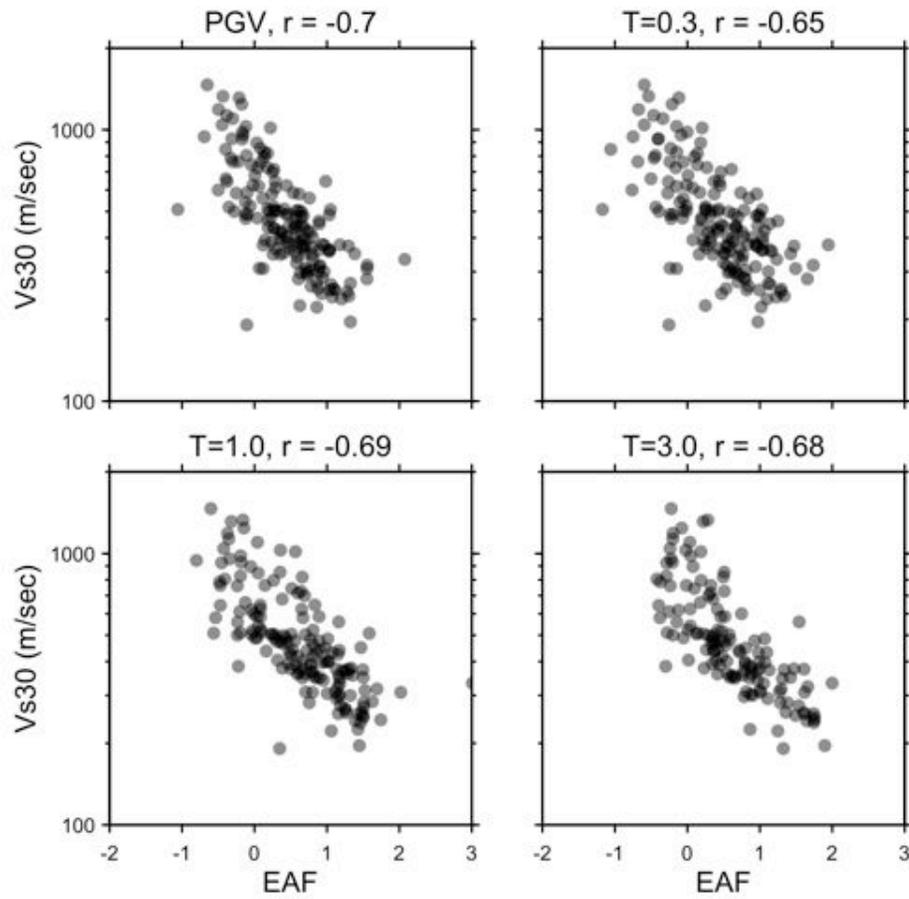


Figure 16. Scatterplots and correlation coefficients for V_{s30} and the EAFs at different spectral periods.

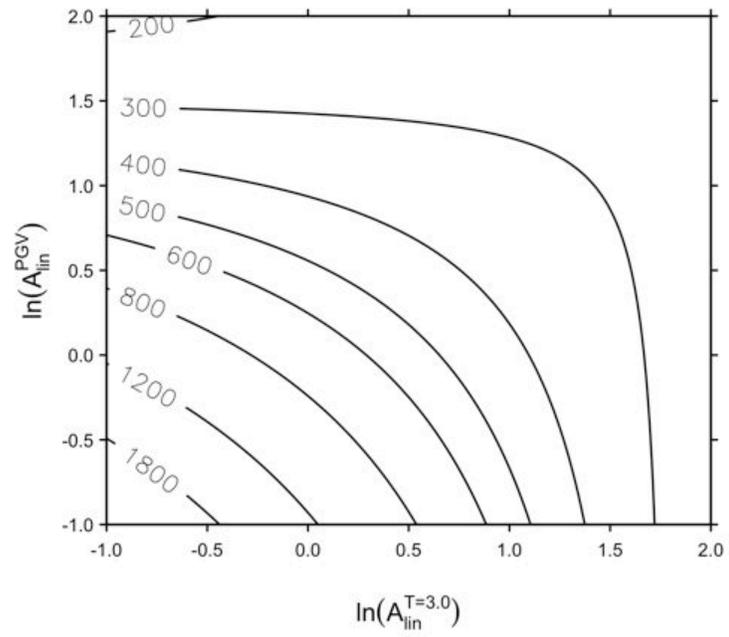


Figure 17. Evaluation of V_{S30} (m/sec) using equation 7.

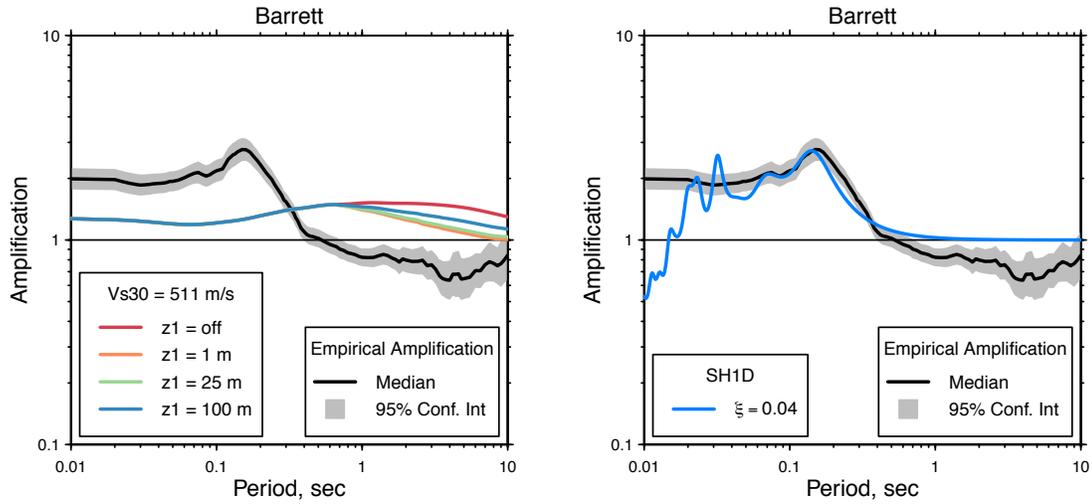


Figure 18. Left: EAFs at Barrett compared to F_S for various values of z_1 . The value of z_1 measured at this site is 25 m (Yong et al., 2013). Right: comparison of the EAFs with the SH1D transfer function with the profile for this site reported by Yong et al. (2013). Note that the SH1D amplification is for Fourier spectra, and so it is not exactly correct to compare to the EAF amplifications, which are from spectral accelerations.

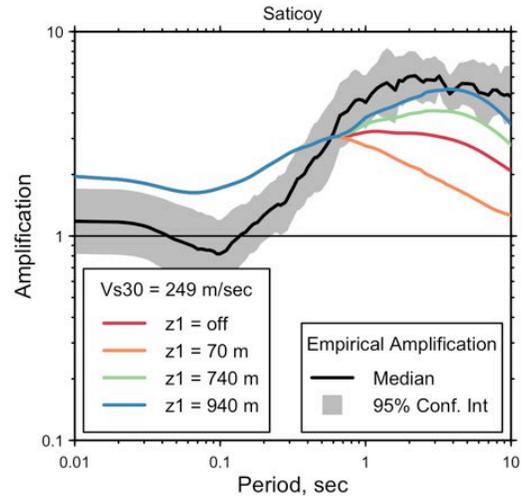


Figure 19. EAFs at Saticoy compared to F_S for various values of z_1 . The available profile does not reach z_1 (Yong et al., 2013).

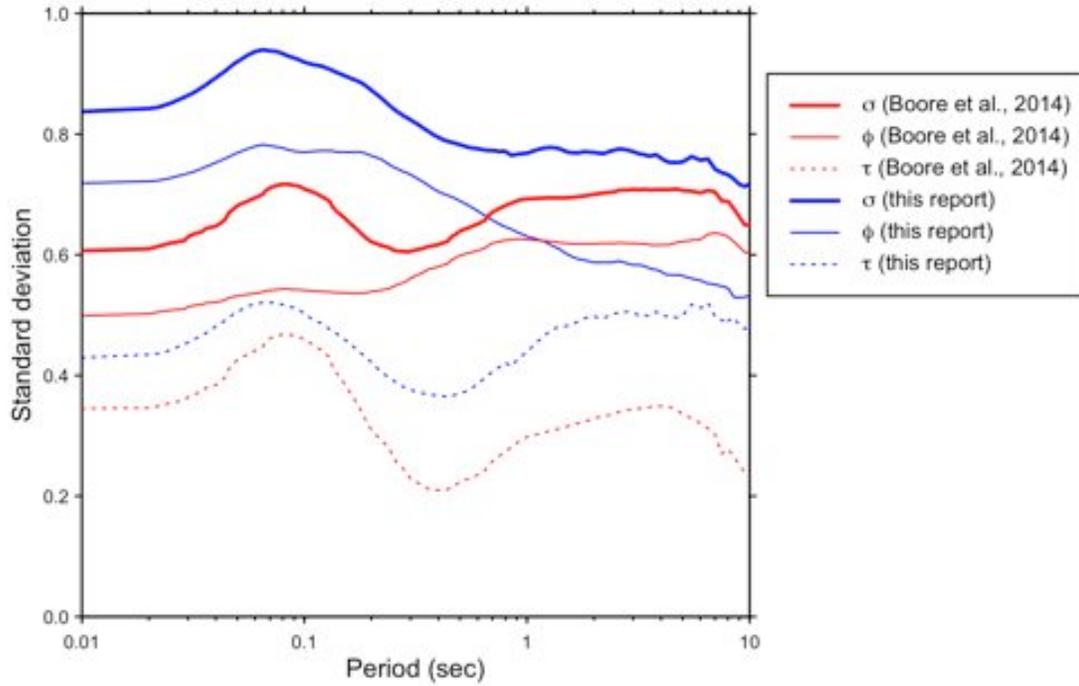


Figure 20. Comparison of standard deviations (defined in equation 11) reported by Boore et al. (2014) and the analogous values computed for the data processed for this report. Note that the Boore et al. (2014) curves are for magnitudes greater than 5.5, distances less than 80 km, and $V_{S30} > 300$ m/s.

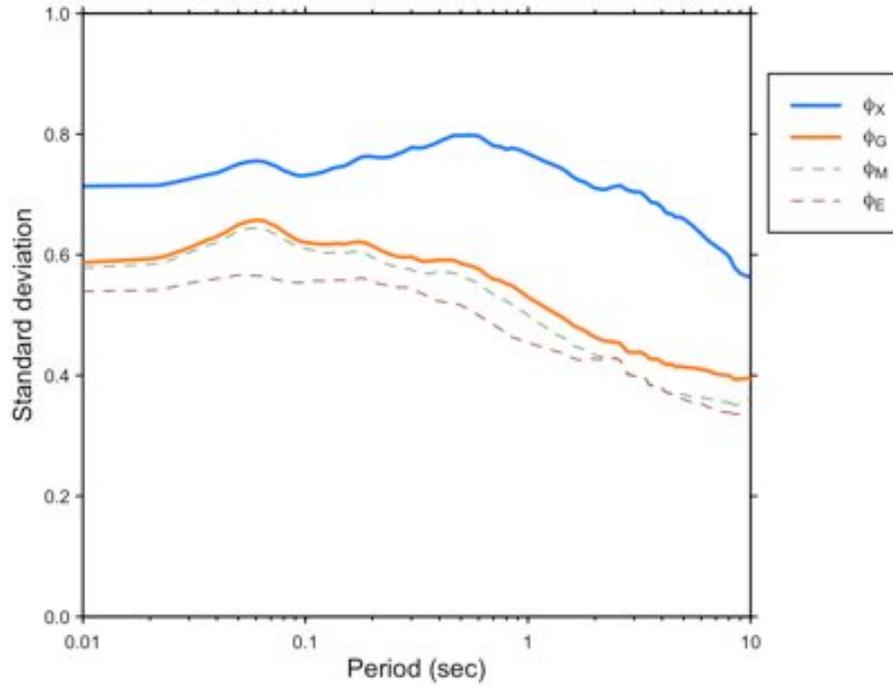


Figure 21. Comparison of the intra-event standard deviation estimates without a site term (ϕ_X), with the site term as F_{lin} with V_{S30} from geology (ϕ_G), with the site term as F_{lin} with measured V_{S30} (ϕ_M), and with $\overline{R_{ec}}$ as the site term (ϕ_E).

Appendix 1: Empirical Amplification Factors for Each Site.

