USGS Award Number: G09AP00062.

Near-surface velocity structure and site resopnse at Parkfield, CA

Term: 1 June 2009 – 31 May 2010

Eric M. Thompson and Laurie G. Baise *

Department of Civil and Environmental Engineering, Tufts University, USA

July 20, 2010

*Corresponding author.
113 Anderson Hall, Tufts University, Medford, MA 02155
email: laurie.baise@tufts.edu
Fax: 617-627-3994

Abstract

The spatial scale of previously proposed methods for mapping site response ranges from the global scale down to individual urban regions exposed to heightened seismic risk. Typically, spatial coverage and accuracy are inversely related. We use the densely spaced strong motion stations in Parkfield, California to estimate the accuracy of different site response mapping methods and demonstrate a method for integrating multiple site response estimates from the site to the global scale. This method is simply a weighted mean of a suite of different estimates, where the weights are the inverse of the variance of the individual estimates. Thus, the dominant site response model varies in space as a function of the accuracy of the different models. Site response models should be judged in terms of both the degree of correlation with observed amplifications and in terms of spatial coverage. Performance varies with period, but in general the Parkfield data show that: (1) Where a velocity profile is available, the square-root-of-impedance method outperforms V_{S30} (30 m divided by the S wave travel time to 30 m depth), and (2) Where velocity profiles are unavailable, the topographic slope method outperforms surficial geology. The performance of the topographic slope method in Parkfield motivates us to use the Next Generation of Attenuation database to develop new equations that predict site response from topographic slope. We show that these equations predict site response more accurately than both the previously published topographic slope method and a previously developed ground motion prediction equation for most spectral periods.

1 Introduction

Earthquake hazard maps, such as the U.S. national seismic hazard maps (Frankel *et al.*, 1996, 2002; Petersen *et al.*, 2008) and rapid response maps (Wald *et al.*, 2006, 2005), are fundamentally linked to empirical ground motion prediction equations (GMPEs). In general, earthquake hazard maps use at least one GMPE to compute the intensity of ground shaking throughout the region of interest. GMPEs are generally a function of three types of explanatory variables: source, path, and site. The source and path variables can be computed at an arbitrary location for an assumed earthquake scenario. In this article, we study the site effects term by comparing recorded ground motions to those predicted by a previously developed GMPE and present (1) a new method for integrating multiple estimates of site response into a period dependent map of site response amplifications, and (2) new equations for incorporating site effects into an existing GMPE using topographic slope, which we develop specifically for the purpose of incorporating site effects into earthquake hazard maps.

Many of the most recently developed GMPEs use V_{S30} (30 m divided by the travel time of an S wave to 30 m depth) to model the site term (Power *et al.*, 2008). A major obstacle to incorporating site effects into earthquake hazard maps is that V_{S30} is not known in most locations. Thus, researchers have developed correlations of V_{S30} with quantities that are more spatially continuous, such as surficial geology (Wills and Clahan, 2006) and topographic slope (Wald and Allen, 2007). Other researchers have focused on the limitations of the ability of V_{S30} to model site response (Castellaro *et al.*, 2008). This limitation can be overcome by including additional site response explanatory variables into GMPEs to improve prediction accuracy (Douglas *et al.*, 2009).

Our first task is to assess the performance of alternative site response explanatory variables. We define performance as the degree of correlation with observed site response amplifications. The quantitative assessment of different site response terms will be valuable for decisions regarding future data collection efforts and GMPE development. We use this information to demonstrate a framework for integrating alternative site response explanatory variables into a unified period-dependent site response map.

We study a suite of site response explanatory variables, including V_{S30}, the fundamental site frequency (f_0), topographic slope, surficial geology, the square-root-of-impedance amplification, and the one-dimensional plane *SH*-wave amplification. We judge the different variables on two different scales: (1) the ability to explain GMPE site response residuals, and (2) data availability, both in terms of number of strong motion stations where the data are available and in terms of geographic coverage for creating earthquake hazard maps. These two scales are generally inversely related. For example, Wald and Allen (2007) developed a method for predicting site response from globally available topography data. Thus, the site response explanatory variable (topographic slope) is known with uniform precision at every strong motion station as well as any point on a map where an estimate of ground shaking intensity is desired. The ubiquity of this explanatory variable, compared to direct measurements of V_{S30}, presumably can only be achieved by sacrificing some predictive accuracy.

Douglas (2003) thoroughly described the history of GMPE development, including the various methods of incorporating site effects. Site effects have been most commonly included into GMPEs with multiplicative factors based on site category (e.g., Trifunac, 1976). Some GMPE developers have also used *S*-wave velocity (V_s) profiles to eliminate the subjectivity of assigning site categories (e.g., Ambraseys, 1995). Three of the five recently developed Next Generation of Attenuation (NGA) GMPEs use multiple explanatory variables: V_{S30} plus a soil depth parameter. Abrahamson and Silva (2008) and Chiou and Youngs (2008) define the soil depth parameter as the depth at which $V_s \geq 1.0$ km/sec ($Z_{1.0}$) and Campbell and Bozorgnia (2008) define the soil depth parameter as the depth at which $V_s \geq 2.5$ km/sec ($Z_{2.5}$).

The accuracy with which any site response variable can model site response is a combination of how well that variable correlates with site response (i.e., model uncertainty) and the accuracy with which that variable can be estimated at a specific location (i.e., estimation uncertainty). For example, an estimate of V_{S30} from downhole logging will presumably correlate better with site response than qualitative categories of "soil" and "rock" because the V_{S30} model exhibits less model uncertainty. In contrast, the downhole estimate of V_{S30} would most likely exhibit a stronger correlation with site response than the V_{S30} inferred from surficial geology or topographic slope because the latter estimates introduce additional estimation uncertainty. In this article, we develop a method for mapping site response that accounts for these two types of uncertainties.

Thompson *et al.* (2010b) collected 52 spectral analysis of surface wave (SASW) V_s profiles at strong motion stations in Parkfield, California. This is an ideal dataset for addressing site response mapping because of the number and spatial density of the strong motion stations and V_s profiles. Although other regions have been densely sampled with V_s measurements (e.g., Holzer *et al.*, 2002), the Parkfield dataset is unique because of the number of locations that contain both a V_s profile and a strong motion station within a relatively small region. We focus on the recordings from the 1983 M 6.36 Coalinga earthquake. We do not to include the 2004 M 6.0 Parkfield earthquake to avoid nonlinear and finite fault effects (Liu *et al.*, 2006; Shakal *et al.*, 2006).

We begin by isolating the site response amplifications for a previously developed GMPE and compare the amplifications in Parkfield to various site response explanatory variables. We then quantify the prediction accuracy of these different explanatory variables, incorporating both model and estimation uncertainty and use this information to integrate the different explanatory variables to produce period-dependent maps of site response for the Parkfield region. The Parkfield dataset is too limited (only 36 observations from one earthquake in one geological/geomorphological environment) to develop equations that are applicable to other regions. The performance of the topographic slope method at Parkfield motivates us to expand our sample size with the NGA flatfile. We develop equations for modifying an existing GMPE to estimate effects from topographic slope using the NGA data. Because these equations are regressed against the same database as the GMPE and the explanatory variable is available globally, these equations can be used for incorporating site effects into seismic hazard maps.

2 Data and Resources

Figure 1 shows the 52 strong motion stations that we analyze in this study and the epicenter of the 1983 M 6.36 Coalinga earthquake. Thompson *et al.* (2010b) described the details of the data collection effort. V_s profiles at the strong motion stations were measured with the SASW method and are available at http://gdc.cee.tufts.edu (last accessed May 2010).

The ground motion records for the 1983 M 6.36 Coalinga earthquake are available from the Consortium of Organizations for Strong-Motion Observation Systems (COSMOS) at http://db.cosmos-eq.org (last accessed May 2010).

We digitized the 1:250,000 scale Jennings (1958) map of surficial geology that is available from the California Geological Survey. The digitized version of the map is plotted in Thompson *et al.* (2010b).

The Shuttle Radar Topography Mission (SRTM) 30 sec global topography data (Farr and Kobrick, 2000) are available at http://www2.jpl.nasa.gov/srtm (last accessed May 2010) and we computed the topographic slope with Generic Mapping Tools (GMT; Wessel and Smith, 1991) as described by Wald and Allen (2007). All other computations in this article were completed with the the open source software R (R Development Core Team, 2010).

The NGA flatfile is available at http://peer.berkeley.edu/nga/flatfile.html (last accessed May 2010).

3 Methods

3.1 Response Variable

Isolating site effects from the other processes that influence ground shaking intensity is not a trivial matter (Abercrombie, 1997; Boore and Joyner, 1997; Steidl *et al.*, 1996). The most common approach is to compare the response at two stations for which the source and path effects can be assumed to be identical. Thus, any differences must be due to the local site conditions (i.e., site response). In terms of the pseudospectral acceleration (PSA), which is a

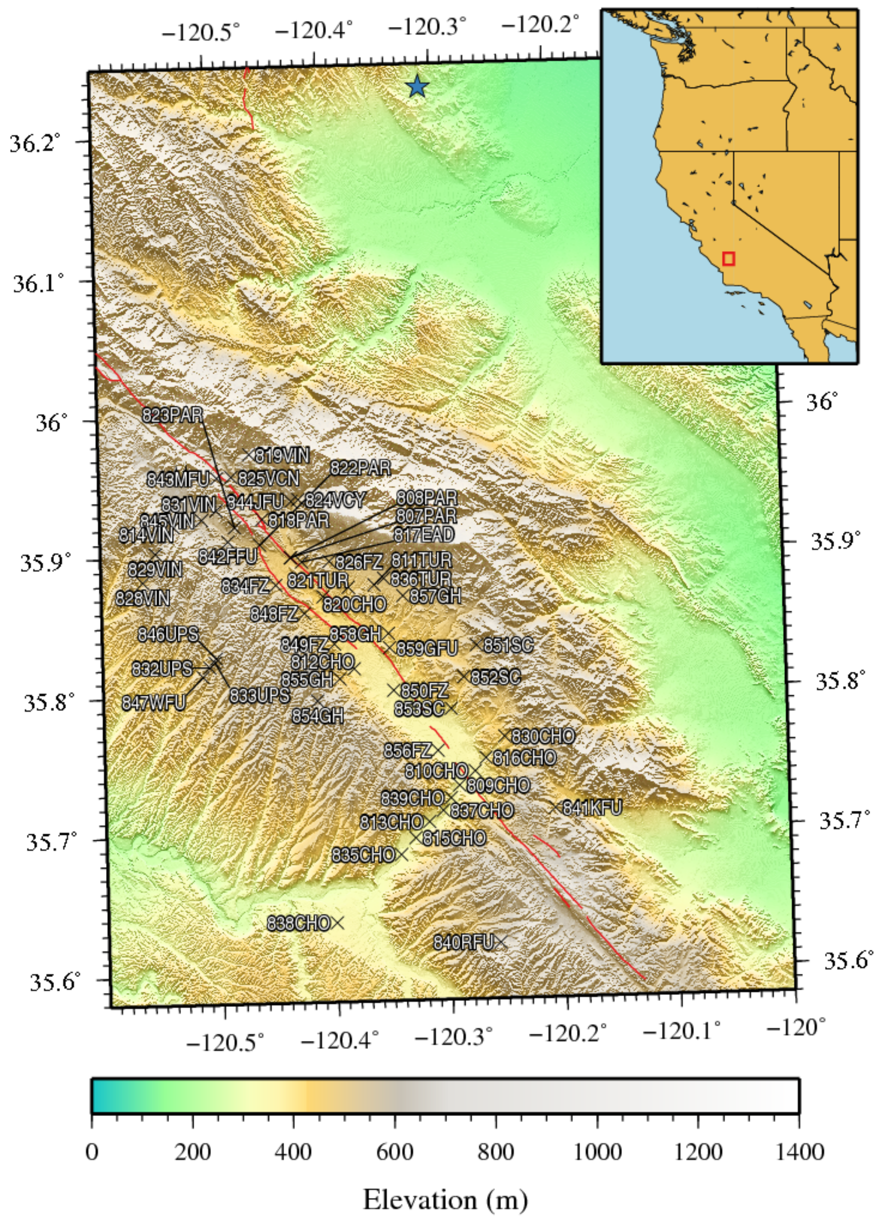

Figure 1: Vicinity map showing the location of the 52 strong motion stations with SASW measurements and the epicenter of the 1983 Coalinga earthquake.

function of the fundamental period (T) of a single-degree-of-freedom system (SDOF), the site response amplification is

$$a(T) = \frac{\mathrm{PSA}^{*}(T)}{\mathrm{PSA}^{\mathrm{ref}}(T)} , \tag{1}$$

where $\mathrm{PSA}^{\mathrm{ref}}(T)$ is the PSA at the reference site and $\mathrm{PSA}^{*}(T)$ is the PSA at the location of interest. All of the PSA values in this article assume 5% viscous damping. Generally, the reference site is chosen to be a site that has minimal site effects, and is termed a "rock" site. For example, Borcherdt (1994) defined site response as the ratio of the spectra of ground motions recorded on soil to the spectra of ground motions recorded on rock. A drawback to this approach is that the choice of the reference site can substantially influence the estimate of $a(T)$ because $\mathrm{PSA}^{\mathrm{ref}}(T)$ may not be known with certainty.

We wish to avoid the somewhat arbitrary choice of reference site since there is no clear rock site that is appropriate for many of the Parkfield strong motion stations. Since our goal is to assess site response explanatory variables in terms of performance within GMPEs, we define $\mathrm{PSA}^{\mathrm{ref}}$ as the PSA computed from an existing GMPE for the reference rock condition. The major drawbacks to this approach are that we must assume all the path and source effects are accurately accounted for by the GMPE. We are comfortable with this assumption for this application because it is unlikely that the unaccounted-for source and path variability will create spurious correlations with any of the site response explanatory variables that we consider in this article. Uncorrelated variability may reduce the degree of correlation between the site response explanatory variables and the site response amplification values, but will not favor one variable over another. Note that we need not worry about the accuracy of how well the GMPE developers captured site response because we remove the effects of the site response term by assuming the reference rock condition.

We use the Boore and Atkinson (2008) GMPE (BA08) to compute the rock (V_{S30} = 760 m/sec) PSA [$\mathrm{PSA}^{r}(T)$] for each station. To implement the BA08 equations, the magnitude (6.36) and rake angle (90°) of the event were obtained from the NGA flatfile, along with the Joyner-Boore distance (R_{JB}) for stations that are included in the NGA flatfile. We cannot compute R_{JB} directly because we are unaware of a publicly available finite fault model for this event. Table 1 gives station summary information and R_{JB} for the 36 stations that recorded the 1983 Coalinga earthquake. One station (853SC; Stone Corral 1E) is not available in the NGA flatfile, so we estimate R_{JB} by linear extrapolation from stations Stone Corral 2E and Stone Corral 3E, which are approximately evenly spaced in a line.

Rewriting equation 1 in terms of natural logs, which is appropriate because of the functional form of the BA08 GMPE, and defining $\mathrm{PSA}^{\mathrm{ref}}(T) = \mathrm{PSA}^{r}(T)$ gives

$$\ln[a(T)] = \ln[\mathrm{PSA}^{*}(T)] - \ln[\mathrm{PSA}^{r}(T)], \tag{2}$$

which is the dependent (or response) variable of interest in this article. The largest usable period ($T_{\max}$) is a function of the corner frequency of the high-pass filter ($f_{\min}$) applied to the Coalinga ground motion records; $f_{\min} = 0.07$ Hz for the 36 records of the Coalinga earthquake that we use in this article. Following Abrahamson and Silva (1997), we assume $T_{\max} = 1/(1.25 \times f_{\min}) \approx 11.4$ sec.

Table 1: Summary of station information and the explanatory variables required to compute the PSA from the Boore and Atkinson (2008) GMPE for the 1983 Coalinga earthquake (where applicable).

ID	Latitude	Longitude	V_{S30} (m/sec)	R_{JB} (km)
807PAR	35.8985	-120.4329	261	27.96
808PAR	35.8986	-120.4325	270	
809CHO	35.7328	-120.2893	173	43.83
810CHO	35.7430	-120.2753	233	42.76
811TUR	35.8779	-120.3599	907	
812CHO	35.8180	-120.3791	214	35.04
813CHO	35.7072	-120.3163	283	46.73
814VIN	35.9220	-120.5348	309	30.91
815CHO	35.6963	-120.3288	237	47.88
816CHO	35.7521	-120.2657	523	41.99
817EAD	35.8952	-120.4229	384	
818PAR	35.9266	-120.4570	384	26.20
819VIN	35.9731	-120.4671	468	22.66
820CHO	35.8701	-120.4051	297	29.91
821TUR	35.8767	-120.3826	309	28.58
822PAR	35.9080	-120.4595	246	28.11
823PAR	35.9211	-120.4813	308	28.00
824VCY	35.9399	-120.4247	657	
825VCN	35.9573	-120.4831	381	24.83
826FZ	35.8955	-120.3989	542	27.10
828VIN	35.8829	-120.5629	320	35.85
829VIN	35.9033	-120.5513	386	33.28
830CHO	35.7670	-120.2488	397	40.01
831VIN	35.9336	-120.4974	284	27.72
832UPS	35.8214	-120.5059	358	
833UPS	35.8248	-120.5011	417	
834FZ	35.8791	-120.4461	372	30.43
835CHO	35.6842	-120.3418	252	49.40
836TUR	35.8824	-120.3510	467	
837CHO	35.7151	-120.3041	410	45.49
838CHO	35.6369	-120.3998	359	55.05
839CHO	35.7239	-120.2970	231	44.82
840RFU	35.6199	-120.2570	239	
841KFU	35.7153	-120.2056	576	
842FFU	35.9108	-120.4873	227	
843MFU	35.9564	-120.4959	398	
844JFU	35.9389	-120.4309	379	
845VIN	35.9268	-120.5100	439	29.01
846UPS	35.8277	-120.5001	342	
847WFU	35.8145	-120.5118	447	
848FZ	35.8586	-120.4214	267	31.64
849FZ	35.8360	-120.3958	221	33.42
850FZ	35.8019	-120.3450	212	36.14
851SC	35.8331	-120.2713	565	32.81
852SC	35.8105	-120.2831	566	35.29
853SC	35.7878	-120.2948	261	37.77
854GH	35.7958	-120.4120	511	38.10
855GH	35.8113	-120.3922	290	35.93
856FZ	35.7580	-120.3075	178	41.04
857GH	35.8695	-120.3346	451	28.72
858GH	35.8428	-120.3485	361	31.85
859GFU	35.8324	-120.3477	558	

3.2 Site Response Explanatory Variables

The site response explanatory variables that we analyze in this article can be categorized in a number of useful ways. We consider both vector and raster spatial data, as well as period dependent and independent data. The only type of raster data that we use is topographic slope, and we use two types of vector data: point data (V_s profiles and strong motion data) and polygon data (surficial geology). These data (with the exception of the strong motion data) can be easily processed to provide an estimate of V_{S30}, which is the most commonly used site response explanatory variable. Since we discuss many different estimates of V_{S30}, it is convenient to use superscripts to specify the different estimation techniques. For example, we refer to the estimate of V_{S30} from an SASW measurement as V_{S30}^{SASW}. We choose not to include the soil depth parameters in our analysis because only eight of the V_s profiles measured $Z_{1.0}$ and none measured $Z_{2.5}$.

We compute the magnitude of the topographic slope (δ) with the method described by Wald and Allen (2007). The slope is estimated from SRTM 30 sec global topography (Farr and Kobrick, 2000) so the pixel size is approximately 1 km by 1 km. From δ, we estimate the V_{S30} (V_{S30}^{δ}) with the empirical relationship presented by Wald and Allen (2007).

We refer to the estimate of V_{S30} predicted by the geologic classification method of Wills and Clahan (2006) as V_{S30}^{geo}. All but one of the 36 strong motion stations that recorded the Coalinga earthquake are included in the NGA Flatfile. Thus, we are able to ensure that we use the same geologic classifications that were used in the NGA GMPE development for 35 of the stations that we consider. The one station that is not in the Flatfile is Stone Corral 1E (853SC in Table 1), which is classified as Quaternary nonmarine terrace deposits (Qt) according to Jennings (1958). The Wills and Clahan (2006) classification that we assign to this station is Quaternary (Pleistocene) alluvium (Qoa).

To map (interpolate) V_{S30}^{geo} we need to assign a Wills and Clahan (2006) classification to locations that are not in the NGA flatfile. Table 2 gives the Wills and Clahan (2006) geologic unit that we assigned to the Jennings (1958) units that are in the vicinity of the stations that we consider in this article.

Lermo and Chávez-García (1993) provided evidence that the horizontal-to-vertical spectral ratios (H/V) may provide valuable site response information. This approach has become very popular due to the ease with which an estimate of the site conditions can be achieved. We use the procedure described by Zhao *et al.* (2006) to estimate f_0 from H/V at the 36 stations in this article that recorded the 1983 Coalinga earthquake. Ideally, we would use multiple earthquakes to achieve a more robust estimate of f_0, but the vast majority of these stations only have records available for the 1983 Coalinga earthquake and the 2004 Parkfield earthquake. H/V ratios are typically estimated from the Fourier spectrum (e.g., Cadet *et al.*, 2010), but the Zhao *et al.* (2006) method uses the 5% damped PSA instead. The advantage of using PSA over the Fourier spectrum is that no additional smoothing is required. From the three orthogonal ground motion components (two horizontal, H1 and H2; and one vertical, V) we compute two estimates of H/V: $\text{PSA}_{H1}(T)/\text{PSA}_V(T)$ and $\text{PSA}_{H2}(T)/\text{PSA}_V(T)$. T_1 and T_2 are the periods at which the two estimates of H/V are maximized and the estimate of the fundamental frequency is $f_0 = 1/\sqrt{T_1 \times T_2}$.

The SASW V_s profiles include information that should be pertinent for site response besides the single value of V_{S30}. Thus, we investigate alternative data reduction methods to simplify

Table 2: Relationship between Wills and Clahan (2006) geology classifications and the Jennings (1958) geologic units.

Jennings (1958)		Wills and Clahan (2006)
Symbol	Description	Symbol
Qal	Alluvium	Qal, thin
Qf	Fan deposits	Qal, thin
Qt	Quaternary nonmarine terrace deposits	Qoa
QP	Plio-Pleistocene nonmarine	QT
Pc	Undivided Pliocene nonmarine	QT
Pu	Upper Pliocene marine	Tsh
Pml	Middle and/or lower Pliocene marine	Tsh
Mc	Undivided Miocene nonmarine	Tsh
Mu	Upper Miocene marine	Tsh
Mv(r)	Miocene volcanic – rhyolite	Tv
Mm	Middle Miocene marine	Tsh
Ml	Lower Miocene marine	Tsh
Oc	Oligocene nonmarine	Tss
E	Eocene marine	Tss
K	Undivided Cretaceous marine	Kss
Ku	Upper Cretaceous marine	Kss
Kl	Lower Cretaceous marine	Kss
KJfv	Franciscan volcanic and metavolcanic rocks	KJf
KJf	Franciscan Formation	KJf
gr	Mesozoic granite rocks	xtaline
bi	Mesozoic basic intrusive rocks	xtaline
ub	Mesozoic ultrabasic intrusive rocks	xtaline
ls	Pre-Cretaceous metamorphic rocks – limestone or dolomite	xtaline

the information in the V_s profile to a single value that can be used as an explanatory variable in a regression analysis. This includes the square-root-of-impedance method and the one-dimensional plane *SH*-wave computation. The key feature of these explanatory variables, is that they are period dependent, unlike V_{S30} and f_0.

The square-root-of-impedance (SRI) method of estimating the site response amplifications computes the amplification that a seismic wave will experience as it travels between two materials with different impedances, where the impedance is defined as the product of density and velocity: $\rho \times V$. The key assumption of this method is that the seismic energy is completely transmitted across the material interface. The SRI method ignores the effects of seismic reflections and refractions that will occur within a stack of laterally constant layers. Joyner *et al.* (1981) found that the impedance effect, as opposed to the effects of reflections and refractions, is the dominant process that contributes to site amplification. The method for estimating the period-dependent amplification with the SRI method [$a^{\mathrm{SRI}}(T)$] has been extensively described in the relevant literature (e.g., Boore, 2003; Douglas *et al.*, 2009; Joyner *et al.*, 1981; Thompson *et al.*, 2010a) so we do not reproduce it here.

Accounting for the effect of reflections and refractions could potentially provide a more realistic model $a(T)$ than $a^{\mathrm{SRI}}(T)$. We accomplish this by assuming plane *SH* wave (the horizontally polarized component of the *S* wave) propagation through a laterally homogeneous medium (SH1D). The resulting amplifications [$a^{\mathrm{SH1D}}(T)$] are computed using the Thomson-Haskell technique (Haskell, 1953; Thomson, 1950).

3.3 Testing for Independence

One of the goals of this paper is to assess the degree of correlation between the different explanatory variables and $\ln[a(T)]$. Pearson's product moment correlation coefficient (r) is the most widely used measure of the linear association between two variables (Draper and Smith, 1981). The hypothesis test for deviation of r from zero requires that the data are assumed to follow the bivariate normal distribution (Fisz, 1963, Section 9.9). Generally, the variables under consideration may be assumed to be lognormally distributed. The one-sided alternative hypothesis depends on the explanatory variable under consideration. For example, we expect $r < 0$ for $\ln[V_{S30}]$ and $\ln[a(T)]$. In contrast, we expect $r > 0$ for $\ln[a^{\text{SRI}}(T)]$ and $\ln[a(T)]$.

3.4 Mapping Site Response

Some of the site response explanatory variables that we consider in this article are available throughout the region of interest, namely V_{S30}^{geo} and δ. In contrast, explanatory variables derived from the SASW measurements [V_{S30}^{SASW}, $a^{\text{SRI}}(T)$, and $a^{\text{SH1D}}(T)$] are only available at 52 locations, and f_0 is only available at the 36 strong motion stations where the Coalinga earthquake was recorded.

If we wish to create maps of site response amplification, we must interpolate the explanatory variables that do not cover the extent of the region of interest. We use the ordinary kriging (OK) method as described by Cressie (1993) and employed by Thompson *et al.* (2010a) to map site response amplifications in the Kobe, Japan sedimentary basin. Due to space limitations, we do not reproduce the mathematics of the OK method, but we adopt the same notation as Cressie (1993) so that it can easily be used as a reference for the spatial statistics (i.e., kriging) material in this article.

Since our goal is to integrate different explanatory variables to predict $a(T)$, we are especially interested in the variance of the OK prediction (σ_k^2), which varies in space and increases with distance from the observed data. Kriging is a generalized least-squares regression algorithm that minimizes σ_k^2. The kriging equations fundamentally depend on the spatial correlation structure of the data, which is typically represented by the variogram [$2\gamma(\mathbf{h})$], where $\mathbf{h}$ is the Euclidean distance between two locations. Often the semivariogram [$\gamma(\mathbf{h})$] is analyzed to diagnose the spatial structure. In this article, we use the "classical" variogram estimator $2\hat{\gamma}(\mathbf{h})$, which uses the method-of-moments technique. The kriging equations require a variogram model [$2\gamma(\mathbf{h}; \boldsymbol{\theta})$], where $\boldsymbol{\theta}$ is the vector of model parameters. We use a three-parameter semivariogram model that is termed the Whittle-Matérn model (Guttorp and Gneiting, 2005).

3.5 Regression Analysis

We apply the classic straight line regression formulation

$$\ln[a(T)] = \beta_0 + \beta_1 X + \epsilon \tag{3}$$

for predicting $\ln[a(T)]$ from an explanatory variable X, where β_0 and β_1 are the parameters of the model, and ϵ is a normally distributed random variable with mean zero and variance σ^2. We notate the least squares estimates of β_0 and β_1 as b_0 and b_1, respectively (Draper and Smith, 1981).

A key assumption of the linear model is that X is not a random variable, and thus has zero variance. Under this assumption, the variance of a predicted value of $\ln[a(T)]$ for a specific value of X (X_0) is given by

$$s_p^2 = s^2 \left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right] , \tag{4}$$

where n is the number of observations of X (X_i), $\bar{X}$ is the sample mean of X_i, and

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} \Big(\ln[a(T)]_i - \ln[\hat{a}(T)]_i \Big)^2 , \tag{5}$$

where $\ln[\hat{a}(T)]$ is the ordinary least squares estimate of $\ln[a(T)]$ (Draper and Smith, 1981).

For the case that $X = \ln[V_{S30}^{\text{geo}}]$, the reliability of the observations of X is variable. The standard deviation of $\ln[V_{S30}^{\text{geo}}]$ clearly depends on geologic unit (see Table 1 in Wills and Clahan, 2006). To account for this, we perform weighted least squares regression, assuming that the observations of X are independent but have different variances. The weights $w_i = 1/\sigma_i^2$, where σ_i^2 is set to the square of the sample standard deviations $\ln[V_{S30}^{\text{geo}}]$ reported in Table 1 of Wills and Clahan (2006).

We use the log transformation of each of the explanatory variables (e.g, $X = \ln[V_{S30}]$ rather than $X = V_{S30}$) because (1) this removes much of the skew in each of the variables, and (2) all of the explanatory variables are necessarily positive.

Note that $X = \ln[a^{\text{SRI}}(T)]$ and $X = \ln[a^{\text{SH1D}}(T)]$ are the only variables for which there is a theoretical expectation that $\beta_0 = 0$. Thus, we remove the β_0 from the regression if it is clearly insignificant (P-value is greater than 0.3).

3.6 Regression Predictions at Unsampled Locations

At locations where X is estimated by kriging (e.g., for $X = \ln[f_0]$ and $X = \ln[a^{\text{SRI}}(T)]$), the assumption that the variance of X is zero is severely violated. Thus, we must add another error term to equation 3 to predict $\ln[a(T)]$ where X is a kriging estimate (X_k)

$$\ln[a_k(T)] = \beta_0 + \beta_1(X_k + \eta) + \epsilon , \tag{6}$$

where η is a normally distributed random variable with mean zero and we assume that the variance is σ_k^2. Equation 4 does not hold under these conditions so we estimate the variance of a prediction at an unsampled location ($s_{p,k}^2$) with the Monte Carlo method as follows:

1. Generate $n_{\text{sim}} = 100,000$ samples of each random variable in equation 6. The random variables are distributed according to:

$$\begin{aligned} \beta_0 &\sim t(\mu_{\beta_0}, n-2, \sigma_{\beta_0}), \\ \beta_1 &\sim t(\mu_{\beta_1}, n-2, \sigma_{\beta_1}), \\ \eta &\sim N(0, \sigma_k), \text{ and} \\ \epsilon &\sim N(0, \sigma), \end{aligned} \tag{7}$$

where $t(\mu, n_{\text{df}}, \sigma)$ is Student's t-distribution with mean μ, n_{df} degrees of freedom, and standard deviation σ; $N(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ. To generate the random sample we substitute the least squares estimates of the means and standard deviations for the theoretical values in equation 7. The vectors of randomly generated samples in equation 7 are denoted $\mathbf{b}_{0,i}$, $\mathbf{b}_{1,i}$, $\mathbf{n}_i$, and $\mathbf{e}_i$, respectively.

2. Compute n_{sim} samples of $\ln[a_k(T)]$:

$$\ln[\mathbf{a}_k(T)]_i = \mathbf{b}_{0,i} + \mathbf{b}_{1,i}(X_k + \mathbf{n}_i) + \mathbf{e}_i \; . \tag{8}$$

3. Compute $s^2_{p,k}$ as the sample variance of $\ln[\mathbf{a}_k(T)]_i$.

4. Repeat the above steps for each value of X_k.

Note that we do not need to adjust the estimators of β_0 and β_1 to account for η because the data used in the regressions are not kriged estimates.

3.7 Integrating Multiple Estimates of Site Response Amplification

The key to integrating multiple estimates of $\ln[a(T)]$ is the prediction variance: $s^2_{p,k}$ for $\hat{a}^{f_0}(T)$ and $\hat{a}^{\text{SRI}}(T)$ are denoted $s^2_{f_0}(T)$ and $s^2_{\text{SRI}}(T)$, respectively; s^2_p for $\hat{a}^{\text{geo}}(T)$ and $\hat{a}^{\delta}(T)$ are denoted $s^2_{\text{geo}}(T)$ and $s^2_{\delta}(T)$, respectively. We choose to combine the four different estimates into a variance-weighted estimate ($\ln[\hat{a}^{\text{vwe}}(T)]$). This is simply a weighted average of the different estimates of $\ln[a(T)]$, where each estimate is given a weight equal to the reciprocal of the variance of the estimate.

$$\ln[\hat{a}^{\text{vwe}}(T)] = \frac{\hat{a}^{f_0}(T)/s^2_{f_0}(T) + \hat{a}^{\text{SRI}}(T)/s^2_{\text{SRI}}(T) + \hat{a}^{\text{geo}}(T)/s^2_{\text{geo}}(T) + \hat{a}^{\delta}(T)/s^2_{\delta}(T)}{1/s^2_{f_0}(T) + 1/s^2_{\text{SRI}}(T) + 1/s^2_{\text{geo}}(T) + 1/s^2_{\delta}(T)} \; . \tag{9}$$

3.8 Topographic Slope

Wald and Allen (2007) correlated topographic slope with V_{S30} and recommended that the site response amplifications then be computed from V_{S30} using the Borcherdt (1994) equations. We hypothesize that a substantial improvement can be achieved by correlating $\ln[a(T)]$ directly with $\ln[\delta]$ using the NGA ground motion database.

BA08 excluded many records in the NGA flatfile from their analysis. This included aftershocks, singly recorded earthquakes, and a number of other criteria. We use exactly those records that were used by BA08 for our analysis of the NGA data in this article. In this way, we ensure that the equation we propose for predicting $a(T)$ from δ is valid for the same conditions as the BA08 equations.

Equation 3 assumes that there is a linear relationship between X and $\ln[a(T)]$, where X is some parameter associated with location of the ground motion recording. Unlike the Parkfield recordings of the Coalinga earthquake, the NGA database contains recordings that span a wider range of PSA, and so must add another term to equation 3 to account for the dependence of $\ln[a(T)]$ on the intensity of the reference ground motion (we avoid the term "nonlinear" because of its different meanings in the context of linear regression and soil dynamics):

$$\ln[a(T)] = \beta_0 + \beta_1 \ln[\delta] + \beta_2 \ln[\text{PSA}^r(T)] + \epsilon \; . \tag{10}$$

Table 3: Correlation coefficients for the different site response explanatory variables as a function of PSA period.

T (sec)	$\ln[f_0]$	$\ln[\delta]$	$\ln[V_{S30}^{\text{geo}}]$	$\ln[V_{S30}^{\text{SASW}}]$	$\ln[a^{\text{SRI}}(T)]$	$\ln[a^{\text{SH1D}}(T)]$
0.05	0.33	-0.26	-0.15	-0.39	0.36	-0.11
0.10	0.42	-0.22	-0.13	-0.26	0.29	-0.05
0.20	0.42	-0.25	-0.07	-0.25	0.30	0.03
0.40	0.39	-0.29	-0.07	-0.33	0.35	0.08
0.50	0.16	-0.23	-0.07	-0.50	0.52	0.18
0.75	0.05	-0.28	-0.29	-0.70	0.70	0.32
1.00	-0.09	-0.07	-0.19	-0.59	0.60	0.26
2.00	-0.39	0.13	0.04	-0.46	0.52	0.17
4.00	-0.40	0.04	-0.08	-0.54	0.63	0.36
5.00	-0.33	0.02	-0.07	-0.51	0.59	0.31
7.50	-0.40	0.04	-0.08	-0.54	0.63	0.36
10.00	-0.01	0.19	0.19	-0.14	0.23	0.04

Note that $\delta = 0$ at a few locations and $\ln(0)$ is undefined, so we apply a minimum "water level" such that $\delta \geq 5 \times 10^{-4}$. We select this value because it is approximately the minimum non-zero value of δ in these data.

This method is only valuable if it improves upon the method proposed by Wald and Allen (2007). We measure the accuracy of the different models with the mean squared error

$$\text{MSE}(T) = \frac{1}{n} \sum_{i=1}^{n} \Big(\ln[a(T)]_i - \ln[\hat{a}(T)]_i \Big)^2 . \qquad (11)$$

We define $\text{MSE}^{\delta}(T)$ for when $\ln[\hat{a}(T)]$ is computed from equation 10, $\text{MSE}^{\text{WA07}}(T)$ for when $\ln[\hat{a}(T)]$ is computed with the Wald and Allen (2007) method, and $\text{MSE}^{\text{BA08}}(T)$ for when $\ln[\hat{a}(T)]$ is computed from the BA08 equations.

4 Results

4.1 Correlation and Regression of Site Response Residuals

The site response residuals for $T = 0.2, 0.5, 1.0, 4.0$ sec are shown in Figures 2, 3, 4, and 5, respectively. Subfigures (a) through (f) in each are for $X = \ln[f_0]$, $\ln[\delta]$, $\ln[V_{S30}^{\text{geo}}]$, $\ln[V_{S30}^{\text{SASW}}]$, $\ln[a^{\text{SRI}}(T)]$, $\ln[a^{\text{SH1D}}(T)]$. The value of r is reported in the top-left corner of each subfigure, and the *P*-value for the null hypothesis that $r = 0$ is reported in the bottom-right corner. Table 3 gives values of r for additional periods.

Figures 2-5 and Table 3 show clear trends in term of the performance of the different explanatory variables. Site response is positively correlated with f_0 at small periods and negatively correlated with f_0 at larger periods. Because of this gradual transition, both r and b_1 become statistically insignificant for intermediate periods (see the Discussion section for more details).

In contrast, $|r|$ for all of the SASW-derived explanatory variables is greater at larger periods (with the exception of the small values at $T = 10$ sec). The period-dependence of the

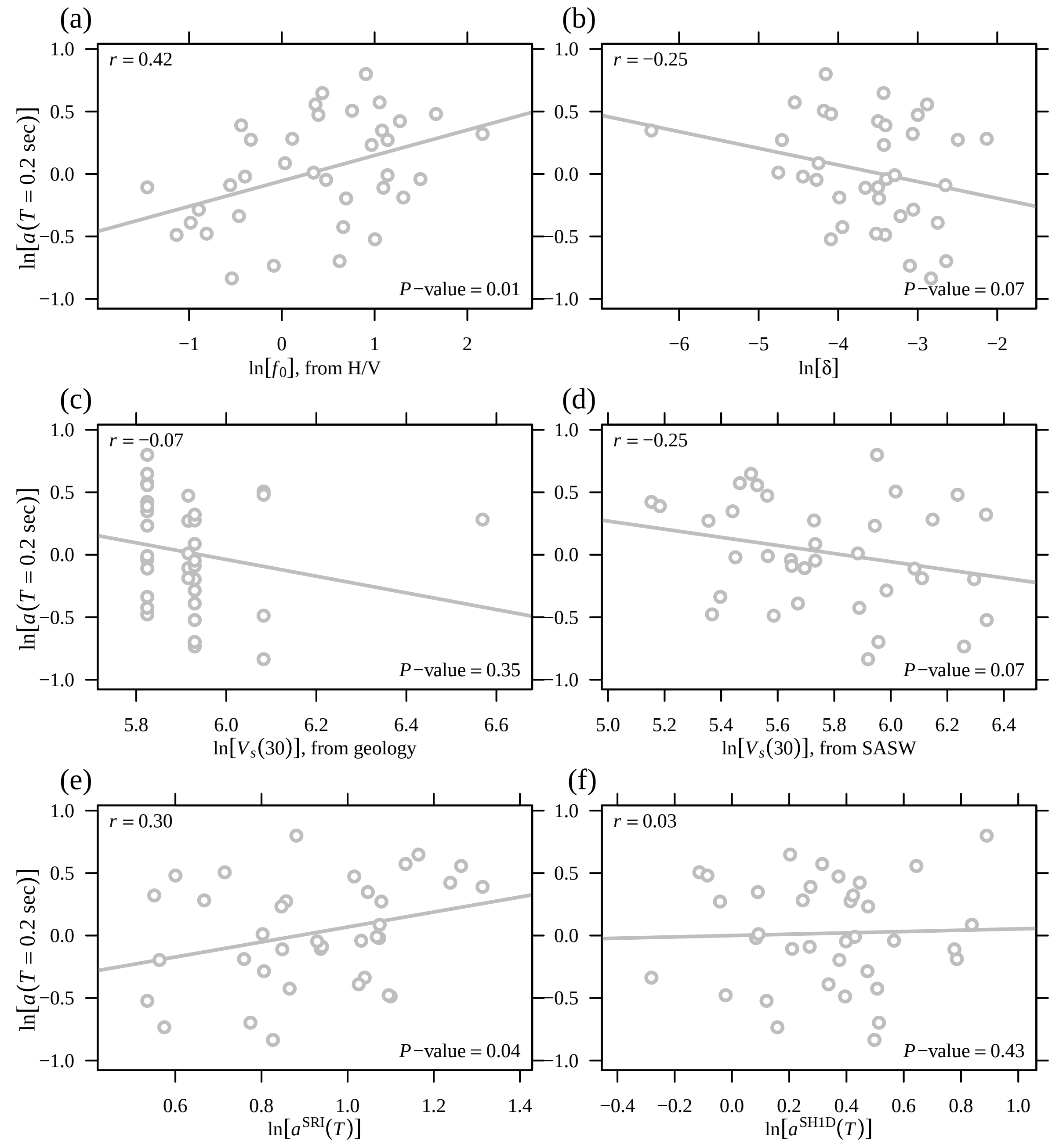

Figure 2: Correlation of $\ln[a(T = 0.2 \text{ sec})]$ with (a) $\ln[f_0]$, (b) $\ln[\delta]$, (c) $\ln[V_{S30}^{\text{geo}}]$, (d) $\ln[V_{S30}^{\text{SASW}}]$, (e) $\ln[a^{\text{SRI}}(T)]$, and (f) $\ln[a^{\text{SH1D}}(T)]$.

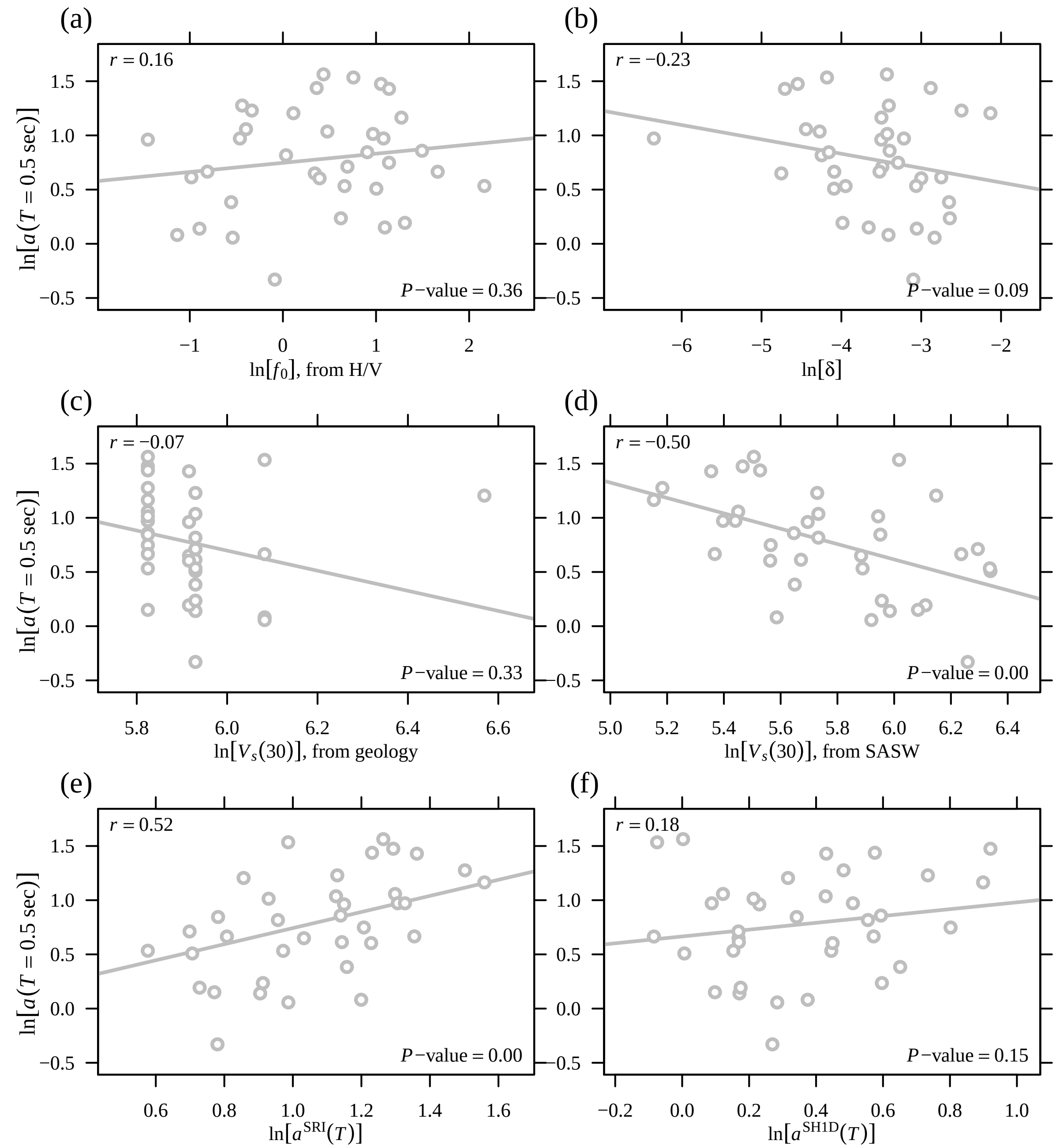

Figure 3: Correlation of $\ln[a(T = 0.5 \text{ sec})]$ with (a) $\ln[f_0]$, (b) $\ln[\delta]$, (c) $\ln[V_{S30}^{\text{geo}}]$, (d) $\ln[V_{S30}^{\text{SASW}}]$, (e) $\ln[a^{\text{SRI}}(T)]$, and (f) $\ln[a^{\text{SH1D}}(T)]$.

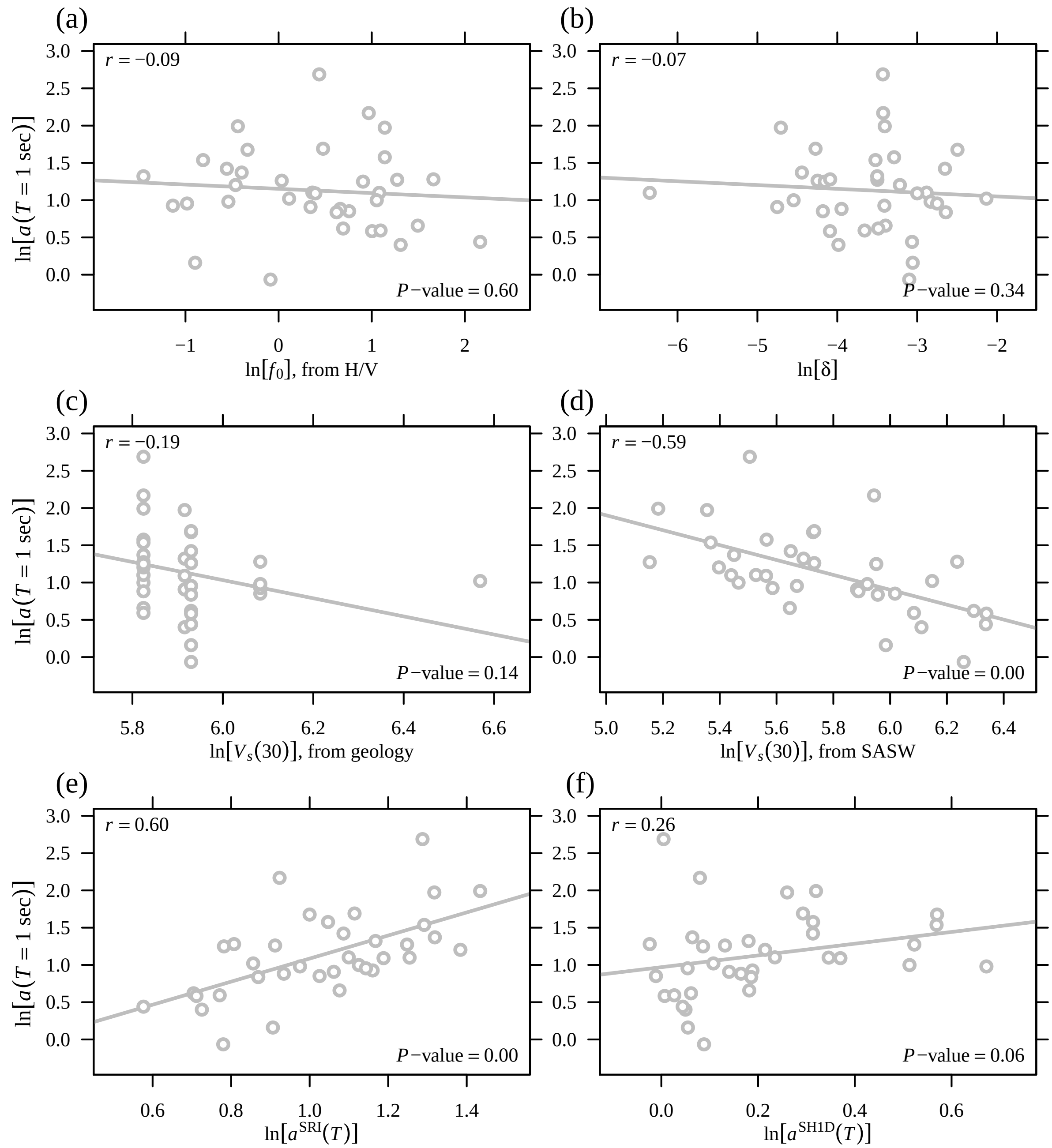

Figure 4: Correlation of $\ln[a(T = 1.0\text{ sec})]$ with (a) $\ln[f_0]$, (b) $\ln[\delta]$, (c) $\ln[V_{S30}^{\text{geo}}]$, (d) $\ln[V_{S30}^{\text{SASW}}]$, (e) $\ln[a^{\text{SRI}}(T)]$, and (f) $\ln[a^{\text{SH1D}}(T)]$.

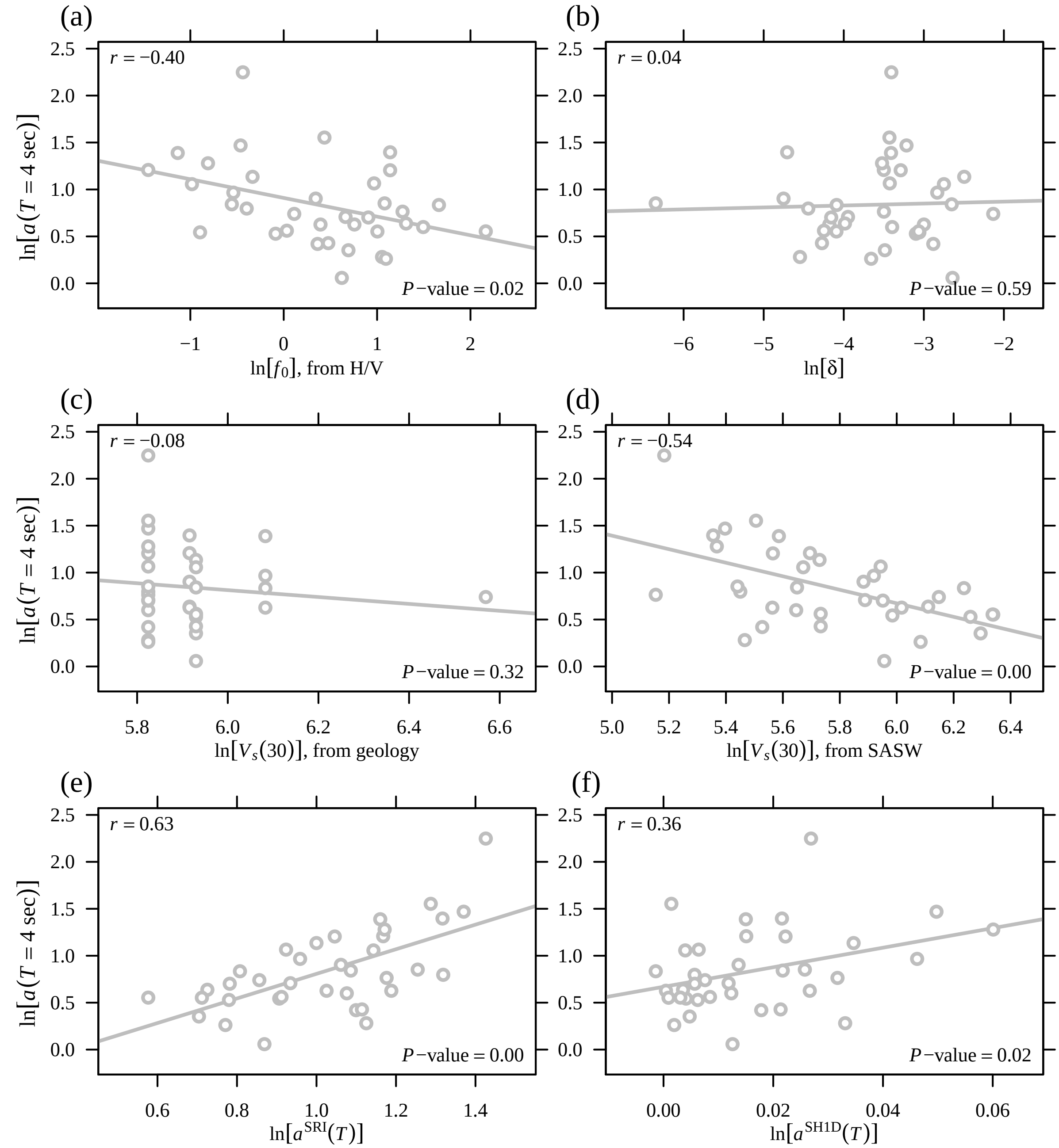

Figure 5: Correlation of $\ln[a(T = 4.0 \text{ sec})]$ with (a) $\ln[f_0]$, (b) $\ln[\delta]$, (c) $\ln[V_{S30}^{\text{geo}}]$, (d) $\ln[V_{S30}^{\text{SASW}}]$, (e) $\ln[a^{\text{SRI}}(T)]$, and (f) $\ln[a^{\text{SH1D}}(T)]$.

correlation of $\ln[\delta]$ and $\ln[V_{S30}^{\text{geo}}]$ is less clear: The performance of $\ln[\delta]$ is relatively constant with respect to period except that the performance is particularly poor for $T \geq 1.0$ sec, while $\ln[V_{S30}^{\text{geo}}]$ performs best at $T = 0.75$ sec but does not exhibit a clear performance trend.

To avoid severe colinearity, we should eliminate all but one of the explanatory variables derived from the SASW data: $\ln[a^{\text{SH1D}}(T)]$ is easily dismissed, but $\ln[V_{S30}^{\text{SASW}}]$ and $\ln[a^{\text{SRI}}(T)]$ perform nearly equally well at most periods. $\ln[a^{\text{SRI}}(T)]$ has a slight edge, which is most significant for $T \geq 2$ sec, and so we drop $\ln[V_{S30}^{\text{SASW}}]$ from our subsequent analysis.

We are then left with four independent explanatory variables that we use to map $a(t)$, each exhibiting distinct spatial characteristics: $\ln[f_0]$, $\ln[\delta]$, $\ln[V_{30}^{\text{geo}}]$, and $\ln[a^{\text{SRI}}(T)]$. Although $\ln[a^{\text{SRI}}(T)]$ generally exhibits the best correlation with $\ln[a(T)]$, the other explanatory variables have other distinct advantages. For example, $\ln[f_0]$ performs better than $\ln[a^{\text{SRI}}(T)]$ for $0.1 \leq T \leq 0.4$. But most importantly, the accuracy of both $\ln[a^{\text{SRI}}(T)]$ and $\ln[f_0]$ will diminish as distance increases from the observed values. Thus, $\ln[\delta]$ and $\ln[V_{30}^{\text{geo}}]$ should be included for locations distant from the SASW V_s profiles and strong motion stations.

4.2 Mapping Explanatory Variables

The raster and polygon data (δ and V_{S30}^{geo}) do not require further analysis for mapping because these values are spatially continuous throughout the region. In contrast, we must interpolate the point data [f_0 and $a^{\text{SRI}}(T)$]. Reliable interpolation with the kriging method is only possible if $\gamma(\mathbf{h})$ increases with $\mathbf{h}$, demonstrating that points that are closer together are more similar than those that are further apart. Figure 6 shows $\hat{\gamma}(\mathbf{h})$ and $\gamma(\mathbf{h};\boldsymbol{\theta})$ for (a) f_0, (b) the average *S*-wave slowness (S_s; inverse of V_s) to 30 m depth (S_{S30}), and (c) $a^{\text{SRI}}(T = 0.5)$ sec.

Here, we must explain why we display the $\hat{\gamma}(\mathbf{h})$ and $\gamma(\mathbf{h};\boldsymbol{\theta})$ for S_{S30} rather than V_{S30} in Figure 6 (b). Although V_{S30} is displayed in Figure 7, we do not krig V_{S30} directly — instead we krig S_{S30} and convert the result to V_{S30}. This is analogous to how V_{S30} is computed from a one-dimensional V_s profile: the S_{S30} is computed and then converted to V_{S30} (see discussions in Boore and Thompson, 2007; Brown *et al.*, 2002, for the motivation of using S_s as the primary variable of interest). Figure 6 demonstrates that kriging is an effective method for spatially interpolating these variables.

Figure 7 compares maps of the different explanatory variables, converted to V_{S30} when applicable. The three different maps of V_{S30} (subfigures b-d) use a single color palette. This figure fulfills two purposes: (1) visualize the spatial variability of the different explanatory variables, and (2) contrast the different methods of V_{S30} estimation. Although $a^{\text{SRI}}(T)$ is not included, the spatial structure of $a^{\text{SRI}}(T)$ is very similar to V_{S30}^{SASW}, with minor variations as a function of T. For the f_0 and V_{S30}^{SASW} maps, we use only the kriged estimates where σ_k^2 is less than the sample variance of the measured data.

The resulting maps of $\hat{a}(T = 0.5 \text{ sec})$ are displayed in Figure 8. Each subfigure uses the same color palette to visually highlight the differences. Analogous maps can easily be created for other periods, but we only display one period as an illustrative example due to space constraints. Note that the $\hat{a}(T)$ for $X = \ln[f_0]$ [$\hat{a}^{f_0}(T)$] and $\hat{a}(T)$ for $X = \ln[a^{\text{SRI}}(T)]$ [$\hat{a}^{\text{SRI}}(T)$] capture a lot of variability near the stations. In contrast, the resolution of $\hat{a}(T)$ for $X = \ln[V_{S30}^{\text{geo}}]$ [$\hat{a}^{\text{geo}}(T)$] and $\hat{a}(T)$ for $X = \ln[\delta]$ [$\hat{a}^{\delta}(T)$] does not vary with distance from the data. This highlights the major challenge that must be addressed to map site response: the

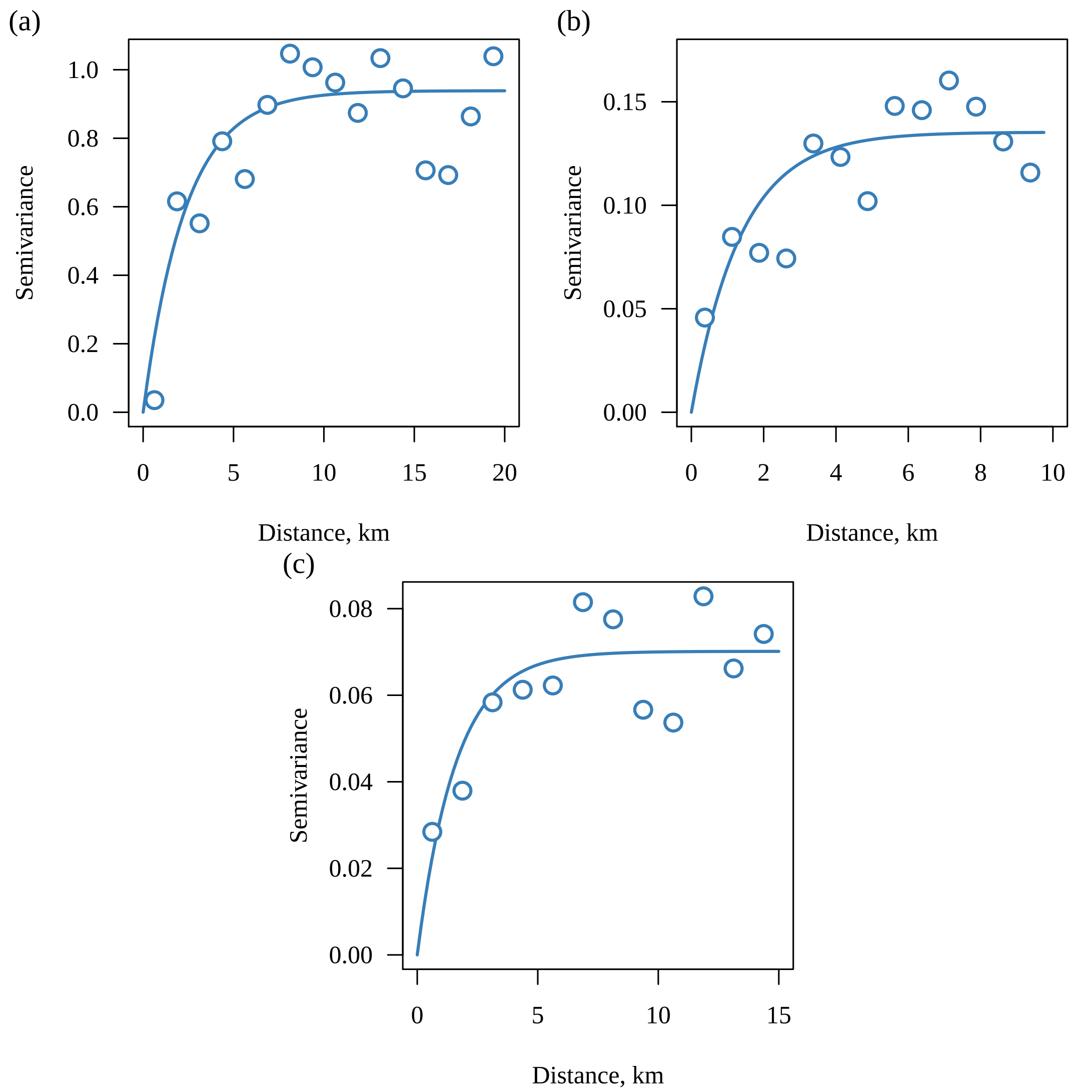

Figure 6: (a) Semivariogram of $\ln[f_0]$ for the 36 Parkfield sites that recorded the 1983 Coalinga earthquake. (b) Semivariograms of $\ln[S_{S30}]$ at the 52 SASW sites. (c) Semivariogram of $\ln[a^{\mathrm{SRI}}(T = 0.5 \text{ sec})]$ at the 52 SASW sites. Points are $\hat{\gamma}(\mathbf{h})$ and the lines are $\gamma(\mathbf{h}; \boldsymbol{\theta})$.

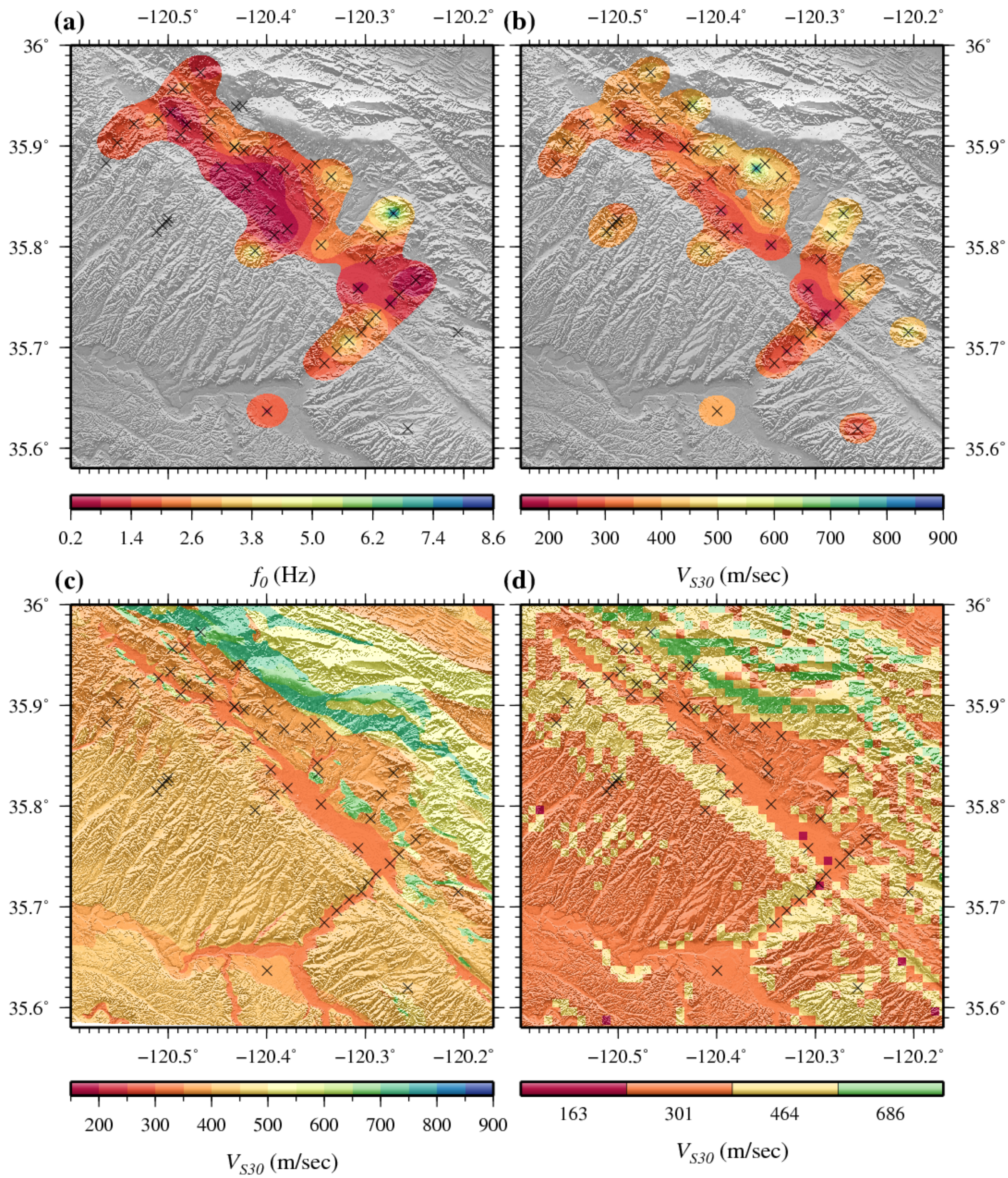

Figure 7: Maps of alternative site response explanatory variables. (a) the fundamental frequency f_0 (Zhao *et al.*, 2006), (b) V_{S30} estimated from SASW measurements (Thompson *et al.*, 2010b), (c) V_{S30} estimated from surficial geology (Wills and Clahan, 2006), and (d) V_{S30} estimated from topographic slope (Wald and Allen, 2007).

performance of the different explanatory variables depends both on location (distance to the observed data) as well as T (as summarized in Table 3).

Figure 9 plots (a) $s^2_{f_0}(T)$, (b) $s^2_{\text{SRI}}(T)$, (c) $s^2_{\text{geo}}(T)$, and (d) $s^2_{\delta}(T)$, again using a single color palette throughout. These figures show that for $T = 0.5$ sec, $\hat{a}^{\text{SRI}}(T)$ has the lowest prediction variance, which is consistent with the fact that it has the largest absolute value of r at this period (see Table 3). The second lowest prediction variance is achieved by $\hat{a}^{\delta}(T)$, which is again consistent with the r values reported in Table 3. The prediction variance of $\hat{a}^{\text{geo}}(T)$ is clearly a function of geologic unit, and is smaller than the prediction variance of $\hat{a}^{f_0}(T)$ in some location.

Figure 10 plots $\hat{a}^{\text{vwe}}(T)$ for (a) $T = 0.2$ sec, (b) $T = 0.5$ sec, (c) $T = 1.0$ sec, and (d) $T = 4.0$ sec. Note that we can recognize which explanatory variables are controlling the spatial variability by comparison to Figure 7. Note that for $T = 0.5$ sec, we can additionally compare the combined result to the predictions of explanatory variables in Figure 8. We see that the variation of $\hat{a}^{\text{vwe}}(T = 0.5 \text{ sec})$ in the vicinity of the stations is controlled by $\hat{a}^{\text{SRI}}(T)$, and in regions distant from the stations is controlled by $\hat{a}^{\delta}(T)$. This is the result that we should expect given the prediction variances in Figure 9 and the r values in Table 3. We also see that for $T = 1.0$ sec and $T = 4.0$ sec, δ contributes less to $\hat{a}^{\text{vwe}}(T)$ than the Jennings (1958) geology polygons at locations distant from the stations. This is consistent with the decrease in r values for $X = \ln[\delta]$ at the larger periods in Table 3.

4.3 NGA Database and Topographic Slope

We represent the performance of the regression of $\ln[a(T)]$ against $\ln[\delta]$ for the NGA database (equation 10) by plotting $\ln[\hat{a}(T)]$ versus $\ln[a(T)]$ in Figure 11 for (a) the peak ground acceleration (PGA) and (b) $T = 4.0$ sec. For comparison, we also plot $\ln[\hat{a}(T)]$ versus $\ln[a(T)]$ where $\ln[\hat{a}(T)]$ is computed with the Wald and Allen (2007) method. These plots qualitatively show that equation 10 accounts for more variability in the site response term than the Wald and Allen (2007) method.

The results are more thoroughly summarized in Table 4 which includes all the periods used by BA08. Table 4 includes the number of records included at each T, the estimates of the regression parameters and their *P*-values, and $\text{MSE}^{\delta}(T)$, $\text{MSE}^{\text{WA07}}(T)$, and $\text{MSE}^{\text{BA08}}(T)$. As expected, the values of b_1 and b_2 are all negative indicating that amplification is inversely related to topographic slope and the reference ground motion intensity. Also, we see that the magnitude of b_2 tends decreases as T increases, indicating that the degree of soil nonlinearity effects is larger for smaller period motion.

In terms of MSE, the Wald and Allen (2007) method slightly outperforms the BA08 site response model for $T \leq 0.25$ sec (by 1 to 3%) and the BA08 model performs better at other periods; The $\text{MSE}^{\text{WA07}}(T)$ is generally 8 to 14% larger than $\text{MSE}^{\text{BA08}}(T)$ for $0.4 \leq T \leq 5$ sec. In contrast, $\text{MSE}^{\delta}(T)$ is 10 to 13% smaller than $\text{MSE}^{\text{BA08}}(T)$ for $T \leq 0.15$ sec (including PGA) and is 6 to 8% smaller for $T \geq 3$ sec. Equation 10 performs poorest relative to the BA08 model for $0.5 \leq T \leq 1.5$ sec, which $\text{MSE}^{\delta}(T)$ is 5 to 8% larger than $\text{MSE}^{\text{BA08}}(T)$. $\text{MSE}^{\delta}(T)$ is smaller than $\text{MSE}^{\text{WA07}}(T)$ for all T except for $T = 0.5$ sec where the differences is approximately 1%. $\text{MSE}^{\delta}(T)$ is 11% smaller than $\text{MSE}^{\text{WA07}}(T)$ for PGA and 13% smaller for PGV.

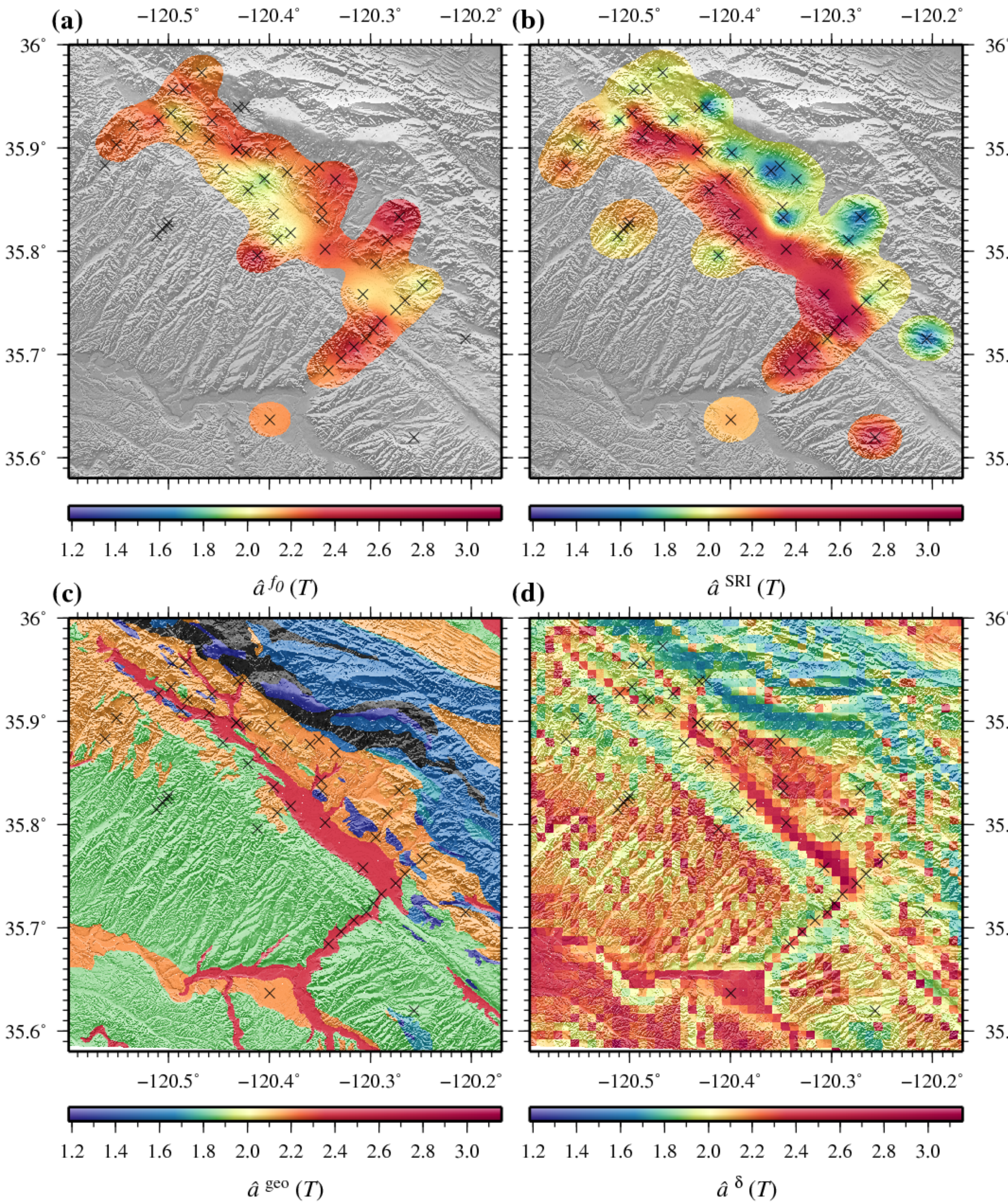

Figure 8: Maps of four different estimates of $a(T = 0.5 \text{ sec})$: (a) $\hat{a}^{f_0}(T)$, (b) $\hat{a}^{\mathrm{SRI}}(T)$, (c) $\hat{a}^{\mathrm{geo}}(T)$, and (d) $\hat{a}^{\delta}(T)$.

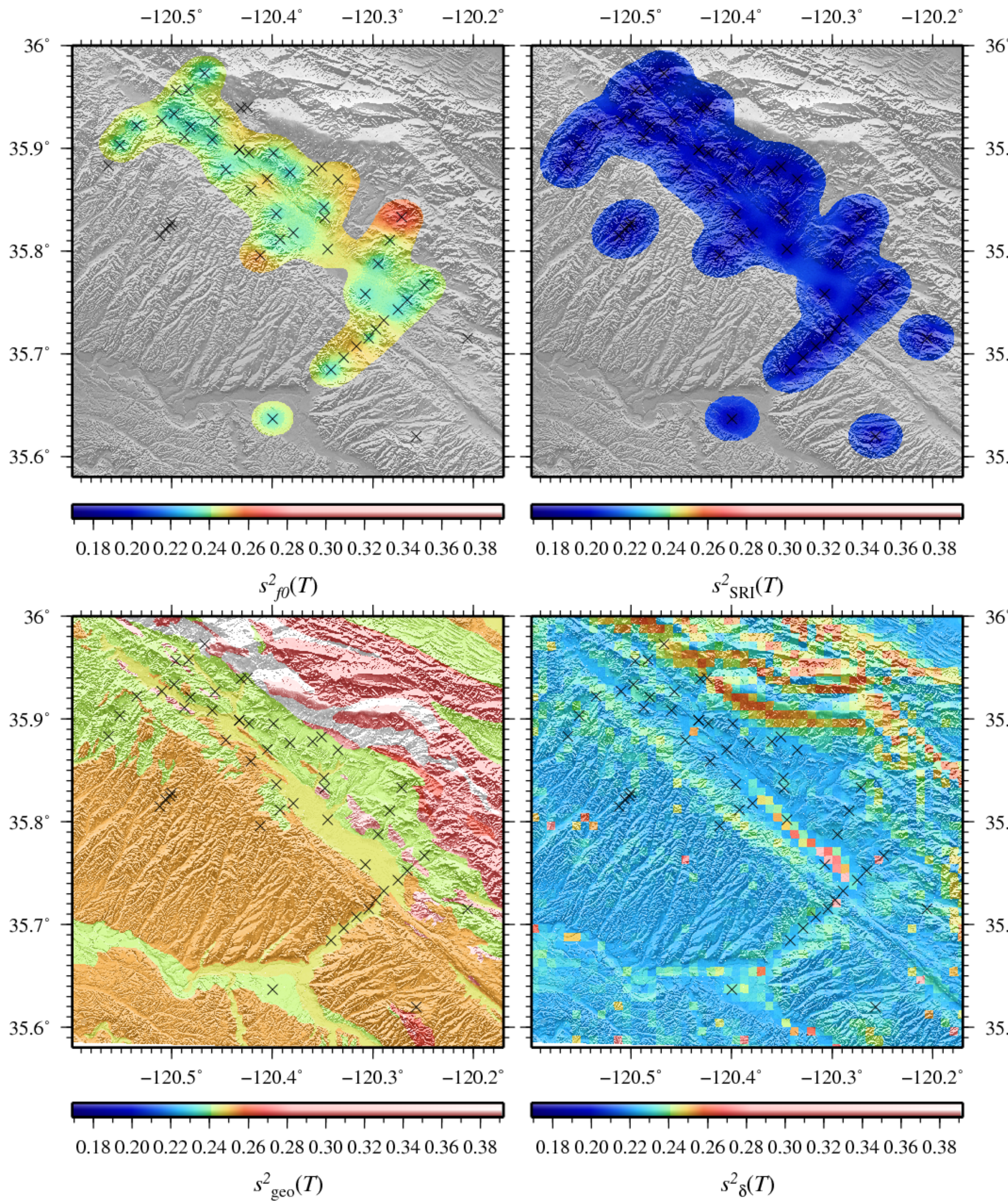

Figure 9: Maps of the variance of the $a(T = 0.5 \text{ sec})$ estimates in Figure 8: (a) $s^2_{f_0}(T)$, (b) $s^2_{\mathrm{SRI}}(T)$, (c) $s^2_{\mathrm{geo}}(T)$, and (d) $s^2_{\delta}(T)$.

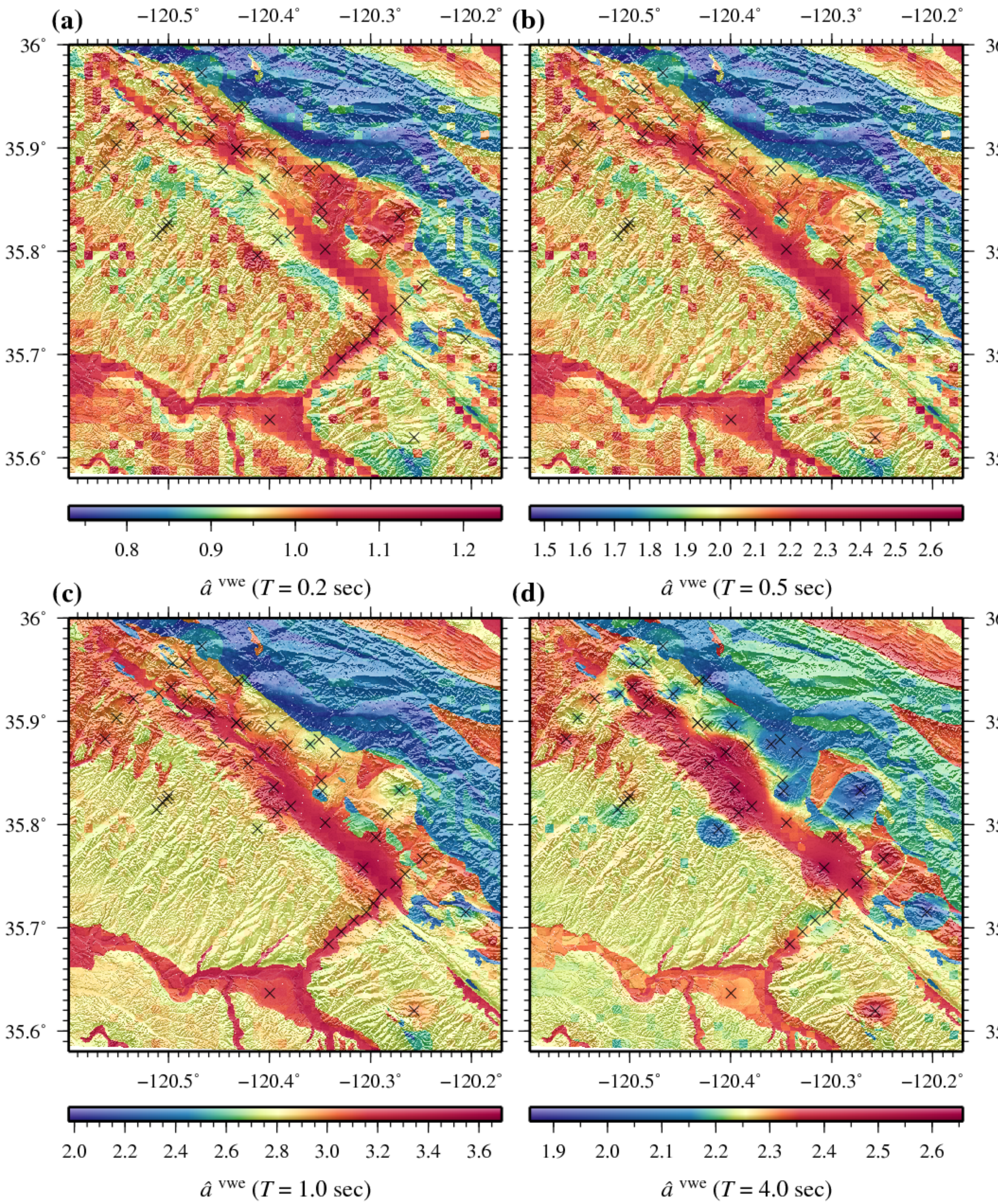

Figure 10: Map of the $\hat{a}^{\text{vwe}}(T)$, integrating the four different explanatory variables in Figure 8 for (a) $T = 0.2$ sec, (b) $T = 0.5$ sec, (c) $T = 1.0$ sec, and (d) $T = 4.0$ sec.

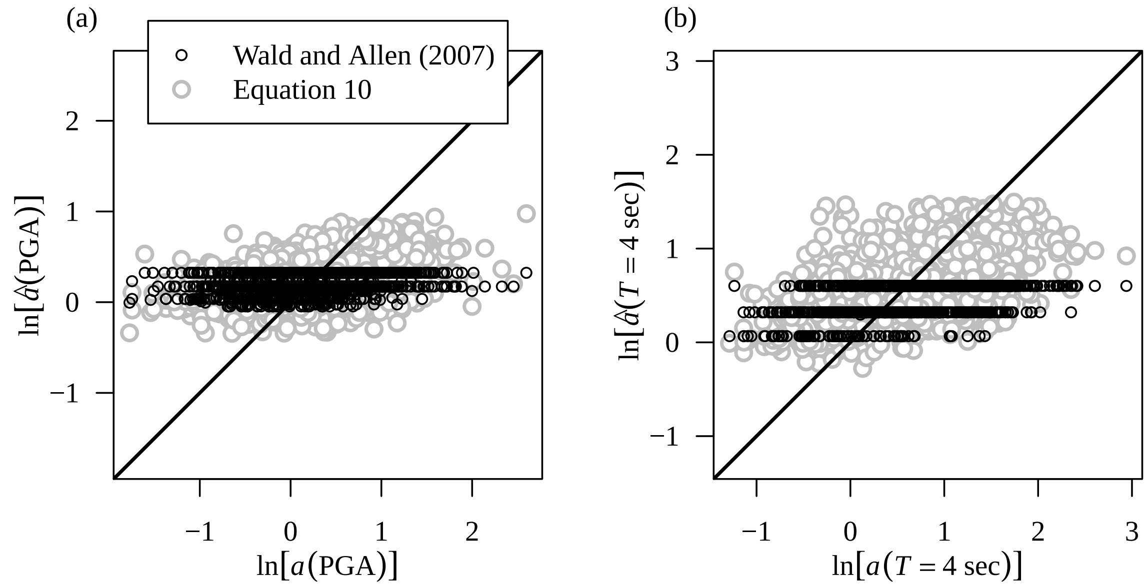

Figure 11: Plots of $\ln[\hat{a}(T)]$ versus $\ln[a(T)]$ for the Wald and Allen (2007) model compared and equation 10 for (a) PGA and (b) $T = 4.0$ sec.

Table 4: Summary of the regression for equation 10 and comparison to alternative models: The number of records available at each period, parameter estimates and their *P*-values, and MSE for eqation 10, the Wald and Allen (2007) model (WA07), and the BA08 model.

	No. of	Intercept		Slope		Intensity		MSE		
T, sec	Records	b_0	*P*-value	b_1	*P*-value	b_2	*P*-value	δ	WA07	BA08
PGA	1599	-0.530	0.000	-0.048	0.000	-0.187	0.000	0.318	0.357	0.365
PGV	1599	-0.003	0.936	-0.143	0.000	-0.085	0.000	0.293	0.336	0.302
0.010	1599	-0.530	0.000	-0.048	0.000	-0.188	0.000	0.320	0.360	0.367
0.020	1599	-0.529	0.000	-0.046	0.000	-0.185	0.000	0.325	0.363	0.372
0.030	1599	-0.518	0.000	-0.042	0.000	-0.186	0.000	0.337	0.377	0.387
0.050	1599	-0.493	0.000	-0.031	0.003	-0.190	0.000	0.363	0.406	0.417
0.075	1599	-0.471	0.000	-0.024	0.028	-0.186	0.000	0.388	0.433	0.443
0.100	1599	-0.415	0.000	-0.026	0.016	-0.177	0.000	0.398	0.430	0.446
0.150	1599	-0.374	0.000	-0.027	0.014	-0.181	0.000	0.393	0.423	0.437
0.200	1599	-0.242	0.000	-0.029	0.006	-0.139	0.000	0.362	0.374	0.383
0.250	1599	-0.206	0.000	-0.046	0.000	-0.119	0.000	0.351	0.363	0.364
0.300	1599	-0.148	0.005	-0.055	0.000	-0.094	0.000	0.359	0.373	0.361
0.400	1599	-0.076	0.154	-0.066	0.000	-0.066	0.000	0.356	0.380	0.347
0.500	1598	-0.015	0.789	-0.088	0.000	-0.036	0.006	0.375	0.373	0.357
0.750	1594	-0.079	0.187	-0.131	0.000	-0.024	0.058	0.399	0.408	0.374
1.000	1579	-0.126	0.043	-0.146	0.000	-0.028	0.023	0.404	0.424	0.377
1.500	1564	-0.247	0.000	-0.187	0.000	-0.012	0.303	0.410	0.444	0.388
2.000	1490	-0.342	0.000	-0.218	0.000	-0.003	0.803	0.410	0.460	0.412
3.000	1341	-0.523	0.000	-0.233	0.000	-0.028	0.015	0.375	0.435	0.399
4.000	1084	-0.404	0.000	-0.241	0.000	-0.009	0.447	0.371	0.460	0.403
5.000	1016	-0.374	0.000	-0.233	0.000	-0.012	0.334	0.392	0.480	0.426
7.500	898	-0.304	0.004	-0.196	0.000	-0.022	0.153	0.446	0.507	0.482
10.000	606	-0.417	0.004	-0.150	0.000	-0.064	0.008	0.429	0.456	0.460

5 Discussion

5.1 Alternative Site Response Variables

This paper has quantified the performance of many different commonly used site response explanatory variables using the Parkfield dataset. These variables should not only be judged in terms of correlations with observed amplifications, but also in terms of data availability; An approximate model that is more widely applicable may be more useful to the scientific and engineering community than a more accurate model that requires inputs that are difficult and expensive to estimate.

Note that we choose to use $X = \ln[\delta]$ rather than $X = \ln[V_{S30}^{\delta}]$. This is because the $\ln[V_{S30}^{\delta}]$ can only take relatively few discreet values in the Wald and Allen (2007) model. We feel that a fair assessment of the topographic slope method would allow the full range in variability of δ.

V_{S30} has become so ubiquitous as a site response explanatory variable that it is sometimes treated as the response (i.e., dependent) variable of interest, rather than just one possible explanatory (independent) variable (e.g., Pilz *et al.*, 2010). A proper assessment of the accuracy of an explanatory variable such as δ should compute correlations with some estimate of the observed site response amplifications rather than V_{S30}. Figures 2 to 5 and Table 3 clearly show that V_{S30} is a valuable but imperfect predictor of site response.

The main advantage of the SRI method over V_{S30} is that it can capture the period dependence of the site response amplifications. The values of r in Table 3 confirm this expectation: the r for both methods are identical at $T = 0.75$ sec, and the SRI method performs better than V_{S30} as the period diverges from this value (with an exception at $T = 0.05$ sec). The substantial decrease in r for $X = a^{\text{SRI}}(T)$ at $T = 10$ sec is expected because the V_s profiles do not extend to depths where $a^{\text{SRI}}(T = 10 \text{ sec})$ can be computed, and so the $a^{\text{SRI}}(T)$ for the largest possible T is used. The performance of the SRI method can be improved at larger periods by collecting profiles that extend to larger depths.

Space limitations require that we focus on a single T for Figures 8 and 9. These figures are necessary for visualizing the method that we propose, but are not helpful for judging the overall performance of one explanatory variable over another. We see that $s^2_{f_0}(T)$ is substantially larger than $s^2_{\text{SRI}}(T)$ and $s^2_{\delta}(T)$ in Figure 9. But notice that $|r|$ for $\ln[f_0]$ is larger at other periods, as indicated by Table 3 which will decrease the the variance of $\hat{a}^{f_0}(T)$. At most periods $\ln[f_0]$ slightly underperforms $\ln[a^{\text{SRI}}(T)]$, but f_0 is an easier explanatory variable to obtain for GMPE development: $\ln[a^{\text{SRI}}(T)]$ requires a V_s profile, which is costly and time consuming, while the computation of f_0 can be automated and only requires that a previous earthquake has been recorded at the station of interest. Thus, f_0 may be preferred over $a^{\text{SRI}}(T)$ in many applications.

5.2 Fundamental Site Frequency

The period dependence of b_1 for $X = \ln[f_0]$ is worth discussing further. Note that b_1 is closely related to r (Table 3) through a positive proportionality constant (though the magnitude of the constant will vary with period). Thus, it is clear that $b_1 < 0$ for larger periods and $b_1 > 0$ for smaller periods. One might expect that $b_1 < 0$ at all periods because larger values of f_0 are associated with stiffer soils, which are in turn associated with less site amplification. This is consistent for b_1 at $T \geq 1.0$ sec, but there is a clear gradual increase in b_1 as T decreases; For

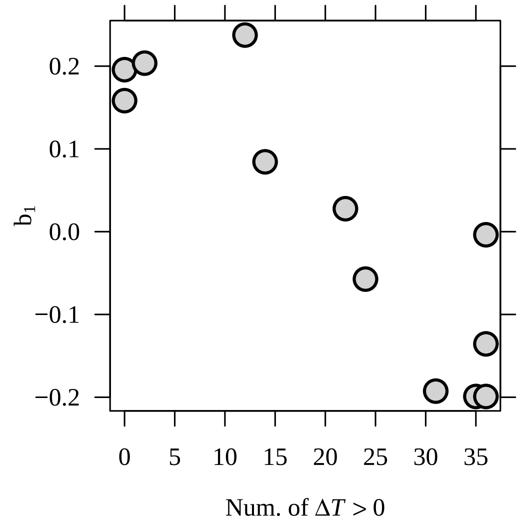

Figure 12: b_1 for $X = \ln[f_0]$ versus the number of sites that fulfill the inequality $\Delta T > 0$.

$T \leq 0.75$ sec, $b_1 > 0$.

This trend can be explained by defining the difference between the fundamental period of the soil ($T_0 = 1/f_0$) and the fundamental period of the SDOF system (T) for which the PSA is computed: $\Delta T = T - T_0$. As the absolute value of ΔT decreases, the resonance of the soil will align with the resonance of the SDOF so $a(T)$ will increase.

Assume $\Delta T < 0$ for all n observations of $a(T)$. Under this condition, ΔT decreases as T_0 increases (and f_0 decreases); Thus $a(T)$ increases as T_0 increases. The inequality $\Delta T < 0$ is more likely to be fulfilled for small values of T.

Now assume $\Delta T > 0$ for all n observations of $a(T)$. Under this condition, ΔT increases as T_0 increases (and f_0 decreases); Thus $a(T)$ decreases as T_0 increases. The inequality $\Delta T > 0$ is more likely to be fulfilled for large values of T.

It follows from these two scenarios that $b_1 < 0$ is expected for large values of T and $b_1 > 0$ is expected for small values of T. For intermediate cases, where the inequalities $\Delta T > 0$ or $\Delta T < 0$ are not true for all n observations of $a(T)$ will give a mixture of the two previously described scenarios. This is qualitatively supported by the trend of r with T for $X = \ln[f_0]$ in Table 3.

To check this in a more quantitative manner, we plot b_1 against the number of sites that fulfill the inequality $\Delta T > 0$ in Figure 12. The trend is what we would expect, showing that as the number of sites where $\Delta T > 0$ increases, the value of b_1 decreases. Note that when the number of sites where $\Delta T > 0$ is approximately $n/2$, then the $b_1 \approx 0$.

5.3 Topographic Slope

Figure 11 shows that directly regressing $\ln[\delta]$ against the observed amplifications in the NGA flatfile is more accurate than the Wald and Allen (2007) model. This improvement comes from removing the intermediate step of estimating V_{S30}, and then using the Borcherdt (1994) equations to estimate $a(T)$.

Table 4 shows the Wald and Allen (2007) model exhibits smaller MSE than the BA08 model for small periods ($T \leq 0.25$ sec). However, the MSE of the Wald and Allen (2007) model is significantly larger than the BA08 model for larger periods and PGV. But the improvements that we achieve by regressing $\ln[\delta]$ directly against $\ln[a(T)]$ have similar or smaller MSEs than the BA08 at all periods except $0.5 \leq T \leq 1.5$. This is consistent with the observations from Parkfield: This is the passband where V_{S30} performs the best at Parkfield (see Table 3). It is particularly encouraging that the regression results reported in Table 4 perform well for PGV because earthquake damage has been shown to be more strongly correlated with PGV than PGA Boatwright *et al.* (2001).

The relatively poor performance of Equation 10 at $0.5 \leq T \leq 1.5$ is a small sacrifice considering (1) the gain in spatial coverage and (2) the savings in terms of time and cost of field investigations that are required to achieve the V_{S30} values in the NGA database. The relatively good performance of Equation 10 leads us to recommend that the regression coefficients in Table 4 can be used to account for site effects in rapid response maps and probabilistic seismic hazards analysis.

The comparison to the BA08 amplifications is not straight forward because site response is estimated from many different sources including downhole V_s profiles and the Wills and Clahan (2006) site conditions map. In fact, only about 35% of the NGA records are associated with site specific measurements of the V_S profile, while the rest are inferred (Douglas *et al.*, 2009). As most in the engineering seismology community would likely expect, the Parkfield data demonstrate that V_{S30} estimated from an SASW survey is a more accurate predictor of site response than the surficial geology. Thus, the MSE reported for the BA08 model is likely a mixture of the model uncertainty with the estimation uncertainty. BA08 will most likely perform better when the estimate of V_{S30} is derived from a site specific measurement of the V_s profile. If these data were available then the MSE values reported in Table 4 for BA08 would be smaller. But this is not possible for hazard mapping, and so the more logical comparison to the topographic slope method would be to use only V_{S30} values that are inferred from the Wills and Clahan (2006) model. This increases the estimation uncertainty, and thus this would increase the MSE values reported in Table 4 for BA08 which would emphasize the value of the topographic slope method further.

6 Conclusions

A wide variety of methods have been proposed for mapping site response amplifications. Each has been developed with a different purpose in mind: Thompson *et al.* (2010a) sought to characterize a single urban region while Wills and Clahan (2006) sought a map with coverage of the entire state of California and Wald and Allen (2007) sought an estimate of site response for the entire globe. Each of these methods uses a different explanatory variable as a site response proxy. The methodology that we propose takes advantage of the inherent benefits of each of these different methods of mapping site response. Addressing the estimation and model uncertainty of these widely varying methods is only possible with the aid of the unique dataset of the densely spaced strong motion stations in Parkfield.

Our analysis shows that the Wald and Allen (2007) method of correlating site response with topographic slope performs surprisingly well considering the simplicity of the method and the spatial coverage that it attains. Indeed, topographic slope correlates better with the Coalinga amplifications than the surficial geology for $T \leq 0.5$. More accurate results can be achieved with local geophysical surveys of the near surface V_s profiles. Linking the V_s profiles to site response amplifications with the square-root-of-impedance method improves the correlation with observed site response amplifications compared to V_{S30}. Although this improvement is not extreme, the effort that is required to achieve this gain is minimal and thus we feel that the effort is warranted. We also find that f_0 provides complementary information to the V_s profile, especially for small periods. Interestingly, we see that f_0 exhibits a direct relationship with amplification for small periods and an inverse relationship with amplification for large periods.

Correlations of topographic slope with the site response amplifications observed in the NGA flatfile allow us to create new regression equations for estimating site response. We show that the Wald and Allen (2007) method agrees well with the observed site response in the NGA flatfile, and we show our newly developed regression equations are a significant improvement over the Wald and Allen (2007) method. These equations provide complete coverage of the globe at a pixel size of approximately 1 km by 1 km with only minimal sacrifice in the accuracy of the predictions compared to the Boore and Atkinson (2008) method.

Acknowledgments

This research is funded by National Earthquake Hazards Reduction Program (NEHRP) Award #G09AP00062. We thank Rich Vogel for valuable discussions and support on statistical modeling.

References

Abercrombie, R. E. (1997). Near-surface attenuation and site effects from comparison of surface and deep borehole recordings, *Bull. Seism. Soc. Am.* **87**(3), 731–744.

Abrahamson, N., and W. Silva (2008). Summary of the Abrahamson and Silva NGA ground-motion relations, *Earthquake Spectra* **24**(1), 67–97.

Abrahamson, N. A., and W. J. Silva (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seism. Res. Letters* **68**(1), 94 – 127.

Ambraseys, N. N. (1995). The prediction of earthquake peak ground acceleration in Europe, *Earthquake Engineering & Structural Dynamics* **24**(4), 467–490.

Boatwright, J., K. Thywissen, and L. C. Seekins (2001). Correlation of ground motion and intensity for the 17 January 1994 Northridge, California, Earthquake, *Bull. Seism. Soc. Am.* **91**(4), 739–752.

Boore, D. M. (2003). Simulation of ground motion using the stochastic method, *Pure Appl. Geophys.* **160**(3-4), 635–676.

Boore, D. M., and G. M. Atkinson (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s, *Earthquake Spectra* **24**(1), 99–138.

Boore, D. M., and W. B. Joyner (1997). Site amplifications for generic rock sites, *Bull. Seism. Soc. Am.* **87**(2), 327–341.

Boore, D. M., and E. M. Thompson (2007). On using surface-source downhole-receiver logging to determine seismic slownesses, *Soil Dyn. Earthquake Eng.* **27**(11), 971–985.

Borcherdt, R. D. (1994). Estimates of site-dependent response spectra for design (methodology and justification), *Earthquake Spectra* **10**, 617–654.

Brown, L. T., D. M. Boore, and K. H. Stokoe (2002). Comparison of shear-wave slowness profiles at 10 strong-motion sites from noninvasive SASW measurements and measurements made in boreholes, *Bull. Seism. Soc. Am.* **52**, 289–300.

Cadet, H., P.-Y. Bard, and A. Rodriguez-Marek (2010). Defining a Standard Rock Site: Propositions Based on the KiK-net Database, *Bull. Seism. Soc. Am.* **100**(1), 172–195, doi: {10.1785/0120090078}.

Campbell, K., and Y. Bozorgnia (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5elastic response spectra for periods ranging from 0.01 to 10 s, *Earthquake Spectra* **24**(1), 139–71.

Castellaro, S., F. Mulargia, and P. L. Rossi (2008). Vs30: Proxy for Seismic Amplification?, *Seism. Res. Letters* **79**(4), 540–543, doi:10.1785/gssrl.79.4.540.

Chiou, B. S.-J., and R. R. Youngs (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthquake Spectra* **24**(1), 173–215.

Cressie, N. A. C. (1993). *Statistics for spatial data*, John Wiley & Sons, Inc., New York, ISBN 0471843369.

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: a review of equations for the estimation of peak ground acceleration and response spectral ordinates, *Earth-Science Reviews* **61**(1-2), 43–104.

Douglas, J., P. Gehl, L. F. Bonilla, O. Scotti, J. Regnier, A.-M. Duval, and E. Bertrand (2009). Making the most of available site information for empirical ground-motion prediction, *Bull. Seism. Soc. Am.* **99**(3), 1502–1520.

Draper, N. R., and H. Smith (1981). *Applied Regression Analysis*, New York: Wiley, 709 pp.

Farr, T. G., and M. Kobrick (2000). Shuttle Radar Topography Mission produces a wealth of data, *EOS* **81**, 583585.

Fisz, M. (1963). *Probability theory and mathematical statistics*, New York: Wiley, 3rd edition, 677 pp.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. V. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). National seismic hazard maps: Documentation june 1996, *U.S. Geol. Surv. Open-File Rept.* **96-532**, 110.

Frankel, A. D., M. D. Petersen, C. S. Mueller, K. M. Haller, R. L. Wheeler, E. V. Leyendecker, R. L. Wesson, S. C. Harmsen, C. H. Cramer, D. M. Perkins, and K. S. Rukstales (2002). Documentation for the 2002 update of the national seismic hazard maps, *U.S. Geol. Surv. Open-File Rept.* **02-420**, 39.

Guttorp, P., and T. Gneiting (2005). On the Whittle-Matérn correlation family, Technical Report 81, University of Washington, NRCSE Technical Report Series.

Haskell, N. A. (1953). The dispersion of surface waves on multilayered media, *Bull. Seism. Soc. Am.* **72**, 17–34.

Holzer, T. L., M. J. Bennett, T. E. Noce, A. C. Padovani, and J. C. Tinsley (2002). Liquefaction hazard and shaking amplification maps of Alameda, Berkeley, Emeryville, Oakland, and Piedmont, California: A digital database, *U.S. Geological Survey Open File Report 02-296* .

Jennings, C. W. (1958). Geologic atlas of California, San Luis Obispo sheet, Calif. Div. Mines Geol. Geologic Atlas Map GAM 018, scale 1:250,000.

Joyner, W. B., R. E. Warrick, and T. E. Fumal (1981). The effect of quaternary alluvium on strong ground motion in the Coyote Lake, California, Earthquake of 1979, *Bull. Seism. Soc. Am.* **71**(4), 1333–1349.

Lermo, J., and F. J. Chávez-García (1993). Site effect evaluation using spectral ratios with only one station, *Bull. Seism. Soc. Am.* **83**(5), 1574–1594.

Liu, P., S. Custodio, and R. J. Archuleta (2006). Kinematic inversion of the 2004 M 6.0 Parkfield earthquake including an approximation to site effects, *Bull. Seism. Soc. Am.* **96**(4, Part B Suppl. S), S143–S158, doi:{10.1785/0120050826}.

Petersen, M. D., A. D. Frankel, S. C. Harmsen, C. S. Mueller, K. M. Haller, R. L. Wheeler, R. L. Wesson, Y. Zeng, O. S. Boyd, D. M. Perkins, N. Luco, E. H. Field, C. J. Wills, and K. S. Rukstales (2008). Documentation for the 2008 Update of the United States National Seismic Hazard Maps, *U.S. Geol. Surv. Open-File Rept.* **2008-1128**, 61.

Pilz, M., S. Parolai, M. Picozzi, R. Wang, F. Leyton, J. Campos, and J. Zschau (2010). Shear wave velocity model of the Santiago de Chile basin derived from ambient noise measurements: a comparison of proxies for seismic site conditions and amplification, *Geophys. J. Int.* **182**(1), 355–367, doi:{10.1111/j.1365-246X.2010.04613.x}.

Power, M., B. Chiou, N. Abrahamson, Y. Bozorgnia, T. Shantz, and C. Roblee (2008). An overview of the NGA project, *Earthquake Spectra* **24**(1), 3–21.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL http://www.R-project.org, ISBN 3-900051-07-0.

Shakal, A., H. Haddadi, V. Graizer, K. Lin, and M. Huang (2006). Some key features of the strong-motion data from the M 6.0 Parkfield, California, earthquake of 28 September 2004, *Bull. Seism. Soc. Am.* **96**(4, Part B Suppl. S), S90–S118, doi:{10.1785/0120050817}.

Steidl, J. H., A. G. Tumarkin, and R. J. Archuleta (1996). What is a reference site?, *Bull. Seism. Soc. Am.* **86**(6), 1733–1748.

Thompson, E. M., L. G. Baise, R. E. Kayen, Y. Tanaka, and H. Tanaka (2010a). A geostatistical approach to mapping site response spectral amplifications, *Engineering Geology* **in press**.

Thompson, E. M., R. E. Kayen, B. Carkin, and H. Tanaka (2010b). Surface Wave Site Characterization at 52 Strong Motion Recording Stations Affected by the Parkfield M6.0 Earthquake of 28 September 2010, *U.S. Geol. Surv. Open-File Rept.* **2010-XXXX**.

Thomson, W. T. (1950). Transmission of elastic waves through a stratified solid, *Journal of Applied Physics* **21**, 89–93.

Trifunac, M. D. (1976). Preliminary analysis of peaks of strong earthquake ground motion - dependence of peaks on earthquake magnitude, epicentral distance, and recording site conditions, *Bull. Seism. Soc. Am.* **66**(1), 189–219.

Wald, D. J., and T. I. Allen (2007). Topographic Slope as a Proxy for Seismic Site Conditions and Amplification, *Bull. Seism. Soc. Am.* **97**(5), 1379–1395.

Wald, D. J., P. S. Earle, K.-W. Lin, V. Quitoriano, and B. Worden (2006). Challenges in rapid ground motion estimation for the prompt assessment of global urban earthquakes, *Bull. Seism. Soc. Am.* **81**(3-4), 273–281.

Wald, D. J., B. C. Worden, V. Quitoriano, and K. L. Pankow (2005). *ShakeMap Manual: Technical Manual, User's Guide, and Software Guide*, U.S. Geological Survey, Techniques and Methods 12A1, http://pubs.usgs.gov/tm/2005/12A01/ (last accessed May 2010).

Wessel, P., and W. H. F. Smith (1991). Generic Mapping Tools, *EOS* **72**, 441.

Wills, C. J., and K. B. Clahan (2006). Developing a map of geologically defined site-condition categories for California, *Bull. Seism. Soc. Am.* **96**(4), 1483–1501.

Zhao, J., K. Irikura, J. Zhang, Y. Fukushima, P. Somerville, A. Asano, Y. Ohno, T. Oouchi, T. Takahashi, and H. Ogawa (2006). An empirical site-classification method for strong-motion stations in japan using h/v response spectral ratio, *Bull. Seism. Soc. Am.* **96**(3), 914–25.