



4th IASPEI / IAEE International Symposium:

Effects of Surface Geology on Seismic Motion

August 23–26, 2011 • University of California Santa Barbara

EFFECTS OF EARTHQUAKE SOURCE GEOMETRY AND SITE CONDITIONS ON SPATIAL CORRELATION OF EARTHQUAKE GROUND MOTION HAZARD

Jack W. Baker

Stanford University
Stanford, CA 94305
USA

Mahalia K. Miller

Stanford University
Stanford, CA 94305
USA

ABSTRACT

Predicting losses to infrastructure systems or portfolios of individual structures in future earthquakes requires an understanding of the variation in ground motion intensity that will be observed over the multiple sites at which the assets of interest are located. This spatial variation can be partially described by linear correlation coefficients in observed logarithmic spectral accelerations at a given pair of sites. That linear correlation can be further decomposed into “correlation in residuals” associated with correlated log spectral values (relative to predicted values) for a given earthquake, and “correlation in means” associated with trends in mean log spectral values at pairs of sites associated with varying earthquake magnitudes and locations of future earthquakes. These two sources can both be calibrated using observed earthquake data, and then combined to characterize overall spatial correlation. It will also be noted, however, that linear correlation is only a partial and imperfect descriptor of joint distributions of spectral values in some cases, and other descriptors such as conditional mean values may be more informative in some cases. Techniques to quantify this variation using observed strong ground motions from past earthquakes will be reviewed, and the utility of this type of analysis for probabilistic seismic risk assessments of spatially-distributed systems will be described. A variety of quantification and analysis approaches are discussed, and the relevance of each for various risk calculations is discussed.

INTRODUCTION

Assessment of risk to spatially distributed infrastructure is a topic of great interest in some parts of the earthquake engineering community. One challenge with these risk assessments is quantifying the spatial variability of ground motion intensities that will be experienced by the various components of the infrastructure system within a given earthquake (Lee and Kiremidjian 2007; Jayaram and Baker 2010), as it is known that the degree of variability can have a significant impact on predicted losses (Adachi and Ellingwood 2007; Lee and Kiremidjian 2007; Park et al. 2007; Shiraki et al. 2007; Sokolov and Wenzel 2010). This spatial variability results from the differing distances from each considered earthquake rupture to the locations of interest, differing site conditions at each location, and variability due to other effects such as source heterogeneity and wave scattering that are difficult to predict in a deterministic manner. The variability in ground motions at any single site is considered in its probabilistic seismic hazard curve, but when considering large numbers of sites for infrastructure risk calculations, more work is needed to characterize spatial correlations.

This paper will present an overview of formulations and observational data related to modeling of spatial variability of ground motions, using the same seismic source models and empirical ground motion prediction models used to produce current seismic hazard maps. Opportunities for refinements of these results using simulated ground motions and site response analyses are also briefly discussed. Ground motion will be characterized here by spectral acceleration at a specified period, $S_a(T)$, although other parameters could be used in the same manner as long as a corresponding ground motion prediction model and spatial correlation model are available (e.g., Foulser-Piggott and Stafford 2011).

To facilitate the discussion below, we introduce new terminology for several sources of spatial correlation. For a given earthquake, mean values and standard deviations of logarithmic spectral acceleration are predicted by ground motion prediction models (e.g., Boore and Atkinson 2008), which account for attenuation of spectral accelerations with distance, as well as the effect of earthquake magnitude, rupture mechanism, site conditions, etc. The difference between observed $\ln Sa$'s and mean predictions is characterized by a prediction residual, and those residuals are seen to have a relatively stable pattern of spatial correlations in well-recorded past earthquakes. We term this effect *correlation in residuals*. When considering future ground motions at pairs of sites in a region, the size and location of a given future earthquake is not known with certainty, but of course both sites will experience shaking from the same earthquake. The commonality in earthquake magnitude, and potential similarities in source-to-site distances and site conditions at the pair of sites will contribute what we term a *correlation in means*. Overall future ground motions at pairs of sites can be characterized by a *total correlation* that includes the effect of both above sources of correlation. These terms will be defined more formally below. These distinctions are arguably artificial, as they are dependent upon the seismic hazard model formulation, but they are nonetheless useful as they decompose the problem into two pieces that can be analyzed and calibrated separately, and the two sources of correlation correspond approximately with separate physical phenomena. The results presented below aim to illustrate the insights that can be gained by looking at the problem in this way. Additionally, the results presented below show that correlation coefficients are only a partial and imperfect descriptor of joint behavior of spectral values at two sites, and preliminary ideas for more detailed descriptions are discussed.

To illustrate the calculations proposed here, numerical results will be presented for the San Francisco Bay Area shown in Figure 1, a dense urban area with many seismically active faults. To illustrate calculations for specific pairs of locations, results will be shown for three example sites (Berkeley Hills, Milpitas and Stanford foothills) also shown in Figure 1.

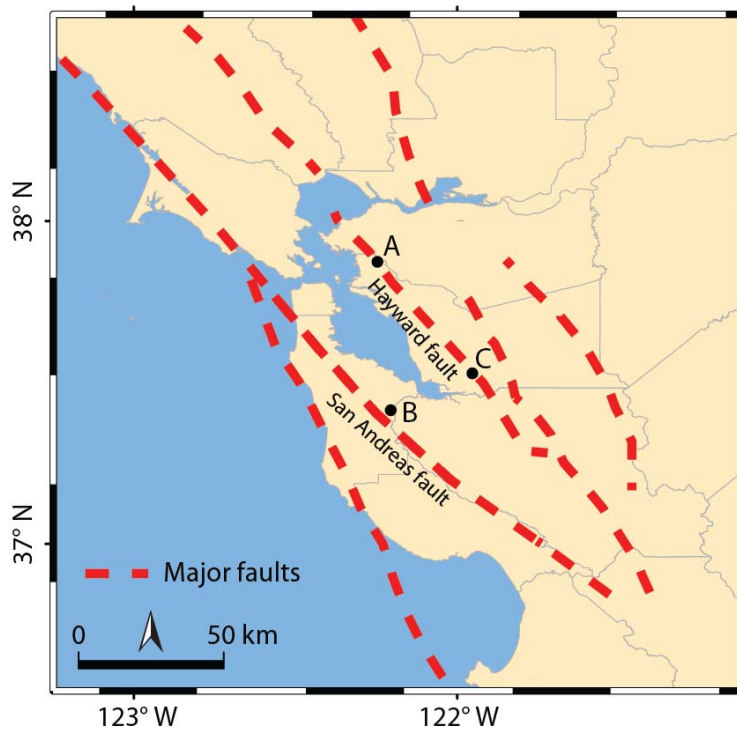


Figure 1. Major seismic faults in the San Francisco Bay Area, and three locations to be used in example calculations below: A – Berkeley, B – Stanford and C – Milpitas.(adapted from Jayaram and Baker 2010).

EMPIRICAL PREDICTION OF STRONG GROUND MOTION

Spatial correlations in ground motion intensity for a given earthquake can be measured empirically from well-recorded past earthquakes. Consider the following popular model for spectral accelerations used by empirical ground motion prediction models (Abrahamson and Youngs 1992; Al Atik et al. 2010)

$$\ln Sa_{ij} = \overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)} + \sigma_{ij}\varepsilon_{ij} + \tau_j\eta_j \quad (1)$$

where Sa_{ij} is the spectral acceleration at period T for site i in earthquake j , $\overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)}$ is the predicted mean value of $\ln Sa_{ij}$, ε_{ij} is the within-event residual representing site-to-site variability in $\ln Sa_{ij}$, η_j is the between-event residual representing event-to-event variability, and σ_{ij} and τ_j are the standard deviations of the within-event and between event residuals, respectively. For a given earthquake magnitude and rupture geometry, $\overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)}$, σ_{ij} and τ_j are deterministic values provided by the ground motion prediction model for every location of interest, and ε_{ij} and η_j are normally distributed random variables with mean zero and unit standard deviation. The η_j will be constant for all sites, while the ε_{ij} will vary from site to site but be correlated for pairs of nearby sites. For a given event, spatial correlations in $\ln Sa$ across the region are thus equivalent to the spatial correlation in the within event residual ε_{ij} . When a specific earthquake scenario is not known, however, correlations are more complex as the $\overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)}$ value in particular will be unknown but spatially correlated. We examine these two sources of uncertainty more carefully in the following subsections.

Correlation in Residuals

We first consider the case of a specified earthquake and rupture geometry, where spatial correlations in the within event residual ε_{ij} completely specifies the spatial correlation in $\ln Sa$. We term this source of correlation “*correlation in residuals*,” and discuss here how its effect can be measured empirically.

The spatial correlations of ε_{ij} can be evaluated from dense observations of strong ground motion in past earthquakes. This is done by first computing ε_{ij} values for each observed ground motion. Since Sa_{ij} is known for a given recording, and the other terms can be computed using a ground motion prediction model, equation (1) can be rearranged to solve for ε_{ij}

$$\varepsilon_{ij} = \left(\ln Sa_{ij} - \overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)} - \tau_j\eta_j \right) / \sigma_{ij} \quad (2)$$

Correlations between ε_{ij} values at a given separation distance can then be estimated by pooling pairs of ε_{ij} values with that separation distance (within some tolerance), and computing an empirical correlation coefficient

$$\hat{\rho}(h) = \frac{1}{n \cdot s_{\varepsilon_j}^2} \sum_{\forall |\varepsilon_{ij} - \varepsilon_{kj}| = h} (\varepsilon_{ij} - \bar{\varepsilon}_j)(\varepsilon_{kj} - \bar{\varepsilon}_j) \quad (3)$$

where $\hat{\rho}(h)$ is the estimated within-event correlation at separation distance h , $\sum_{\forall |\varepsilon_{ij} - \varepsilon_{kj}| = h}$ denotes a summation over all ε_{ij} and ε_{kj} pairs that are separated by distance h , n is the number of such pairs, and $\bar{\varepsilon}_j$ and $s_{\varepsilon_j}^2$ denote the sample mean and variance of all ε_{ij} values from the j th earthquake. These paired ε values are illustrated in Figure 2a, and the resulting correlation coefficients for the 1999 Chi-Chi, Taiwan earthquake as a function of h are illustrated in Figure 2b. A predictive equation can then be fit to the observed correlation coefficients, as is also shown in Figure 2b.

The estimation approach of equation (3) assumes that correlation in ε_{ij} values between two sites is dependent only on the sites' separation distance and not on other factors, allowing all observations of ε_{ij} pairs with the same separation distance to be pooled and used to estimate a correlation coefficient. This requires that ε_{ij} be independent of the mean predictions (i.e., any site effects, distance attenuation, etc. have been accounted for in $\overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)}$ to the extent possible). This is arguably a reasonable assumption, as ground motion modelers always evaluate their predictions to verify this independence, and if some dependence is identified then it is addressed by revising the predictive model. This also requires the source of within-event correlation to be stationary and isotropic; that assumption is more questionable, as the likely sources of this correlation (similar location to asperities, similar wave propagation paths and similar site effects) may not be stationary and isotropic, but the assumption has not yet been invalidated empirically (Jayaram 2010). Given this empirical reasonableness, approaches similar to that described above have been used by several researchers in the past to quantify these correlations (Boore et al. 2003; Wang and Takada 2005; Goda and Hong 2008; Jayaram and Baker 2009), and more procedural details are available in those documents.

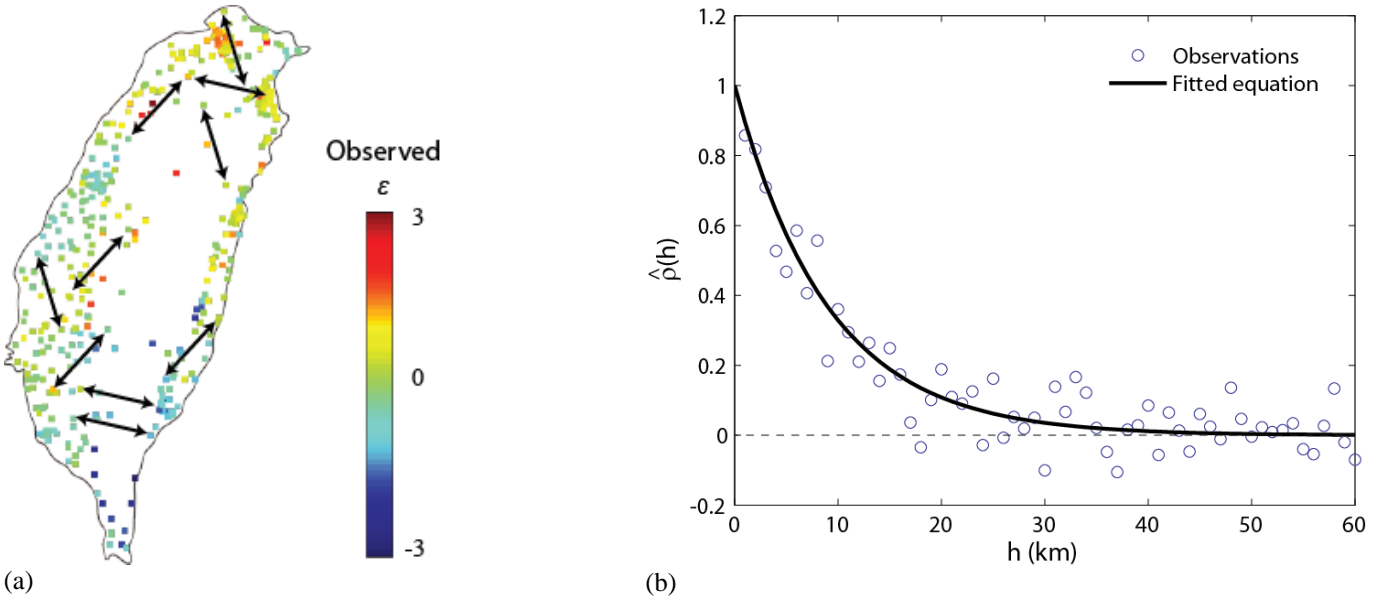


Figure 2: (a) Observed within-event residuals from the 1999 Chi-Chi, Taiwan earthquake, and arrows illustrating paired observations at a given distance used to estimate within-event correlation at that distance. (b) Computed correlation coefficients as a function of separation distance, and fitted equation for predicting correlations in future earthquakes.

Correlation in Means

The calculations of the previous section quantify $\ln Sa$ correlations for a given earthquake with known magnitude and rupture geometry, using prediction residuals from recorded ground motion data where the magnitude and rupture geometry are known. But the magnitudes and locations of future earthquakes are also random and will contribute to correlation in ground motion intensities. We can see this by considering equation (1), evaluated for two sites indexed as $i=1$ and $i=2$:

$$\ln Sa_{1j} = \overline{\ln Sa(M_j, R_{1j}, V_{s30,1}, T, \dots)} + \sigma_{1j} \varepsilon_{1j} + \tau_j \eta_j \quad (4)$$

$$\ln Sa_{2j} = \overline{\ln Sa(M_j, R_{2j}, V_{s30,2}, T, \dots)} + \sigma_{2j} \varepsilon_{2j} + \tau_j \eta_j \quad (5)$$

The $\overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)}$ term will now contribute uncertainty to values of $\ln Sa_{ij}$ that will be observed in future earthquakes, because the inputs to that function (e.g. M and R) are now random variables. In a given earthquake, the magnitude M_j will be identical for both sites, so that will create a source of high correlation in $\ln Sa$ values at the two sites (loosely speaking, the fact that the two sites will both experience either a high magnitude or low magnitude in a given earthquake will tend to make the resulting $\ln Sa$ values highly correlated). The distance values R_{ij} will in general differ for the two sites, and the degree of difference will depend upon the orientation of the two sites relative to the earthquake sources, as we will study later. Finally, the V_{s30} values for the sites may differ, although they are likely to be spatially correlated due to commonalities in geologic conditions at nearby sites (Thompson et al. 2007). Additional predictor parameters, such as hanging wall and foot wall terms, or depth to bedrock differ as well, but the effect of these terms are not discussed here or included in the notation for the sake of brevity.

The induced correlation caused by commonality in earthquake events and in site conditions is here termed “*correlation in means*” as it manifests itself in correlation of the mean $\overline{\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)}$ term when empirical ground motion prediction models are used to model distributions of future ground motion intensities.

NUMERICAL RESULTS AND JOINT DISTRIBUTIONS OF SPECTRAL VALUES

We can study the combined effect of the above two sources of spatial correlation by performing Monte Carlo simulation to generate a synthetic catalog of future ground motion intensities in a region. To do this, we first simulate a set of earthquake events using a seismic source model (the identical model used in a probabilistic seismic hazard analysis calculation), evaluating

$\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)$, σ_{ij} and τ_j for every earthquake event, simulating one η_j for each event from a normal distribution, and simulating ε_{ij} values for each event and each site using the residual correlation model discussed above. A more detailed discussion of this simulation procedure is provided by, e.g., Crowley and Bommer (2006) and Jayaram and Baker (2010).

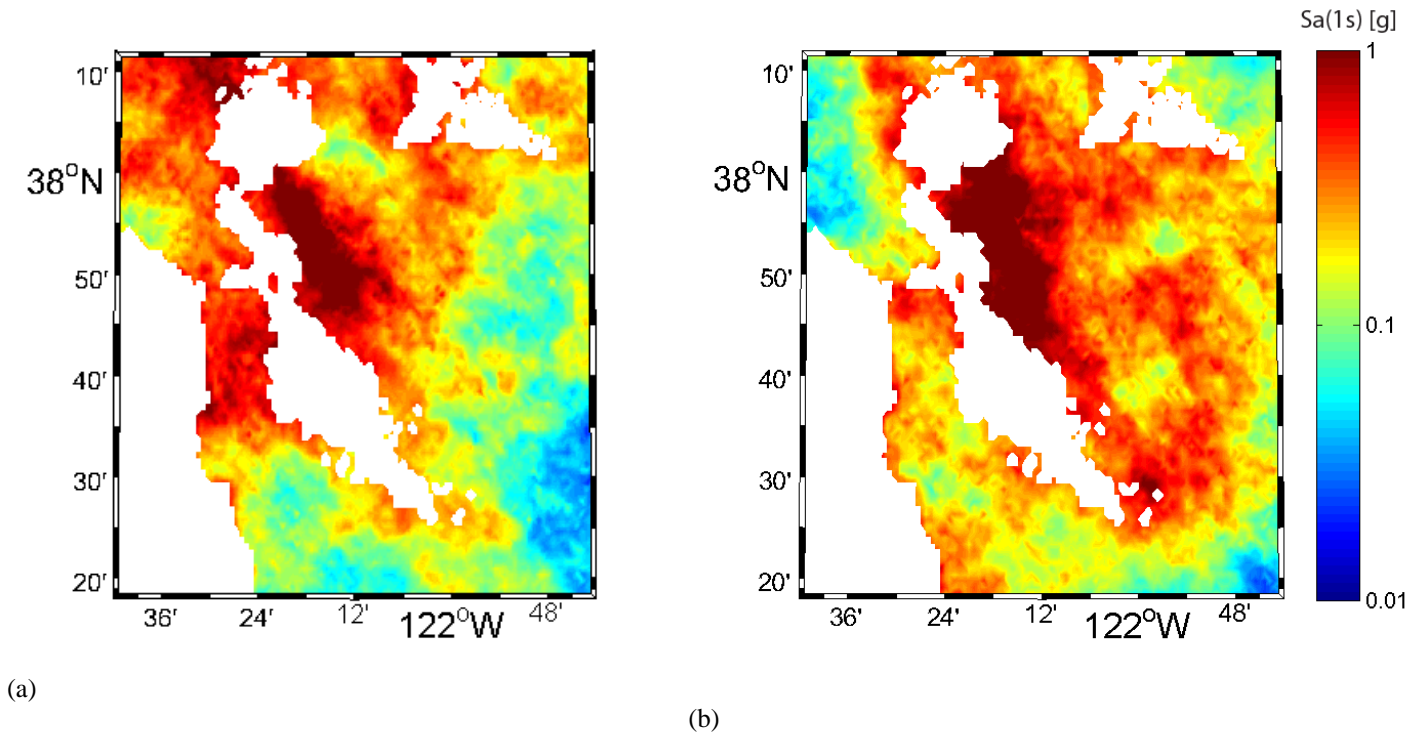


Figure 3: Two simulations spectral acceleration values associated with an $M = 7.05$ rupture of Northern Hayward fault. The mean predicted $\ln Sa$ values at each site are identical in (a) and (b), and only the ε_{ij} values vary between the two.

This calculation is facilitated by the Event Set Calculator in the OpenSHA software package (Field et al. 2005), as this tool automatically computes the required $\ln Sa(M_j, R_{ij}, V_{s30,i}, T, \dots)$, σ_{ij} and τ_j values for all specified sites in a region and for each earthquake scenario considered in a given source model¹. The user then only needs to provide simulated ε_{ij} and η_j values to complete the simulation procedure used here. For the results below, this procedure was used to produce spectral acceleration simulations for a set of 9195 sites in the San Francisco Bay Area, using the Boore and Atkinson (2008) ground motion prediction model and the UCERF2 seismic source model (Field et al. 2009). V_{s30} values for each of the 9195 sites were determined using OpenSHA's V_{s30} map server (Wald and Allen 2007). For each of these sites, 4039 $Sa(1s)$ simulations of mean values were produced (corresponding to 4039 earthquake scenarios), and three sets of (spatially correlated) residuals per event were simulated using the above model, for a total of 12,117 simulations. Examples of two resulting maps of regional ground motion are shown in Figure 3.

Using this catalog of simulations, Sa values at pairs of sites can be studied to evaluate their correlations. Because the Monte Carlo procedure produces a large suite of ground motions for each pair of sites, and because each input to the simulation procedure has been empirically validated, it is possible to study these results on a site-by-site basis to identify the degree of correlation implied by the input models to this procedure. The correlation in means is captured implicitly through the prediction of mean $\ln Sa$ values for each site and event, and the previously calibrated correlation in residuals has been incorporated through simulation of appropriately correlated ε_{ij} values.

Example data from this simulated catalog is shown in Figure 4. Observed spectral acceleration values at Berkeley and Stanford (points A and B in Figure 1) for an $M = 7.05$ rupture of the Northern Hayward fault are plotted in Figure 4a. All points in this plot thus

¹ In the OpenSHA terminology, the ground motion prediction model is termed an Intensity Measure Relationship, and the source model is termed an Earthquake Rupture Forecast.

correspond to fixed values of M_j , R_{ij} , etc. in equations (4) and (5), and only the ε_{ij} values vary². In this case, because the two sites are approximately 50 km apart, the ε_{ij} values are effectively uncorrelated at this large separation distance based on the empirical model for correlation in residuals shown in Figure 2b. Figure 4b shows spectral values for the same two sites, but now considering all 12,117 events in the synthetic catalog. Here we see that, even though the correlation in residuals is effectively zero for these two sites, the correlation in means contributes to produce a strong overall correlation in spectral values at these two sites. Qualitatively speaking, the data in the upper right corner come from very large earthquakes that will produce strong shaking throughout the region, and the data in the lower left corner come from small earthquakes that do not produce strong shaking. Note that the simulated events were produced using an importance sampling procedure that preferentially simulates the large earthquakes of interest for risk assessment, and then applies weights to each simulation to correct for this preferential sampling. The points plotted in Figure 4b are thus not all equally weighted and so visual inspection does not provide a perfect representation of the relative frequency of observing spectral values of given amplitudes, but qualitative visual inspections still provide reasonable interpretations.

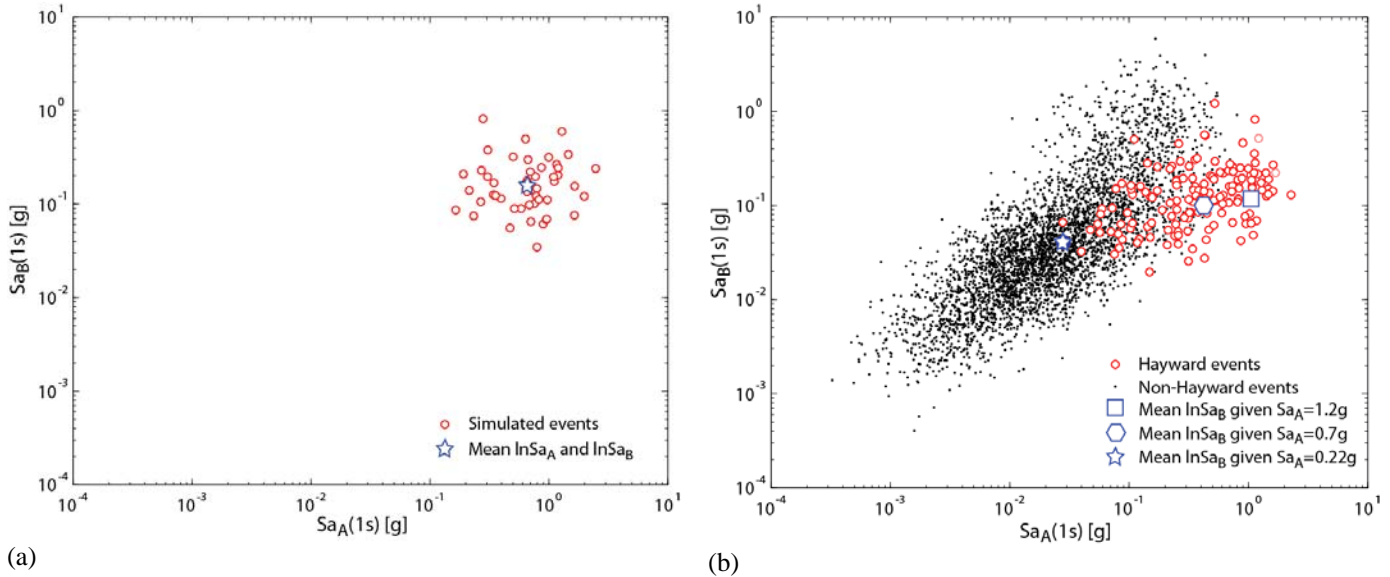


Figure 4: (a) Simulated one-second spectral acceleration values at Berkeley and Stanford, given an $M = 7.05$ rupture of the Northern Hayward segment. (b) Simulated one-second spectral acceleration values at Berkeley and Stanford, considering all events in the synthetic catalog.

For the case of logarithmic spectral accelerations conditional on an event, the joint distribution of $\ln Sa$'s are seen to be well-described by a joint Gaussian distribution (Jayaram and Baker 2009), so a correlation coefficient is sufficient to completely describe the joint distribution of intensities. Once random earthquake sources are considered, however, this model is no longer valid in general, as non-Gaussian distributions of earthquake magnitudes and distances, and nonlinear functional forms in the $\ln Sa(M_j, R_{ij}, T, \dots)$ will in general lead to non-Gaussianity of the resulting $\ln Sa$ values. This is apparent in Figure 4b, where the simulated $\ln Sa$ values do not fall in elliptical shapes that would suggest joint Gaussian distributions. In particular, the large Hayward events produce disproportionately large $\ln Sa$ values in Berkeley relative to Stanford, and large San Andreas events do the reverse (the Hayward events are plotted in a different style in Figure 4b to illustrate this effect). The non-Gaussianity of the joint distribution means that a linear correlation coefficient is an imperfect descriptor of the joint behavior of the $\ln Sa$ values at a given pair of sites.

The potential inadequacy of linear correlation coefficients can also be seen in Figure 5a, which shows the correlation coefficients of $\ln Sa$ values at Berkeley with all other sites in the region (for reference, the correlation coefficient of the data in Figure 4b is 0.65, and that value is plotted at the location of Stanford in Figure 5a). Also shown in Figure 5b-d are mean values of $\ln Sa$ throughout the region, conditional on $Sa(1s)$ values at Berkeley of 0.22g, 0.7g, and 1.2g, respectively. These are the amplitudes exceeded at Berkeley with 50%, 10% and 2% probabilities in 50 years, respectively. These maps are obtained by finding all simulations in the synthetic catalog with Sa values equal to the target amplitude at Berkeley (within some small tolerance), and taking the mean values of the simulated $\ln Sa$'s at each other site in the region. The three conditional mean values at the Stanford site are noted in Figure 4b above.

² Extra ε_{ij} values for this event and pair of sites have been simulated to illustrate this joint behavior in more detail, but the main synthetic catalog only has three such pairs of spectral values for this particular earthquake event

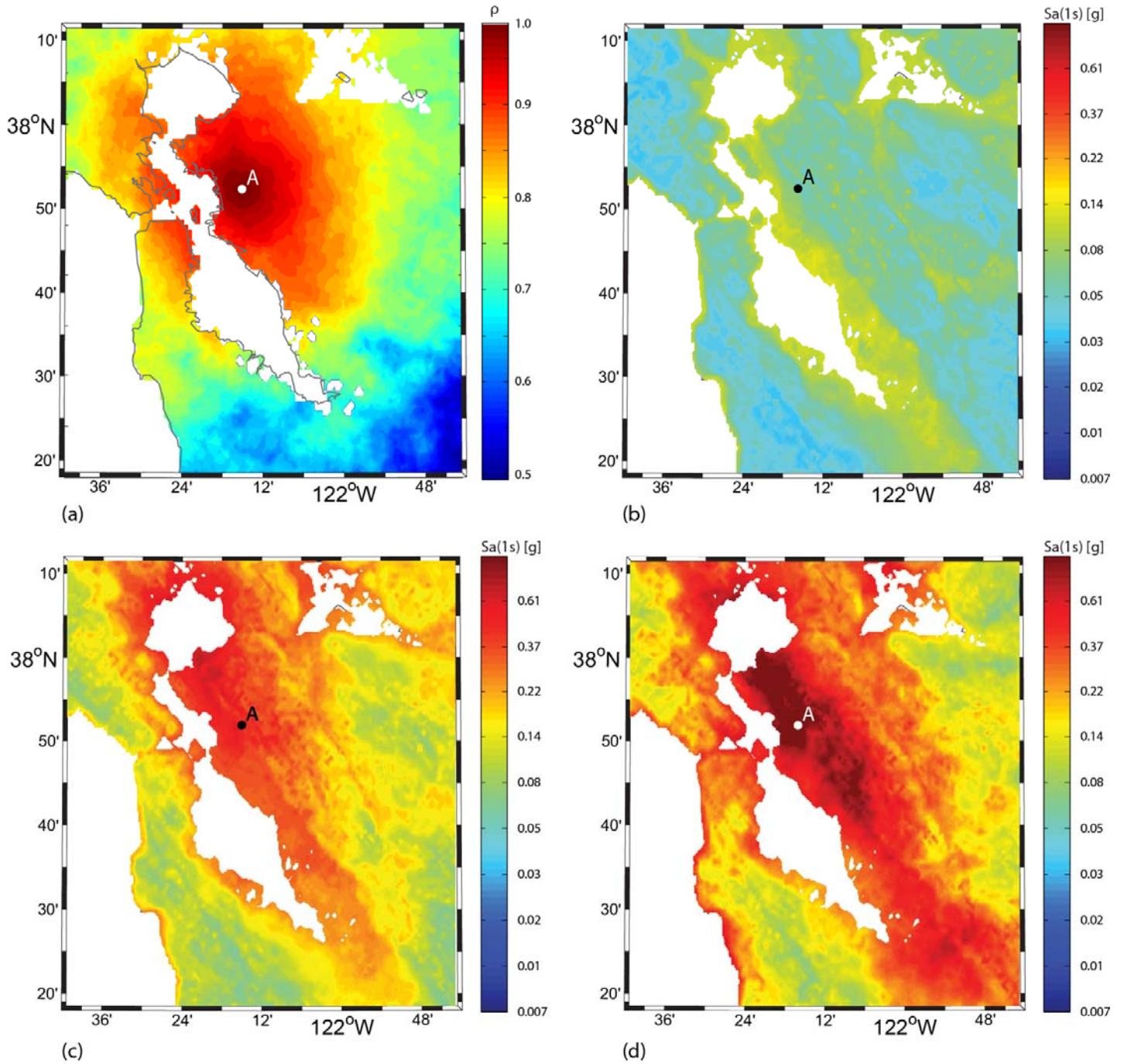


Figure 5: Measures of joint characteristics of ground motions at Berkeley (labeled as point “A” on the figures) and other locations throughout the San Francisco Bay Area (a) Correlation of $\ln Sa$ values between Berkeley and other locations, (b) Conditional mean $\ln Sa(1s)$ values given $Sa_A(1s)=0.22g$, (c) Conditional mean $\ln Sa(1s)$ values given $Sa_A(1s)=0.7g$, (d) Conditional mean $\ln Sa(1s)$ values given $Sa_A(1s)=1.2$.

Several observations can be made from these figures. It is notable that correlation coefficients decay as you move away from the conditioning site in a manner that is dependent almost exclusively on distance, as indicated by the approximately concentric circles of equal correlation. This is similar to the way in which correlation in residuals decay (although here the decay is slower due to the added effect of correlation in means). In contrast, the conditional mean values show additional structure, with elevated mean values tracing out regions along the Hayward fault, and also being present at sites with low V_{s30} values. The difference between the two cases suggests that correlation coefficients are insufficient to characterize the joint behavior of spectral values at pairs of sites, indicating non-Gaussianity of the data (if the $\ln Sa$ values were Gaussian, then plots of correlation and conditional means would be nearly

identical, with the only differences being due to variation in marginal mean values of $\ln Sa$ from site to site). This is not surprising, given Figure 4b, which shows non-Gaussian behavior in the joint distribution of $\ln Sa$ values at sites A and B. There are other cases, however, where perhaps joint Gaussian behavior is a reasonable model for $\ln Sa$ distributions. Figure 6 shows $\ln Sa$ values from the stochastic catalog for sites A and C (Berkeley and Milpitas), and the $\ln Sa$ values here lie in more approximately elliptical patterns, which could be represented by contours of a Gaussian distribution.

To quantify our comparisons of these two cases, we can note that the Stanford and Milpitas sites are both 50 km away from Berkeley. The total correlation coefficient between Berkeley and Stanford is 0.65, while the correlation between Berkeley and Milpitas is 0.71, corresponding to our observation in Figure 5a that correlation coefficients are approximately constant at a given separation distance from Berkeley. On the other hand, the exponential of the mean $\ln Sa$ values at the two sites conditional on Sa at Berkeley equal to 1.2g are more dramatically different: 0.12g at Stanford and 0.38g at Milpitas. The more approximate Gaussianity of the Berkeley and Milpitas pair, and the higher conditional mean $\ln Sa$ values at Milpitas, are due to Berkeley and Milpitas both having high- Sa ground motions coming from Hayward Fault earthquakes, while the Stanford site has high- Sa values that come from San Andreas Fault earthquakes (as expected from inspection of Figure 1, and confirmed from seismic hazard analysis deaggregation at each site). This suggests that when pairs of sites have hazard dominated by a common seismic source, that simple models for joint distributions of Sa values may be realistic and simple to calibrate, while if the pairs of sites have hazard driven by differing sources, the picture is more complicated. Ongoing work aims to quantify these results in a form that will facilitate decision making without requiring the extensive simulations performed here.

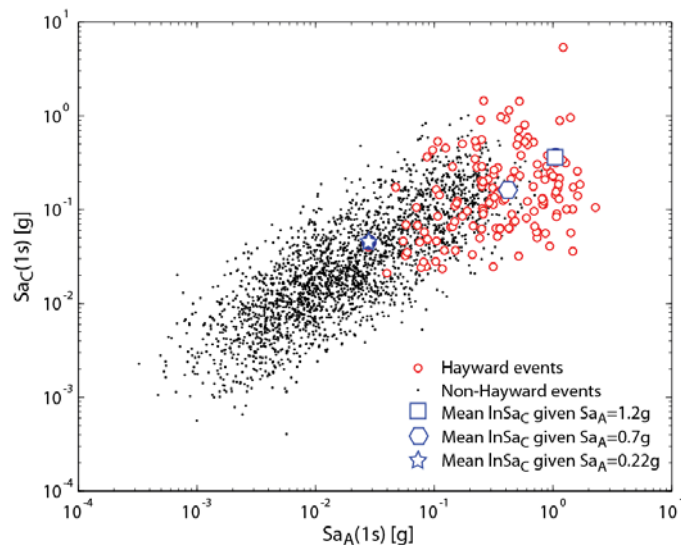


Figure 6: Simulated one-second spectral acceleration values for Berkeley and Milpitas, considering all events in the synthetic catalog.

REGIONAL VARIATION OF CORRELATIONS

We have seen from the above examples comparing two pairs of sites that the joint distribution of spectral values depends upon orientation of faults relative to the sites of interest, rather than simply separation distance between the pair of sites. This implies that trends in spatial correlation of $\ln Sa$ values will vary from region to region, since fault structures and geometries will vary from region to region.

To illustrate, Figure 7a shows correlation coefficients in $\ln Sa$ values from pairs of locations, plotted versus their separation distance. These are the same data shown in Figure 5a. There is variation in measured correlation at a given separation distance, due to variation in site conditions and proximities to faults among the various pairs of sites shown here. To discern a general trend in these data, the mean value of correlation at a given distance is estimated using a moving-average fit, and is also shown in Figure 7a. A comparable synthetic catalog of $\ln Sa(1s)$ values was then produced for the Los Angeles basin region, and correlation coefficients in those data were computed versus distance. The mean correlation versus distance in Los Angeles is compared to the comparable result from San Francisco in Figure 7b. It is seen that correlations in Los Angeles tend to be lower, presumably because the San Andreas and Hayward

faults are the dominant sources of seismic hazard in the San Francisco Area, and they extend long distances, meaning that high Sa values at distant sites may occur in the same earthquake. Los Angeles, in contrast, has major contributions to hazard from local thrust faults that may not extend over as great of lengths, thus reducing the correlation in means in that region. The differences in correlations may or may not be important, depending upon the application of interest; it is also worth recalling that linear correlation is an imperfect measure of joint distributions of $\ln Sa$ values in many cases, so more work is needed to fully understand the patterns we are interested in. Identification of improved metrics for capturing these joint distributions is a topic of ongoing work. Also shown in Figure 7b for reference is the model for correlation in residuals from Figure 2b, showing that the correlation in means greatly increases overall correlations relative the consideration of only correlation in residuals.

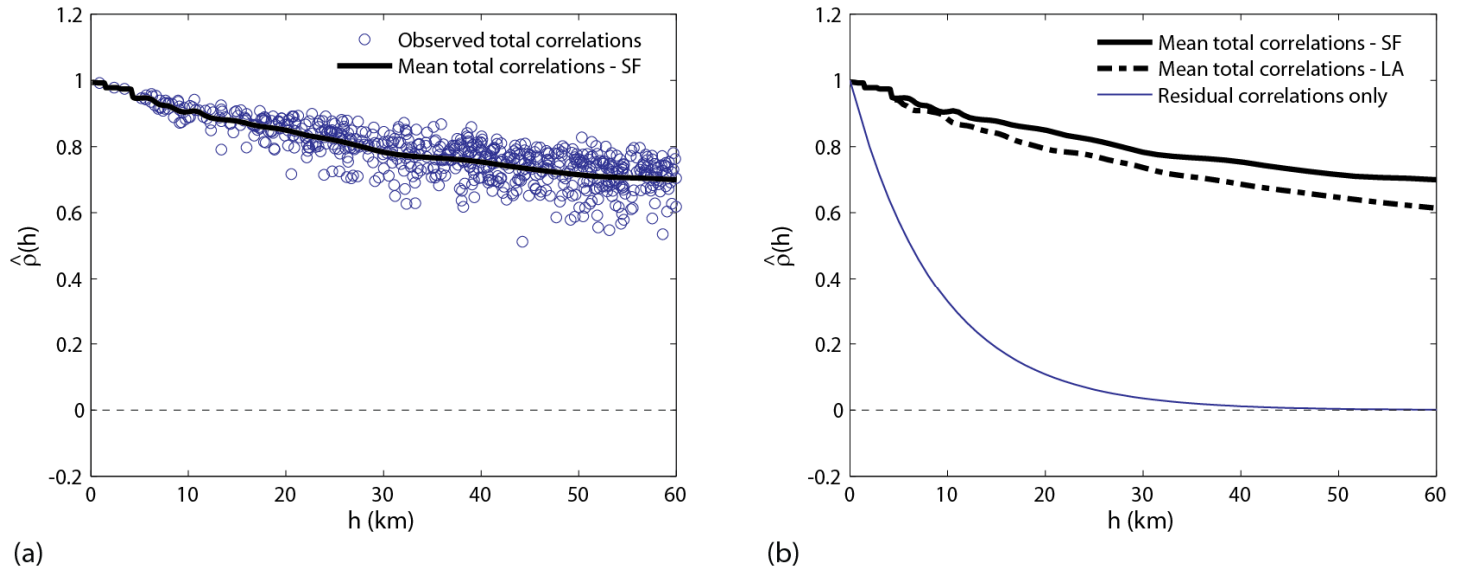


Figure 7: (a) Correlations in one-second log spectral accelerations at pairs of locations with a specified separation distance for the San Francisco Bay Area, and the mean trend in correlations obtained from a moving average fit. (b) Mean trend in correlations with distance from the San Francisco Bay Area (from a) compared with mean correlations from the Los Angeles area, and the residual correlations model of Figure 2b.

INFLUENCE OF SITE CONDITIONS

Surface geology plays an important role in strong ground motion, and it plays an important role in these spatial correlation results as well. This effect can be seen in several ways. While the correlation in residuals measured from past earthquakes does not appear to vary strongly across regions or with earthquake magnitude, Jayaram and Baker (2009) did note a potential dependence on V_{s30} heterogeneity in the region being studied, with strongly varying V_{s30} indicating lower correlation in residuals. This was speculated to be due to two effects. First, V_{s30} is understood to be an imperfect predictor of the effect of near-surface geology on ground motion, and when V_{s30} is strongly varying spatially, this is likely to be a proxy for strongly varying geology and thus variations in ground motion that are not well explained by the ground motion prediction model, leading to lower correlations in residuals. Second, when V_{s30} values are inferred, they vary less across a region than when they are directly measured. These inferred V_{s30} values are likely to be imperfect estimates, and likely introduce a (spatially correlated) error in ground motion predictions, leading to spatially correlated residuals. Thus inferred site conditions are likely to be associated with homogeneous V_{s30} values and greater correlation in residuals (Jayaram and Baker 2009).

In addition to correlation in residuals, varying site conditions produce a correlation in means, as can be understood from inspection of equations (4) and (5). When the two sites of interest have the same near surface geology, they will have the same V_{s30} value and thus the site conditions will have the same effect on predicted mean $\ln Sa$ values. When the sites have difference near surface geology, the effect of V_{s30} on predicted mean values will differ between the two sites, and to the extent that the site condition term in the predictive model is nonlinear as a function of ground motion intensity, the ground motions at the two sites will differ across the various events considered. The same effect, to a lesser extent, will occur from other site-related terms (such as depth to bedrock) that might be present in the predictive models. The role of this effect is seen in Figure 8, which shows conditional mean values of $\ln Sa$, given Sa at

Berkeley = 1.2g. In Figure 8a, best estimates of site V_{s30} are used for each site in the region (this figure is identical to Figure 5d). Figure 8b shows the same result (mean values of $\ln Sa$, given Sa at Berkeley = 1.2g), but using a constant value of $V_{s30}=698$ m/s for all sites (the value at the Berkeley Hills site used for conditioning, so that the hazard at that site is identical in both cases). The difference between these two cases indicates quantitatively the qualitatively-intuitive impact that varying site conditions have on this joint behavior.

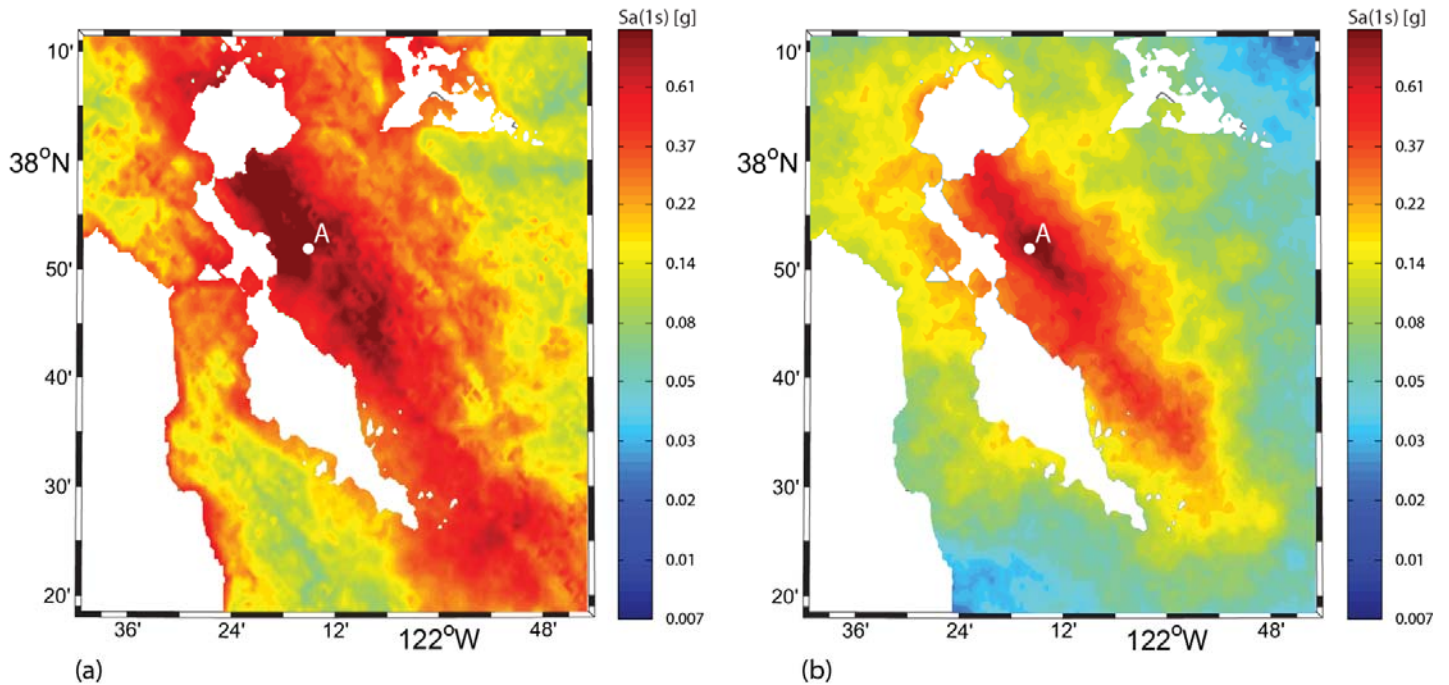


Figure 8: (a) Conditional mean $\ln Sa(1s)$ values given $Sa_A(1s)=1.2g$, including site-specific V_{s30} values. (b) Conditional mean $\ln Sa(1s)$ values given $Sa_A(1s)=1.2g$, using a constant $V_{s30}=698$ m/s value for all sites.

Note that these results only account for site conditions to the extent that the V_{s30} -related component of the ground motion prediction model accounts for the true effect of site conditions on ground motion. This is admittedly an imperfect treatment of site effects, but is likely the only practical approach available at the moment. Potentially superior alternatives include estimation of site-specific site response factors for use with the ground motion prediction model (Baturay and Stewart 2003; Bazzurro and Cornell 2004), or use of simulated ground motions that account for near-surface site-effects (Li et al. 2011). The former approach may be impractical when considering large numbers of sites, and the latter requires the modeling of nonlinear near-surface effects that are not yet commonly considered in many physics-based simulation approaches. These improved options will likely become more feasible, however, in the near future.

ROLE OF SIMULATED GROUND MOTIONS

The above analysis was all performed in the context of empirical ground motion prediction models. An active research area at the moment is in the use of numerically simulated ground motions to replace such empirical models in seismic hazard analysis (Graves et al. 2010). Those simulations could be used to perform similar calculations to these, by stochastically simulating earthquake ruptures (including detailed source models necessary for the simulations), and then computing resulting ground motions at each location of interest. By repeating this process over a large number of earthquake ruptures, a stochastic catalog like that used above could be produced. With this approach, the distinction between correlation in residuals and correlation in means would not exist, and while the simulations would require careful validation, the resulting catalog could be used to help evaluate the validity of the empirical-based formulation above. This simulation-based approach would potentially allow for the study of the effect of topography, V_{s30} , directivity, etc. on correlations, since these effects can in principle be accounted for in simulations but are not as well constrained by the empirical ground motion prediction models used above.

Despite the potential of this approach, producing the required stochastic catalog is a notably non-trivial task. These calculations require the simulations to be performed for many or all possible earthquake scenarios in a region, rather than just a single scenario rupture; this would necessitate not only having a source model for the region, but also being able to randomize, e.g., slip distributions for a given rupture extent. Such efforts are taking place (Graves et al. 2010), and despite the considerable effort required to produce such data, this type of modeling will eventually facilitate a greatly improved understanding of spatial variation of ground motions.

While ground motion simulations have the potential to improve understanding of this problem, conversely the empirical-based formulation above also offers some insights to ground motion simulators. First, the comparison of correlations from recordings and simulations does serve as an interesting validation of simulations (Jayaram et al. 2010), as within-event correlations appear to be relatively stable across earthquake magnitudes, at least within active tectonic regions. This suggests that evaluating within-event correlations is one possible way to validate ground motion simulations. Second, it will be harder to study regional variations in spatial variation using simulations, due to non-uniform quality of the required detailed source models and crustal velocity models across regions at the moment. However, as those needed inputs improve for a greater range of regions, comparisons with empirical-based results may provide a useful validation point.

CONCLUSIONS

Spatial variation in ground motion spectral accelerations is an important property for assessing seismic risk at a regional scale, but has received relatively less attention than the distribution of ground motion intensity at single sites. Several results have been presented to demonstrate how joint distributions in Sa values at pairs of sites can be quantified in the context of empirical ground motion prediction by studying correlations in spectral acceleration residuals (which can be estimated from strong motion data), and correlation in mean $\ln Sa$ predictions (which are implicitly defined by seismic source models and ground motion prediction models). Both of these terms are empirically calibrated, and most of this information is already used in standard single-site probabilistic seismic hazard analysis. Correlations in residuals as a function of distance appear to be relatively consistent across the set of well-recorded earthquakes for which such calculations can be performed (since most clear predictors of ground motion variability are already accounted for in the ground motion prediction model), while the correlations in means will vary by region, as they will depend upon the geometry of earthquake ruptures and the number of earthquake sources in the region. Variations in site conditions across a region also play an important role in these spatial correlations. It is observed that non-negligible correlations in residuals exist in sites that are separated by up to approximately 25 km, and that non-negligible correlations in total Sa 's exist at separation distances of 100 km or more, depending upon the seismic source model for the region. These variations have implications for prediction of infrastructure disruption and insured losses to portfolios of buildings. Analysis results of the type shown here will also help facility owners deciding on the placement of redundant facilities and infrastructure agencies planning to mitigate system-wide damage. Looking forward, advancements in modeling the effect of site conditions on ground motions, and in physics-based ground motion simulation and hazard analysis, will improve our ability to understand and predict joint dependence of spectral accelerations at multiple sites, leading to even greater benefits for these stakeholders.

ACKNOWLEDGEMENTS

Thanks to Nirmal Jayaram for insightful discussions and feedback that motivated some of this work. This work was supported by the National Science Foundation under NSF grant number CMMI 0952402, the NSF Graduate Research Fellowship and the Stanford Graduate Fellowship. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the National Science Foundation.

REFERENCES

- Abrahamson, N. A., and Youngs, R. R. (1992). "A Stable Algorithm for Regression Analysis Using the Random Effects Model." *Bulletin of the Seismological Society of America*, 82(1), 505-510.
- Adachi, T., and Ellingwood, B. R. (2007). "Impact of infrastructure interdependency and spatial correlation of seismic intensities on performance assessment of a water distribution system." *10th International Conference on Application of Statistics and Probability in Civil Engineering (ICASP10)*, Tokyo, Japan, 9.

- Al Atik, L., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N. (2010). "The Variability of Ground-Motion Prediction Models and Its Components." *Seismological Research Letters*, 81(5), 794-801.
- Baturay, M. B., and Stewart, J. P. (2003). "Uncertainty and Bias in Ground-Motion Estimates from Ground Response Analyses." *Bulletin of the Seismological Society of America*, 93(5), 2025-2042.
- Bazzurro, P., and Cornell, C. A. (2004). "Nonlinear Soil-Site Effects in Probabilistic Seismic-Hazard Analysis." *Bulletin of the Seismological Society of America*, 94(6), 2110-2123.
- Boore, D. M., and Atkinson, G. M. (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., Gibbs, J. F., Joyner, W. B., Tinsley, J. C., and Ponti, D. J. (2003). "Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California." *Bulletin of the Seismological Society of America*, 93(6), 2737-2751.
- Crowley, H., and Bommer, J. J. (2006). "Modelling Seismic Hazard in Earthquake Loss Models with Spatially Distributed Exposure." *Bulletin of Earthquake Engineering*, 4(3), 249-273.
- Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., Parsons, T., Petersen, M. D., Stein, R. S., Weldon, R. J., and Wills, C. J. (2009). "Uniform California Earthquake Rupture Forecast, Version 2 (UCERF 2)." *Bulletin of the Seismological Society of America*, 99(4), 2053-2107.
- Field, E. H., Gupta, N., Gupta, V., Blanpied, M., Maechling, P., and Jordan, T. H. (2005). "Hazard Calculations for the WGCEP-2002 Earthquake Forecast Using OpenSHA and Distributed Object Technologies." *Seismological Research Letters*, 76(2), 161-167.
- Foulser-Piggott, R., and Stafford, P. J. (2011). "A predictive model for Arias intensity at multiple sites and consideration of spatial correlations." *Earthquake Engineering & Structural Dynamics*, (in press).
- Goda, K., and Hong, H. P. (2008). "Spatial Correlation of Peak Ground Motions and Response Spectra." *Bulletin of the Seismological Society of America*, 98(1), 354-365.
- Graves, R., Jordan, T. H., Callaghan, S., Deelman, E., Field, E., Juve, G., Kesselman, C., Maechling, P., Mehta, G., Milner, K., and others. (2010). "CyberShake: A physics-based seismic hazard model for southern California." *Pure and Applied Geophysics*, 1-15.
- Jayaram, N. (2010). *Probabilistic seismic lifeline risk assessment using efficient sampling and data reduction techniques*. PhD Thesis, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA.
- Jayaram, N., and Baker, J. W. (2010). "Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment." *Earthquake Engineering & Structural Dynamics*, 39(10), 1109-1131.
- Jayaram, N., and Baker, J. W. (2009). "Correlation model for spatially distributed ground-motion intensities." *Earthquake Engineering & Structural Dynamics*, 38(15), 1687-1708.
- Jayaram, N., Park, J., Bazzurro, P., and Tothong, P. (2010). "Estimation of spatial correlation between spectral accelerations using simulated ground-motion time histories." *9th US National and 10th Canadian Conference on Earthquake Engineering*, Toronto, Canada, 10p.
- Lee, R., and Kiremidjian, A. S. (2007). "Uncertainty and Correlation for Loss Assessment of Spatially Distributed Systems." *Earthquake Spectra*, 23(4), 753-770.
- Li, W., Assimaki, D., and Fragiadakis, M. (2011). "Site response modeling uncertainty in 'rupture-to-rafter' broadband ground motion simulations." *Earthquake Spectra*, (in press).
- Park, J., Bazzurro, P., and Baker, J. W. (2007). "Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation." *10th International Conference on Application of Statistics and Probability in Civil Engineering (ICASP10)*, Tokyo, Japan, 8.
- Shiraki, N., Shinozuka, M., Moore, J. E., Ii, Chang, S. E., Kameda, H., and Tanaka, S. (2007). "System Risk Curves: Probabilistic Performance Scenarios for Highway Networks Subject to Earthquake Damage." *Journal of Infrastructure Systems*, 13(1), 43-54.
- Sokolov, V., and Wenzel, F. (2010). "Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation." *Earthquake Engineering & Structural Dynamics*, n/a-n/a.
- Thompson, E. M., Baise, L. G., and Kayen, R. E. (2007). "Spatial correlation of shear-wave velocity in the San Francisco Bay Area sediments." *Soil Dynamics and Earthquake Engineering*, 27, 144-152.
- Wald, D. J., and Allen, T. I. (2007). "Topographic Slope as a Proxy for Seismic Site Conditions and Amplification." *Bulletin of the Seismological Society of America*, 97(5), 1379-1395.
- Wang, M., and Takada, T. (2005). "Macrospectral Correlation Model of Seismic Ground Motions." *Earthquake Spectra*, 21(4), 1137-1156.