Seismic risk assessment of spatially-distributed systems using ground-motion models fitted considering spatial correlation

Nirmal Jayaram
Risk Management Solutions Inc., CA, USA

Jack W. Baker
Stanford University, CA, USA

ABSTRACT: Ground-motion models are commonly used in earthquake engineering to predict the probability distribution of the ground-motion intensity at a given site due to a particular earthquake event. These models are often built using regression on observed ground-motion intensities, and are fitted using either the one-stage mixed-effects regression algorithm proposed by Abrahamson and Youngs (1992) or the two-stage algorithm of Joyner and Boore (1993). In their current forms, these algorithms ignore the spatial correlation between intra-event residuals. Recently, Jayaram and Baker (2010a) and Hong et al. (2009) observed that considering spatial correlation while fitting the models does not impact the model coefficients that are used for predicting median ground-motion intensities, but significantly increases the variance of the intra-event residual and decreases the variance of the inter-event residual. These changes have implications for risk assessments of spatially-distributed systems, because a smaller inter-event residual variance implies lesser likelihood of observing large ground-motion intensities at all sites in a region.

This manuscript explores the impact of considering spatial correlation on the ground-motion model in situations where the models are fitted using only a few recordings or closely-spaced recordings, which is often the case in low to moderately seismic regions such as the eastern United States. This is done by quantifying the changes to the variances of the inter-event and the intra-event residual in a variety of situations where the models are fitted using earthquakes with a small to moderate number of recordings that are separated by short to medium distances. It is seen that the changes to the variances of the residuals are more significant as the number of recordings per earthquake reduces, though the trend with the average station separation distance is not as clear. Finally, sample risk assessments are carried out for a hypothetical portfolio of buildings in order to illustrate the potential impact on the seismic risk of spatially-distributed systems. Overall, this work serves to illustrate the need to consider spatial correlation in the regression for ground-motion models in upcoming projects such as the NGA east project.

1 INTRODUCTION

Ground-motion models are commonly used in earthquake engineering to predict the probability distribution of the ground-motion intensity at a given site due to a particular earthquake event. Typically, a ground-motion model takes the following form:

$$ln\left(Y_{ij}\right) = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij} + \eta_i \qquad (1)$$

where Y_{ij} denotes the ground-motion intensity parameter of interest (e.g., $S_a(T)$, the spectral acceleration at period T) at site j during earthquake i; $f(\boldsymbol{P}_{ij}, \boldsymbol{\theta})$ denotes the ground-motion prediction function with predictive parameters $\boldsymbol{P}_{ij}$ (e.g., magnitude, distance of source from site, site condition) and coefficient set $\boldsymbol{\theta}$; ε_{ij} denotes the intra-event residual, which is a zero mean random variable with standard deviation σ_{ij}; η_i denotes the inter-event residual, which is a random variable with zero mean and standard deviation τ_{ij}.

Ground-motion models are primarily fitted using either the two-stage regression algorithm of Joyner and Boore (1993) (e.g., Boore and Atkinson, 2008) or the one-stage mixed-effects model regression algorithm of Abrahamson and Youngs (1992) (e.g., Abrahamson and Silva, 2008; Campbell and Bozorgnia, 2008; Chiou and Youngs, 2008). Both these algorithms, in their current forms, assume that the intra-event residuals are independent of each other, although the intra-event residuals are known to be spatially correlated (e.g., Jayaram and Baker, 2009). Recently, Jayaram and Baker (2010a) and Hong et al. (2009) observed that considering spatial correlation while fitting the models does not impact the model coefficients that are used for predicting median ground-motion intensities, but significantly increases the variance of the intra-event residual and decreases the variance of the inter-event residual. These changes have implications for risk assessments of spatially-distributed systems, because a smaller inter-event residual variance implies lesser likelihood of observing large ground-motion intensities at all sites in a region.

This study explores the impact of considering spatial correlation on the variances in situations where the models are fitted using only a few closely-spaced recordings. These situations often arise in low to moderately seismic regions such as the eastern United States (Cramer et al., 2010). This manuscript first summarizes the procedure proposed by Jayaram and Baker (2010a) for fitting ground-motion models with consideration of spatial correlation. Subsequently, the ground-motion models are fitted for a variety of sample earthquake databases comprising of earthquakes with a small to moderate number of recordings that are separated by short to medium distances, and the impact of considering the spatial correlation on the residual variances is discussed. Finally, sample risk assessments are carried out for a hypothetical portfolio of buildings in order to illustrate the potential impact of the changes in the residual variances on the seismic risk of spatially-distributed systems.

2 REGRESSION ALGORITHM FOR FITTING GROUND-MOTION MODELS CONSIDERING SPATIAL CORRELATION

The regression algorithm proposed by Jayaram and Baker (2010a) for fitting ground-motion models considering spatial correlation is summarized below.

In the first step of the algorithm, it is assumed that the random-effects terms $\eta_1, \eta_2, \cdots, \eta_M$ equal zero, in which case Equation 1 simplifies to $ln\left(Y_{ij}\right) = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij}$. The coefficient set $\boldsymbol{\theta}$ is then estimated based on the observed Y_{ij}'s using a fixed-effects regression algorithm. In the next step, the standard deviations σ (for the intra-event residuals) and τ (for the inter-event residuals) are computed using the likelihood maximization approach described below.

The total residuals (i.e., the sum of the inter-event and the intra-event residuals), denoted $\varepsilon_{ij}^{(t)}$, can be computed using the $\boldsymbol{\theta}$ estimated in the previous step as follows:

$$\varepsilon_{ij}^{(t)} = \varepsilon_{ij} + \eta_i = ln(Y_{ij}) - f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) \tag{2}$$

The likelihood (L_1) of having observed the set of total residuals $\boldsymbol{\varepsilon}^{(t)} = \left(\varepsilon_{ij}^{(t)}\right)$ can be estimated as follows:

$$ln(L_1) = -\frac{N}{2} ln(2\pi) - \frac{1}{2} ln|\boldsymbol{C}| - \frac{1}{2}\left(\boldsymbol{\varepsilon}^{(t)}\right)' \boldsymbol{C}^{-1}\left(\boldsymbol{\varepsilon}^{(t)}\right) \tag{3}$$

where N is the total number of data points, $\boldsymbol{C}$ is the covariance matrix of the total residuals and $\left(\boldsymbol{\varepsilon}^{(t)}\right)'$ denotes the transpose of $\boldsymbol{\varepsilon}^{(t)}$.

The covariance matrix $\boldsymbol{C}$ can be estimated as

$$\begin{aligned} \boldsymbol{C}(\varepsilon_{ij}^{(t)}, \varepsilon_{ij'}^{(t)}) &= \boldsymbol{C}(\varepsilon_{ij} + \eta_i, \varepsilon_{ij'} + \eta_i) \\ &= \rho(d_{jj'})\sigma^2 + \tau^2 \quad \forall \quad i, j, j' \end{aligned} \tag{4a}$$

$$\boldsymbol{C}\left(\varepsilon_{ij}^{(t)}, \varepsilon_{i'j'}^{(t)}\right) = 0 \quad \forall \quad j, j', i \neq i' \tag{4b}$$

where $\rho(d_{jj'})$ denote the spatial correlation between intra-event residuals at two sites j and j' as a function of $d_{jj'}$, the separation distance between j and j'.

The maximum likelihood estimates of σ and τ are those that maximize the likelihood function L_1, and are obtained using numerical optimization.

Given $\boldsymbol{\theta}$ and the maximum likelihood estimates of σ and τ, η_i can be estimated as follows:

$$\hat{\eta}_i = \frac{\boldsymbol{1}'_{n_i,1} \boldsymbol{C}_{\boldsymbol{c}}^{-1} \boldsymbol{\varepsilon}_i^{(t)}}{\frac{1}{\tau^2} + \boldsymbol{1}'_{n_i,1} \boldsymbol{C}_{\boldsymbol{c}}^{-1} \boldsymbol{1}_{n_i,1}} \tag{5}$$

Finally, using the estimated value of η_i, a new set of coefficients $\boldsymbol{\theta}$ is obtained using a fixed-effects algorithm for $ln(Y_{ij}) - \eta_i$ (i.e., considering $ln\left(Y_{ij}\right) - \eta_i = f(\boldsymbol{P}_{ij}, \boldsymbol{\theta}) + \varepsilon_{ij}$). The new set $\boldsymbol{\theta}$ is then used to reestimate σ, τ and $\boldsymbol{\eta}$, and this iterative algorithm is continued until the coefficient estimates converge.

3 IMPACT OF CONSIDERING SPATIAL CORRELATION ON INTER-EVENT AND INTRA-EVENT RESIDUAL VARIANCES

In the current study, the algorithm described in the previous section is used to refit the Campbell and Bozorgnia (2008) ground-motion prediction model (henceforth referred to as the CB08 model) for illustration. Both Jayaram and Baker (2010a) and Hong et al. (2009) reported that the consideration of spatial correlation does not significantly change the coefficients of the ground-motion model that are used for predicting median intensities. This was also observed in this study, and hence this manuscript only reports the variances of the inter-event and the intra-event residuals estimated after fitting the ground-motion models with and without consideration of spatial correlation. First, in order to provide a baseline model for comparison, the coefficients of the CB08 model are reestimated while ignoring spatial correlation. For consistency, only records in the Pacific Earthquake Engineering Research (PEER) Next Generation Attenuation (NGA) database used by CB08 are used for estimating the coefficients. Table 1 shows the standard deviations estimated in this study for predicting spectral accelerations at 1 second (denoted $S_a(1s)$) in the uncorrelated case. Also shown in the table for comparison are the corresponding published CB08 model coefficients. The value of the published intra-event residual standard deviation reported here corresponds to that at large V_s30's (The V_s30 is set above a threshold value beyond which the ground-motion model no longer consider soil non-linearity effects, wherein the intra-event residuals have a constant variance at any given period.) The refitted variance estimates obtained in this work are similar, but not identical, to those reported by CB08. As reported by Jayaram and Baker (2010a), these small discrepancies are likely due to the manual coefficient smoothing carried out by the authors of the CB08 model (Campbell, 2009). For consistency, the refitted model variances are treated as the benchmark values, for comparison to variances obtained considering spatial correlation.

Table 1: Standard deviations of residuals corresponding to $S_a(1s)$

Case	σ	τ
1	0.568	0.255
2	0.578	0.223
3	0.654	0.157

Case 1: Published CB08 results (Campbell and Bozorgnia, 2008)
Case 2: Estimated in this study without considering spatial correlation
Case 3: Estimated in this study considering spatial correlation

The model coefficients and variances are reestimated considering spatial correlation. The spatial correlation model is obtained from Jayaram and Baker (2009), and is shown below.

$$\rho(h) = e^{-3h/b} \tag{6}$$

where h (km) denotes the separation distance between the sites of interest, and b denotes the 'range' parameter which determines the rate of decay of correlation. This range is a function of the spectral period, and equals 26km when $S_a(1s)$ is considered.

After the incorporation of spatial correlation, significant changes are seen in the estimates of the variance of the residuals (Table 1). In particular, the value of σ increases from 0.578 to 0.654 and the value of τ decreases from 0.223 to 0.157 after incorporating the spatial correlation.

In certain situations, it becomes necessary to fit ground-motion models using data from earthquakes instrumented with only a few potentially closely-spaced recording stations. This is, for example, the case while fitting ground-motion models in low to moderately seismically active regions such as the eastern United States. The impact of considering spatial correlation in such situations is discussed subsequently in this section.

First, the CB08 ground-motion model is fitted to predict spectral accelerations at six different periods (0, 1, 2, 4, 7.5, 10 seconds) with and without consideration of spatial correlation using subsets of earthquakes with specified maximum average spacings between recordings. The percent change in the residual standard deviations on incorporating spatial correlation in various cases is shown in Figure 1. In general, the consideration of spatial correlation results in an increase in the value of σ and a decrease in the value of τ. These changes are more significant when smaller average separation distance thresholds are used, with an exception being the change in the long-period σ's when the threshold is set at 50km. In fact, when the threshold is set to below 50km (in which case, the average separation distance falls to below 12km), the value of τ drops close to zero at periods longer than 1s, as indicated by a near 100% change in the value of τ in these cases.

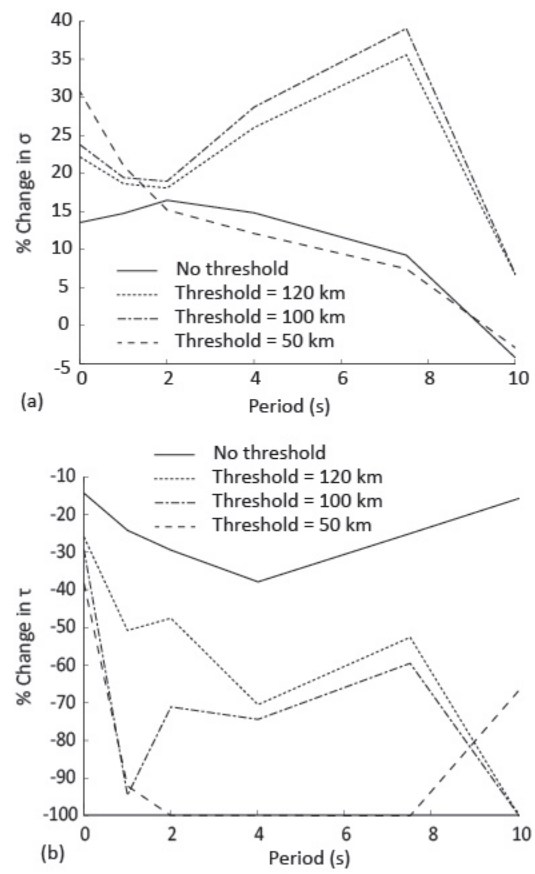

Figure 1: Impact of the average site-separation distance on (a) σ (b) τ

In order to study the impact of fitting ground-motion models using earthquakes that are not well instrumented, a subset of the NGA database is chosen that only comprises of earthquakes with a specified maximum number of recordings. The CB08 ground-motion model is then fitted with and without consideration of spatial correlation using the selected subset. The percent change in the residual standard deviations on incorporating spatial correlation in various cases is shown in Figure 2. In this case, the impact of the considering spatial correlation does not increase monotonically with a reduction in the threshold number of sites (significant changes are observed when the threshold is set at 200). In all the cases, however, significant impacts are seen on both σ and τ when spatial correlation is considered.

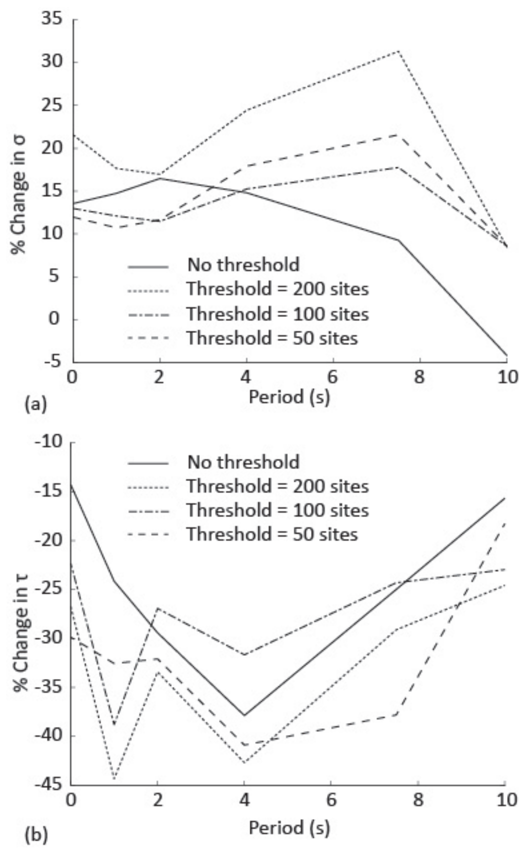

Figure 2: Impact of the number of recordings on (a) σ (b) τ

As discussed by Jayaram and Baker (2010a), ignoring spatial correlation while fitting the ground-motion model does not significantly affect the estimates of the ground-motion medians ($f(\boldsymbol{\theta})$) or the standard deviation of the total residuals, and therefore hazard and loss analyses for single structures will produce accurate results if the existing ground-motion models are used. Risk assessments for spatially-distributed systems, however, are influenced by the standard deviation of the inter-event and the intra-event residuals and not just by the medians and the standard deviation of the total residuals (this is discussed in more detail in the following section). Therefore, risk assessments of such systems carried out using ground-motion models fitted with and without consideration of spatial correlation could result in different loss estimates. Based on Figure 1, it can be concluded that the error in the loss estimates obtained using ground-motion models fitted without the consideration of spatial correlation will, in particular, be more severe when the models are fitted using closely-spaced recorded ground motions. In the next section, this is illustrated using risk assessments carried out on a hypothetical portfolio of buildings located in the San Francisco Bay Area.

Recently, many researchers have focused on estimating a 'site-specific variance' term, with the goal of reducing the intra-event residual variance by explicitly quantifying the site-effects which if not modeled can contribute to the variance (e.g., Al Atik et al., 2010). The increase in the variance of the intra-event residual while considering spatial correlation lends this issue additional importance and potential.

4 RISK ASSESSMENT FOR A HYPOTHETICAL PORTFOLIO OF BUILDINGS

Consider a hypothetical portfolio of 100 buildings in the San Francisco Bay Area located on a 10 by 10 grid with a grid spacing of 20km. Each building in the portfolio is assumed to have a replacement value of \$1,000,000. The seismic risk of this portfolio is estimated by modeling the seismic hazard due to 10 different faults and fault segments. (The source model is obtained from USGS (2003)). The risk assessment is carried out using a simulation-based procedure described in Crowley and Bommer (2006) and Jayaram and Baker (2010b). The steps involved in this procedure are summarized below.

Step 1: Simulate earthquakes of different magnitudes on the active faults in the region, using appropriate magnitude-recurrence relationships.

Step 2: Using the ground-motion model, compute the median ground-motion intensities ($f(\boldsymbol{\theta})$) and the standard deviations of the inter-event and the intra-event residuals (σ and τ respectively) at the sites of interest.

Step 3: Simulate the inter-event residual (i.e., η_j) by sampling from the univariate normal distribution with mean zero and standard deviation τ.

Step 4: Simulate the intra-event residuals (i.e., ε_{ij}'s) by sampling from a multivariate normal distribution with mean $\mathbf{0}_{p,1}$ (zero vector of size p) and covariance matrix given by Equation 4. Here, the spatial correlation ($\rho_{jj'}$) is defined by the exponential model in Equation 6 with a range of 26 km.

Step 5: Combine the medians, inter-event residuals and intra-event residuals using Equation 1 to obtain realizations of the ground-motion intensity at all sites of interest. In the rest of the paper, each set of ground-motion intensities is referred to as a ground-motion intensity map. The collection of all simulated ground-motion intensity maps quantifies the total ground-motion hazard in the region.

Step 6: Simulate the damage to the buildings due to each ground-motion intensity map. Here, this is done using damage functions which provide the probability of the building damage being in or exceeding various damage states (no damage, minor damage, moderate damage, extensive damage and collapse) as a function of the spectral acceleration at 1 second at the building location. The damage functions were assumed to be cumulative lognormal distribution functions with median values 0.4, 0.5, 0.7 and 0.9 for the minor, moderate, extensive and collapse damage states respectively. The lognormal standard deviation was assumed to be 0.6 in all these cases.

Step 7: Compute the total monetary loss associated with the damage to the portfolio due to each ground-motion intensity map. This is computed by assuming the damage ratio (ratio of repair cost to replacement cost) to be 0.03, 0.08, 0.25 and 1.00 for the minor, moderate, severe and collapse damage states respectively.

Step 8: Obtain the loss exceedance curve which provides the annual rate of exceedance of various monetary loss values. The loss exceedance curve is obtained as the product of the recurrence rates of all earthquakes in the region and the probability of exceedance of various monetary loss values. The exceedance probabilities are calculated as follows:

$$P(L \geq l) = \frac{1}{n} \sum_{i=1}^{n} I(L_i \geq l) \quad (7)$$

where $P(L \geq l)$ is the probability that the loss exceeds l, n denotes the number of simulated ground-motion intensity maps, L_i is the monetary loss associated with ground-motion intensity map i, and $I(L_i \geq l)$ is an indicator variable that equals one if L_i exceeds l and zero otherwise.

The above-mentioned risk assessment process is carried out using the values of σ and τ estimated in this work with and without consideration of spatial correlation for two cases: (a) case 1: ground-motion model fitted using all the NGA database records used by CB08 (b) case 2: ground-motion model fitted using earthquakes where the average station-station separation distance is less than 100km. In both cases, the CB08 median model coefficients are used for estimating median intensities, since the objective of this manuscript is to demonstrate the impact of the changes in the variances on the risk estimates. The resulting loss exceedance curves are shown in Figure 3.

It can be seen from Figure 3 that the recurrence rates of extreme losses are overestimated while using the variances obtained without consideration of spatial correlation. This is a result of the fact that the value of τ gets overestimated and σ gets underestimated on ignoring spatial correlation. A large value of τ increases the likelihood of observing large positive inter-event residuals, which will simultaneously increase the ground-motion intensity at all the sites in the region. If spatial correlations are large, a large value of σ will have a similar effect and can result in large ground-motion intensities at multiple sites. In such a case, the effect of underestimating σ is compensated by the effect of overestimating τ. If the spatial correlations are small, however, underestimating

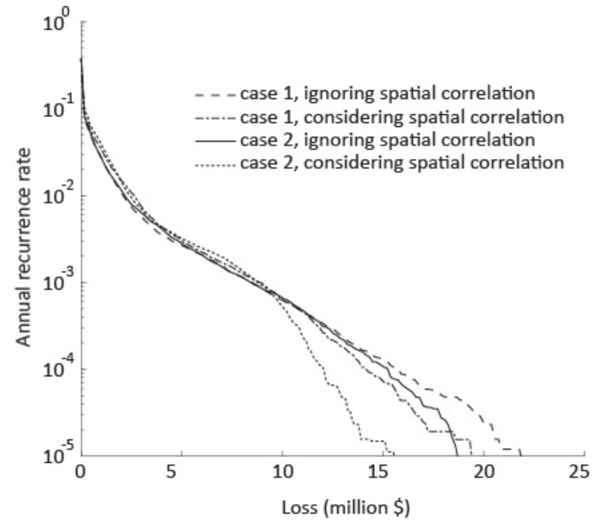

Figure 3: Risk assessment for a hypothetical portfolio of buildings performed using ground-motion models developed with and without the proposed refinement

σ and overestimating τ will have the net effect of jointly producing more extreme ground-motion intensities at multiple sites than is probable in reality. It can be inferred from Equation 6 that the spatial correlation will be small if h is large or if b is small. Therefore, when the components of a spatially-distributed system are well separated (large h) or if the correlation range is small, the ground-motion models fitted without considering spatial correlation will overestimate the likelihood of jointly observing extreme ground-motion intensities at multiple sites. It is to be noted that the separation between the buildings in the hypothetical portfolio considered in this work is substantial, which leads to differences between the loss curves obtained with and without consideration of spatial correlation.

Further, the difference between the risk estimates obtained with and without consideration of spatial correlation is higher in case 2 than in case 1 directly as a consequence of the more significant differences in the values of σ and τ in case 2.

It is difficult to make general conclusions about the size of this effect, but it is clear that seismic risk analysis calculations using existing ground-motion model estimates of σ and τ will bias the estimated risk.

5 CONCLUSIONS

This work illustrated the impact of considering spatial correlation between intra-event residuals while developing ground-motion models in situations where the models are fitted using only a few recordings or closely-spaced recordings, which is often the case in low to moderately seismic regions such as the eastern United States. Ignoring spatial correlation while fitting ground-motion models is seen to underestimate the intra-event residual variance while overestimating the inter-event residual variance. The changes to the variances are more severe in situations where the models are fitted using only a few recordings, though the trend with average station separation distance is not clear. The changes to the variances have implications for risk assessments of spatially-distributed systems, because a smaller inter-event residual variance implies lesser likelihood of observing large ground-motion intensities at all sites in a region. This is illustrated using sample risk assessments for a hypothetical portfolio of buildings, which show that the portfolio loss estimates are biased while fitting models obtained without consideration of spatial correlation.

This work serves to illustrate the need to consider spatial correlation in the regression for ground-motion models in upcoming projects such as the NGA east project. Further, many researchers have focused recently on estimating a 'site-specific variance' term, with the goal of reducing the intra-event residual variance by explicitly quantifying the site-effects which if not modeled can contribute to the variance. The increase in the variance of the intra-event residual while considering spatial correlation lends this issue additional importance and potential.

REFERENCES

Abrahamson, N. A. & W. J. Silva (2008). Summary of the Abrahamson & Silva NGA ground-motion relations. *Earthquake Spectra 24*(1), 99–138.

Abrahamson, N. A. & R. R. Youngs (1992). A stable algorithm for regression analyses using the random effects model. *Bulletin of the Seismological Society of America 82*(1), 505–510.

Al Atik, L., N. Abrahamson, J. Bommer, F. Scherbaum, F. Cotton, & N. Kuehn (2010). The variability of ground-motion prediction models and its components. *Seismological Research Letters 81*(5), 794–801.

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

Campbell, K. (2009). Personal Communication.

Campbell, K. W. & Y. Bozorgnia (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.01 to 10s. *Earthquake Spectra 24*(1), 139–171.

Chiou, B. S.-J. & 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.

Cramer, C., J. Kutliroff, & D. Dangkua (2010). A database of eastern north American ground motions for the next generation attenuation east project. In *9th U.S. National and 10th Canadian Conference on Earthquake Engineering, Toronto, Canada*.

Crowley, H. & J. Bommer (2006). Modelling seismic hazard in earthquake loss models with spatially distributed exposure. *Bulletin of Earthquake Engineering 4*(3), 249–273.

Hong, H. P., Y. Zhang, & K. Goda (2009). Effect of spatial correlation on estimated ground-motion prediction equations. *Bulletin of the Seismological Society of America 99*(2A), 928–934.

Jayaram, N. & J. Baker (2010a). Considering spatial correlation in mixed-effects regression, and impact on ground-motion models. *Bulletin of the Seismological Society of America, in press*.

Jayaram, N. & J. Baker (2010b). Efficient sampling and data reduction techniques for probabilistic seismic lifeline risk assessment. *Earthquake Engineering and Structural Dynamics 39*(10), 1109–1131.

Jayaram, N. & J. W. Baker (2009). Correlation model for spatially-distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics 38*(15), 1687–1708.

Joyner, W. & D. Boore (1993). Methods for regression analysis of strong-motion data. *Bulletin of the Seismological Society of America 83*(2), 469–487.

USGS (2003). Earthquake probabilities in the San Francisco bay region: 2002-2031. Technical report, Open file report 03-214, USGS.