Proceedings of the 9th U.S. National and 10th Canadian Conference on Earthquake Engineering
Compte Rendu de la 9ième Conférence Nationale Américaine et
10ième Conférence Canadienne de Génie Parasismique
July 25-29, 2010, Toronto, Ontario, Canada • Paper No 1226

SPATIAL CORRELATION BETWEEN SPECTRAL ACCELERATIONS USING SIMULATED GROUND-MOTION TIME HISTORIES

N. Jayaram[1], J. Park[2], P. Bazzurro[2], and P. Tothong[2]

ABSTRACT

Many private and public stakeholders are strongly affected by the impact of earthquakes on a regional basis rather than on just a single property at a specific site. The stakeholders include government and relief organizations that are in charge of preparing for future earthquakes and managing emergency response after an earthquake, and private organizations that have to manage spatially-distributed assets. The assessment of regional earthquake impact requires a probabilistic description of the ground-motion intensity field that a future event can generate or an event that has just occurred has generated. In order to quantify ground-motion intensities over a region, it is important to accurately assess and model the spatial correlation between the intensities. Statistical models that describe the spatial correlation between intensity measures are available in the literature, and the mathematics behind simple models that estimate the spatial correlation as a function of site separation distance has already been developed. In this study we intend to investigate whether a more sophisticated model of spatial variability that incorporates features such as non-stationarity (variation of correlation with spatial location), anisotropy (directional dependence) and directivity effects (different correlation models for pulse-like and non-pulse-like ground motions) is warranted. Testing the need for these additional features, however, requires a large number of ground-motion time histories. Since real data are sparse, the current study uses simulated ground-motion time histories instead. Overall, this study tests and provides a basis for some of the subtle assumptions commonly used in spatial correlation models.

Introduction

Many private and public stakeholders are strongly affected by the impact of earthquakes on a region rather than on just a single property at a specific site. In the aftermath of large events, public bodies, such as government agencies and relief organizations, and private entities, such as corporations and utility companies, need to assess the potential damage on a regional scale in order to plan their emergency response in a timely manner. These organizations also need to assess risks from future earthquakes in order to take mitigation actions such as retrofitting and acquiring insurance coverage. The impact of an event that just happened, or might happen in the future, can only be accurately evaluated by considering the distribution of ground-motion intensities at multiple sites throughout the region of interest.

Many predictive equations have been developed for estimating the distribution of the ground-motion intensity that an earthquake can cause at a single site. Much less attention has

[1]Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA – 94305
[2]AIR Worldwide, San Francisco, CA – 94111

been devoted, however, to estimating the statistical dependence between ground-motion intensities generated by an earthquake at multiple sites. In general, the values of a ground-motion intensity parameter at two sites are correlated due to the following reasons[3]: a) they have been generated by the same earthquake (e.g., a high stress-drop earthquake may generate ground-motion intensities that are, on average, higher than the median values from events of the same magnitude); and b) the seismic waves travel over a similar path from source to site. Modern ground-motion models implicitly account for the first cause of dependence via a specific inter-event error term, η_i, as follows:

$$\ln Y_{i,j} = \ln \overline{Y}_{i,j} + \sigma\varepsilon_{i,j} + \tau\eta_i \tag{1}$$

where $Y_{i,j}$ is the ground-motion intensity parameter of interest (e.g., $S_a(T_1)$, the spectral acceleration at period T_1), $\overline{Y}_{i,j}$ is the median value of Y predicted by the ground-motion model at site j for earthquake i (which depends on parameters such as the magnitude, distance of the site from the rupture and local site conditions), η_i is the normalized inter-event standard normal residual, $\varepsilon_{i,j}$ is the site-to-site normalized intra-event standard normal residual, and τ and σ are the corresponding standard deviations of the two residuals. While the ground-motion model in Eq. 1 partially accounts for the correlation of $Y_{i,j}$ at different sites via a common η_i, there is a significant amount of unaccounted correlation in the $\varepsilon_{i,j}$'s for different values of j. Of course, the ground-motion models do not provide an estimate of the correlation between $\varepsilon_{i,j}$'s at two different sites.

An alternative formulation for Eq. 1, which was common in older prediction equations, is given by

$$\ln Y_{i,j} = \ln \overline{Y}_{i,j} + \tilde{\sigma}\tilde{\varepsilon}_{i,j} \tag{2}$$

where $\tilde{\varepsilon}_{i,j}$ is a random variable called the normalized total residual representing both the inter-event and the intra-event variation at site j from earthquake i. By comparing Eqs. 1 and 2, it is can be seen (Park *et al*., 2007) that $\tilde{\sigma}$ must equal $\sqrt{\sigma^2 + \tau^2}$ and

$$\tilde{\varepsilon}_{i,j} = \frac{\tau\eta_i + \sigma\varepsilon_{i,j}}{\tilde{\sigma}} \tag{3}$$

In this study, we intend to empirically estimate the correlation between the intra-event residuals ($\varepsilon_{i,j}$) using ground-motion time histories. Since the inter-event residual is a constant across all sites during a given earthquake, however, it can be seen that the correlation between $\varepsilon_{i,j}$'s equals the correlation between $\tilde{\varepsilon}_{i,j}$'s.While estimating spatial correlations, it is convenient to directly work with total residuals (Eq. 3) since the values of $\tilde{\varepsilon}_{i,j}$ can be directly computed from the ground motion observations without the knowledge of η_i.

In the past, researchers have estimated the spatial correlations between the total residuals

[3] To avoid any possible misunderstanding we emphasize here that this study is not concerned with the similarity, or coherence, at a point in time of ground motion signals at closely spaced sites but rather investigates the correlation of peak values of oscillator response (or of the ground) at different sites observed over the entire duration of the shaking.

using recorded ground-motion data (e.g., Jayaram and Baker 2009, Wang and Takada 2005). Uisng geostatistical tools, Jayaram and Baker (2009) u identified various factors influencing the extent of the spatial correlation, and developed a predictive model that can be used to select appropriate correlation estimates. While recorded ground motions represent the natural source for estimating the extent of correlation between ground-motion intensities at two sites, they do not suffice for investigating the validity of assumptions such as second-order stationarity (i.e., dependence of correlation on just the separation between sites, and not on the actual location of the sites) and isotropy (i.e., invariance of correlation with the orientation of the sites) that are commonly used in the spatial correlation models developed so far. This is on account of the scarcity of ground-motion recordings for any particular earthquake. This limitation can be partially overcome by using simulated ground motions. Although the simulations may not be complete substitutes for recorded data, they are still extremely useful for testing and refining existing correlation models that requires a large amount of data. To verify the commonly-used assumptions of stationarity and isotropy in this study we utilized ground motions simulated by Dr. Brad Aagaard of the United States Geological Survey for the 1989 Loma Prieta earthquake source model (Aagaard et al. 2008). For information about other tests carried out using additional sets of simulated ground motions, the reader is requested to refer to Bazzurro et al. (2008).

Statistical estimation of spatial correlation

The current work uses geostatistical tools previously used by Jayaram and Baker (2009) to empirically estimate the spatial correlations of residuals from simulated ground-motion time histories. These tools are described briefly in this section; a detailed discussion can be found in, for example, Deutsch and Journel (1998) and Jayaram and Baker (2009).

Let $\tilde{\boldsymbol{\varepsilon}}$ denote the set of normalized total residuals distributed over space. The semivariogram of $\tilde{\boldsymbol{\varepsilon}}$ is a measure of the dissimilarity between the residuals and is useful in computing the spatial correlation between the residuals. Let u and u' denote two sites separated by $\mathbf{h}$. The semivariogram ($\gamma(u,u')$) is defined as follows:

$$\gamma(u,u') = \frac{1}{2} E\left[\left\{ \tilde{\varepsilon}_u - \tilde{\varepsilon}_{u'} \right\}^2 \right] \tag{4}$$

The semivariogram defined in Eq. 2 is location-dependent and its inference requires repetitive realizations of $\tilde{\varepsilon}$ at locations u and u'. Such repetitive measurements are, however, never available in practice. Hence, it is typically assumed that the semivariogram does not depend on site locations u and u', but only on their separation $\mathbf{h}$ to obtain a stationary semivariogram. The stationary semivariogram ($\gamma(\mathbf{h})$) can then be estimated as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[\left\{ \tilde{\varepsilon}_u - \tilde{\varepsilon}_{u+\mathbf{h}} \right\}^2 \right] \tag{5}$$

A stationary semivariogram is said to be isotropic if it is a function of the separation distance ($h = |\mathbf{h}|$) rather than the separation vector $\mathbf{h}$. An isotropic, stationary semivariogram can be empirically estimated from a data set as follows:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\tilde{\varepsilon}_{u_\alpha} - \tilde{\varepsilon}_{u_\alpha + h} \right]^2 \quad (6)$$

where $\hat{\gamma}$(h) is the experimental stationary isotropic semivariogram (estimated from a data set); *N(h)* denotes the number of pairs of sites separated by *h*; and $\left\{ \varepsilon_{u_\alpha}, \varepsilon_{u_\alpha + h} \right\}$ denotes the α^{th} such pair. Since $\hat{\gamma}(h)$ only provides semivariogram values at discrete values of *h*, a continuous function is usually fitted to the discrete values to obtain the semivariogram for continuous values of *h*. The exponential function shown below is commonly used for this purpose.

$$\hat{\gamma}(h) = a\left[1 - \exp\left(-3h / b\right)\right] \quad (7)$$

where *a* denotes the "*sill*" of the semivariogram (which equals the variance of the normalized total residuals (=1)) and *b* denotes the "*range*" of the semivariogram (which equals the separation distance *h* at which $\hat{\gamma}(h)$ equals 0.95a).

It can be theoretically shown that the spatial correlation function ($\hat{\rho}(h)$) for normalized total residuals (and therefore, for normalized intra-event residuals) can be computed from the semivariogram function as follows:

$$\hat{\rho}(h) = 1 - \hat{\gamma}(h) \quad (8)$$

Therefore, the correlations are completely defined by the semivariogram, which in turn, is a function only of the range. (The sill is known to equal one, which is the variance of the normalized total residuals for which the semivariogram is constructed.)

Results and discussion

For the 1989 Loma Prieta data set, ground-motion time histories are available at 35,547 sites. Soft soil sites with $V_{s30} \leq 500$m/s, however, are excluded from the computations, due to limitations of the simulation methodology in capturing nonlinear soil behavior. Also, current limitations in the simulation procedure only allow us to investigate the spatial correlation of spectral accelerations, S_a, at periods, *T*, longer than 2s. Simulated $S_a(T)$'s values for $T<2$s are considered unreliable.

The total residuals, $\tilde{\varepsilon}$'s, are computed from the fault normal $S_a(T)$ values with T=2s, 5s, 7.5s, and 10s using the Boore and Atkinson (2008) ground-motion model. Using the geostatistical procedure described in the previous section, discrete semivariogram values are estimated for these residuals, and an exponential function is subsequently fitted to the discrete values.

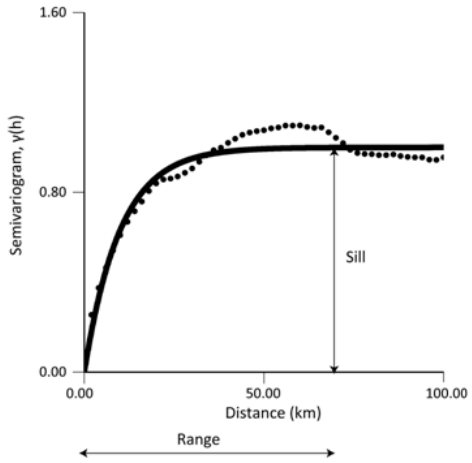

Figure 1. Semivariogram computed using the $S_a(T=2s)$ residuals.

Fig. 1 shows a sample semivariogram obtained using the residuals computed for $S_a(T=2s)$. The ranges of the semivariograms obtained using the fault normal residuals at all the periods are plotted in Fig. 2a. As mentioned earlier, the range is an indicator of the extent of spatial correlation, and a larger range implies a larger amount of spatial correlation. Fig. 2a shows that the range and, therefore, the amount of spatial correlation increase with oscillator period. This trend is to be expected given that the coherency between the period components of the ground motion increases with period (Der Kiureghian 1996). Note that the ranges obtained from this simulated 1989 Loma Prieta data set are slightly larger than those from recorded ground motions computed by Jayaram and Baker (2009) shown in Fig. 2b. This means that this simulated ground motion data set is more spatially correlated than real, recorded data sets analyzed so far. While uncovering the reasons of this apparent discrepancy is beyond the scope of this study, this finding can perhaps be used to enhance the simulation technique. Again, despite this limitation we assume that the large number of simulated ground-motions contains useful information to study the isotropy and second-order stationarity of ground-motion intensities. This exercise can be carried out irrespective of the extent of correlations observed.

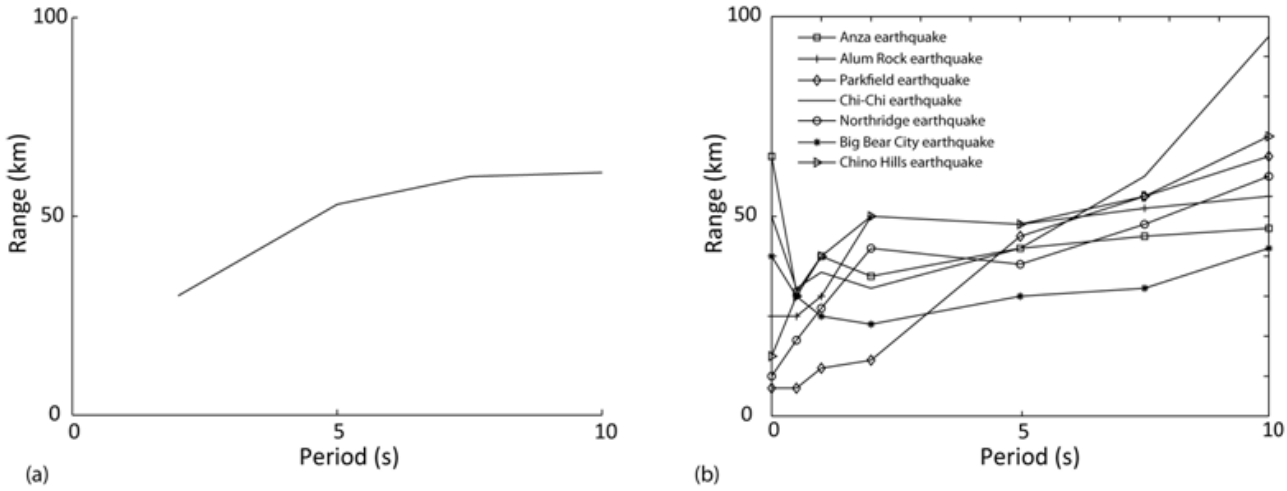

Figure 2. Ranges of semivariograms obtained using residuals computed from the (a) 1989 Loma Prieta simulations and (b) recorded ground motions (Jayaram and Baker 2009).

Effect of ground-motion component orientation on semivariogram's range

In order to test if the orientation of the ground-motion component used has an influence on the estimates of spatial correlation, semivariograms of residuals are estimated using the fault normal, fault parallel, north-south and east-west components of the simulated ground motions. The ranges of these semivariograms are shown in Fig. 3a. The range estimates are essentially identical for S_a at T=2s and do not show a significant variation with the component used at longer periods. Hence, most of the following analyses in this section are based on the fault normal components of the simulated ground motions.

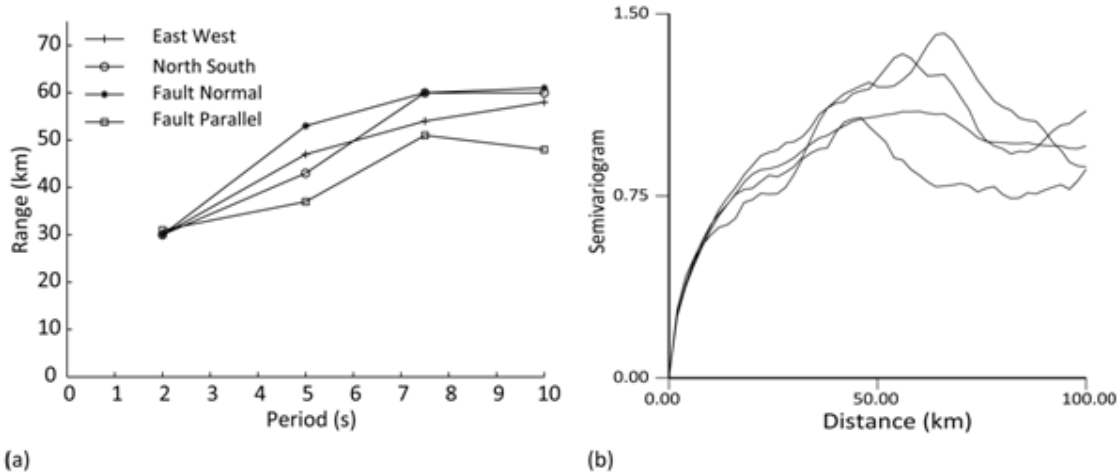

Figure 3. (a) Ranges are computed using residuals at different orientations (b) Omni-directional (i.e., obtained using all pairs of points irrespective of the azimuth) and directional semivariograms computed using residuals for S_a(T=2s).

Testing the assumption of isotropy using directional semivariograms

Directional semivariograms of residuals (Deutsch and Journel, 1998 and Jayaram and Baker, 2009) are obtained as shown in Eq. 5 except that the estimates are obtained using only pairs of $(\tilde{\varepsilon}_{u_\alpha}, \tilde{\varepsilon}_{u_\alpha+\boldsymbol{h}})$ such that the azimuth of the vector $\boldsymbol{h}$ is identical (or, strictly speaking, within a narrow band of azimuths) for all the pairs utilized. In this study we consider azimuth angles of 0°, 45° and 90°. If anisotropy is present in the data, the semivariograms along the pre-specified azimuths will differ from each other and from the omni-directional semivariogram (i.e., the semivariogram obtained using all pairs of points irrespective of the azimuth). Fig. 3b compares the omni-directional semivariogram with the semivariograms obtained by considering azimuths of 0°, 45° and 90° for residuals for S_a(T=2s). All the semivariograms are almost identical for separation distances below 10 km and are reasonably close for separation distances between 10 km and 20km. Recall that during the characterization of the distribution of ground-motion intensities over a region, it is more important to capture the effects of the spatial correlation at short separation distances since the extent of spatial correlation decreases rapidly with separation distance. Also, in addition to having low correlation, widely separated sites also have little impact on each other due to an effective 'shielding' of their influence by more closely-located sites (Deutsch and Journel 1998).

As a result, since the semivariograms in Fig. 3b are nearly identical at short separation

distances, it can be reasonably concluded that, at least for this data set, the spatial correlations can be adequately represented using an isotropic model. Tests carried out using this Loma Prieta simulated data set for residuals computed for S_a at longer periods showed similar results as well (Bazzurro et al. 2008).

Testing the assumption of second-order stationarity

A spatial random function Z is said to be second-order stationary if the random variable Z_u and Z_v (i.e., the random variables that represent the values of Z at locations u and v, respectively) have constant means and second-order statistics (i.e., the covariance) that depend only on the distance vector between u and v and not on the actual locations. In other words, the covariance is the same between any two sites that are separated by the same distance and direction (direction is not a concern for isotropic semivariograms), no matter where the sites are located with respect to the causative fault. The assumption of second-order stationarity is convenient while developing correlation models since it allows the data available over the entire region of interest to be pooled together and because it considerably simplifies the application of the spatial correlation models.

We know that the means of the residuals equal zero irrespective of the location of the residuals. Therefore, second-order stationarity can be tested by comparing the spatial correlation estimates obtained using residuals located in different spatial domains (i.e., using data from two groups of sites, one close to the fault and one far from it) Similar semivariograms imply that the actual spatial location of the sites where the ground-motion intensities are measured does not matter. In the current work, seven spatial domains are defined based on the distance of the sites from the rupture: Domain 1 includes sites between 0-20km while Domains 2-7 consist of sites between 20-40km, 40-60km, 60-80km, 80-120km, 120-160km and 160-200km of the rupture, respectively. Note that, as with histograms, the selection of the distance bins is somewhat arbitrary. Very narrow bins may provide results that are both unstable because of scarcity of data and potentially influenced by local effects (e.g., a cluster of sites with large residuals). Conversely, very broad bins may not detect any trend in the data, even if there is one. Here, the width of the domains is selected judiciously to avoid both the above pitfalls.

The 1989 Loma Prieta fault normal ground motions are used to compute $\tilde{\varepsilon}$ values at four different periods, namely, 2s, 5s, 7.5s and 10s. Semivariograms are constructed for each spatial domain using only the residuals at sites that belong to that domain, and the estimated ranges are reported in Fig. 4a. It can be seen that the ranges estimated using residuals at sites within 20-160km of the rupture are reasonably close to the range estimated using all fault normal residuals ('all-site ranges'). There are more significant differences, however, between the ranges computed using residuals at sites that are very close to or very far away from the rupture from the all-site ranges. Semivariograms computed using sites that are farther than 160 km from the rupture show significantly smaller ranges, as do the semivariograms computed using sites that are within 20 km of the rupture. The ground-motion intensities at sites farther than 160 km from the rupture are generally very small and, therefore, accounting for the reduced correlations at these extremely far-off sites is certainly not critical. It is, however, important to study the smaller correlations observed at near-fault locations. Intuitively, it is reasonable to expect small-scale variations to reduce spatial correlation between ground motions at near-fault sites. At sites farther than 20km, the small-scale variations have less influence, thereby resulting in larger ranges and, therefore, larger correlations.

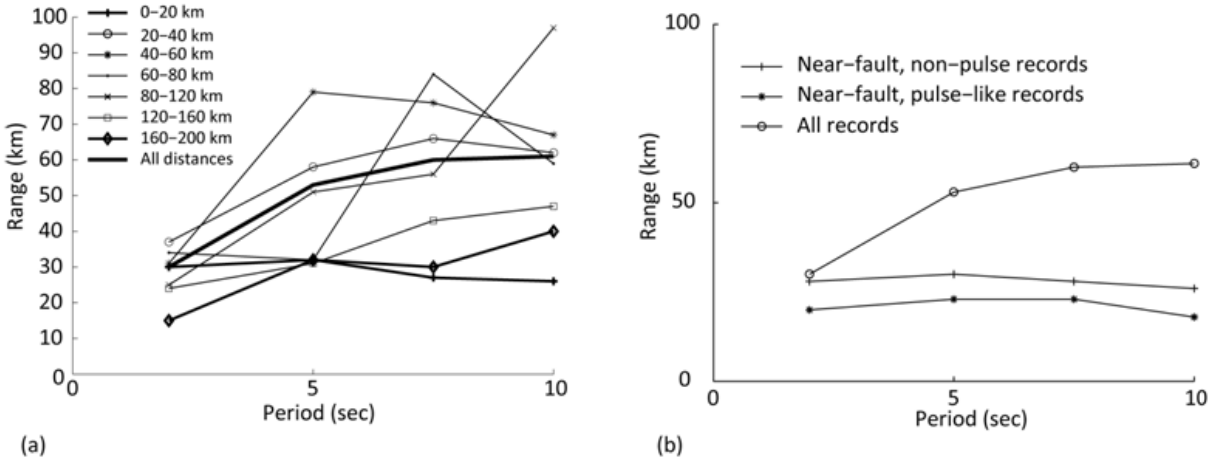

Figure 4: (a) Ranges are computed using residuals from different spatial domains (b) Ranges are computed using pulse-like and non-pulse-like near fault ground motions.

Effect of directivity on spatial correlation

Ground motions at near-fault sites are typically influenced by directivity effects, resulting often times in large amplitude pulse-like ground motions in the forward-directivity region. Most ground-motion models, however, do not explicitly capture this effect. Therefore, the residuals in such cases may be more correlated because of the additional prediction errors at sites influenced by directivity that are not captured by the ground-motion model. In this study, we intend to verify whether the spatial correlation between pulse-like ground motions is different from that between non-pulse-like ground motions.

Baker (2007) developed a technique that uses wavelet analysis to identify ground motions with pulses. Although not all the pulses identified by this technique are due to directivity effects, this approach provides a reasonable data set for studying the potential impact of directivity. The wavelet analysis procedure of Baker (2007) is used to identify 434 pulses in the fault normal components of 1989 Loma Prieta simulations (incidentally, the wavelet analysis procedure also identified 121 pulses in the fault parallel direction, which are not utilized here). Residuals at four different periods are computed based on these ground motions and semivariograms of the residuals are developed. The estimated ranges (shown in Fig. 4b) of these semivariograms are smaller than those estimated based on all the fault normal residuals, but similar to those estimated based on ground motions at all the sites that are within 20 km from the rupture (Fig. 4a). For a comparison, Fig. 4b also shows the ranges obtained using ground motions at all the sites that do not have pulse-like ground motions, but are within 20 km from the rupture (called near-fault non-pulse records in the legend). It is seen that the ranges obtained in this case are similar to the ranges obtained using pulses. This indicates that the effect of directivity does not substantially alter the ranges of the semivariograms. It is to be noted that the ranges based on near fault pulse-like and non-pulse-like ground motions have been computed separately only for pedagogical purposes. For most practical applications (e.g., risk assessment of portfolios of buildings), the only information required are the ranges computed based on all near-fault ground motions, unless sites where pulse-like ground motions will be experienced can be accurately predicted.

Conclusions

This study investigated the validity of commonly-used assumptions such as non-stationarity (variation of correlation with spatial location) and anisotropy (directional dependence) in spatial correlation models. Testing the need for these additional features, however, requires a large number of ground-motion time histories which are unavailable from historical earthquakes. Hence, as a proxy the current study used ground-motion time histories simulated by Dr. Brad Aagaard for the 1989 Loma Prieta earthquake source model instead. Other data sets were considered in Bazzurro et al.(2008).

Geostatistical tools were used to measure the extent of spatial correlation between spectral accelerations using the simulated ground motion data set. The correlations were estimated using various orientations of the time histories, namely, fault normal, fault parallel, north-south and east-west, and were found to be similar to one another. The spatial correlations were also seen to be essentially isotropic (unidirectional). In other words, the spatial correlation between two sites was seen to be independent of the orientation of the sites. The correlations were seen to be smaller than average between sites located extremely close to the fault rupture. . Intuitively, it is reasonable to expect small-scale variations to reduce spatial correlation between ground motions at near-fault sites. It is, however, important to further study the smaller correlations observed at near-fault locations. Incidentally, the ground-motion intensities at sites very far away from the rupture was also found to be less spatially correlated than average, but this finding is of no practical importance. We also used the pulse-identification algorithm of Baker (2007) for identifying pulse-like ground motions, and to compare the correlations between pulse-like and non-pulse-like ground motions. The study, however, did not find significant differences between the correlations in these two cases. Although additional investigation is needed, this study tests and provides a preliminary basis for some of the subtle assumptions commonly used in spatial correlation models.

Acknowledgments

We are very thankful to Dr. Brad Aagaard for providing the simulated ground-motion data set that forms the basis of this study. We are also thankful to Prof. Jack Backer for the fruitful discussions on the subject of spatial correlation that occurred over time and for providing his software for identifying pulse-like ground motions. The funding for this study was provided by USGS under the 07HQGR0032 Research Program Award.

References

B. T. Aagaard, T. M. Brocher, D. Dolenc, D. Dreger, R. W. Graves, S. Harmsen, S. Hartzell, S. Larsen, and M. L. Zoback (2008). Ground-motion modeling of the 1906 San Francisco earthquake, part I: Validation using the 1989 Loma Prieta earthquake, *Bulletin of the Seismological Society of America* 98(2), 989–1011.

J. W. Baker (2007). Quantitative classification of near-fault ground motion using wavelet analysis, *Bulletin of the Seismological Society of America* 97(5), 1486–1501.

P. Bazzurro, Park, J., Tothong, P., and Jayaram, N. (2008). Effects of spatial correlation of ground-motion parameters for multi-site risk assessment: Collaborative research with Stanford University and AIR.

Technical report, *Report for U.S. Geological Survey National Earthquake Hazards Reduction Program (NEHRP)* External Research Program Awards 07HQGR0032.

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 and 10.0 s, *Earthquake Spectra* 24, 99–138.

A. Der Kiureghian (1996). A coherency model for spatially varying ground motions, *Earthquake Engineering and Structural Dynamics*, 25, 99–111.

C. V. Deutsch and A. G. Journel (1998). *Geostatistical Software Library and User's Guide,* Oxford University Press, Oxford, New York.

N. Jayaram and J. W. Baker (2009). Correlation model for spatially-distributed ground-motion intensities, *Earthquake Engineering and Structural Dynamics* (in press).

Park, J., Bazzurro, P., and J.W. Baker (2007). "Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation", *Proceedings of ICASP10*, Tokyo, Japan, July 31-August 4.

Wang M. and Takada T., (2005). Macrospatial correlation model of seismic ground motions, *Earthquake Spectra* 21 (4), 1137-1156.