Effects of spatial correlation of ground motion parameters for multi-site seismic risk assessment: Collaborative Research with Stanford University (Prof. Jack Baker) and AIR

AWARD No. **07HQGR0032/07HQGR0031**

Dr. Paolo Bazzurro, Dr. Jaesung Park, Dr. Polsak Tothong, and Nirmal Jayaram
AIR Worldwide Corporation
388 Market Street, Suite 750
San Francisco, CA 94111
415/912-3111 (phone)
415/912-3112 (fax)
pbazzurro@air-worldwide.com
jpark@air-worldwide.com
ptothong@air-worldwide.com

in collaboration with
Prof. Jack Baker(Stanford University)

December 2008

Keywords: Spatial correlation of ground motion, vector probabilistic seismic hazard analysis, portfolio loss estimation.

TABLE OF CONTENTS

ABSTRACT

Many private and public stakeholders are strongly affected by the impact of earthquakes on a regional basis rather than on a single property at a specific site. The stakeholders could be government and relief organizations that need to prepare for future events and manage emergency response, or private organizations that have spatially distributed assets. Whether for mitigating future seismic risk or managing response after an earthquake, regional assessment of the earthquake impact requires a probabilistic description of the ground motion field that an event is capable of generating or has just generated. With knowledge, albeit probabilistic, of the level of ground shaking at a regional level, one could more accurately estimate, for example, 1) the monetary losses caused to specific structures owned by a corporation or insured by an insurance company, 2) the number of injuries, casualties, and homeless people in a region; 3) whether the access to certain critical buildings, such as hospitals, might be restricted due to yellow or red tagging; and 4) the probability that distributed lifeline networks for power, water, and transportation may be interrupted.

The probabilistic assessment of ground motion parameters (e.g., peak ground acceleration or spectral quantities) at an individual site based on the event magnitude, the source-to-site distance, and the local soil conditions is a consolidated practice that started in the late 60's. Much less attention has been devoted, however, to estimating the statistical dependence of ground motion intensities from a single event at multiple sites. (Note that here we do *not* intend to study the similarity, or coherence, in the time domain or frequency domain of ground motion signals at a point in time but rather the correlation of two peak values of oscillator response observed over the entire duration of the ground motion.) In general, three effects account for correlation of ground motion parameters at two sites: a) they have been generated by the same earthquake (e.g., a high stress-drop earthquake may generate ground motions in the region that are, on average, higher than the median values from events of the same magnitude at all sites); b) the seismic waves travel over a similar path from source to site; and c) similar soil conditions (due to possible bias in the prediction equation used to estimate the median ground motion at the sites). Modern ground motion attenuation equations implicitly recognize the first cause of dependency via a specific inter-event error term and, naturally, minimize the effect of the third. The second source of correlation, which is not addressed in attenuation relationships for single sites, is crucial for the spatially distributed applications addressed here. Limited research on this topic to date indicates that correlation of peak ground acceleration or velocity values decreases with increasing spacing between two sites. The few published models, however, do not agree on the amount of correlation and on how fast the correlation dies down with distance. The site-to-site correlation of other ground motion parameters that are good predictors of structural response, such as elastic spectral acceleration, $S_a(T)$ at a period T, to date has been only partially investigated.

This study is divided in three parts. Firstly, we investigate the nature of spatial correlation in synthetic ground motion fields both to support results found using real recordings (see companion report by Jayaram and Baker, 2008b), and also to explore issues, such as isotropy and second-order stationarity of the correlation function, that cannot be studied with real data because the number of recordings for any single event is never sufficiently large. Secondly, we show how the Vector Probabilistic Seismic Hazard Analysis (VPSHA) tool can incorporate the spatial correlation of ground motion and be successfully used for computing the joint seismic hazard a multiple sites. The joint hazard forms the basis for assessing the likelihood that a portfolio of structures at those sites may incur monetary losses due to ground shaking effects. Thirdly, we explicitly model the ground motion spatial correlation in assessing earthquake losses using a Monte Carlo simulation technique for portfolios comprising a number of structures that is too large for the VPSHA tool to handle via numerical integration.

1 SCOPE OF WORK AND MOTIVATION

Many private and public stakeholders are strongly affected by the impact of earthquakes on a regional basis rather than on 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 level in order to plan their emergency responses in a timely manner. The same organizations also need to assess risks from future earthquakes before they occur 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 intensity at multiple sites throughout the affected region.

Earthquakes that occur in some seismically active areas of the world such as California, Japan and Taiwan, are recorded by extended networks of stations, while other areas such Mexico and Turkey are less well instrumented. In any of these regions, however, the severity of the shaking is only known at the station locations, while the shaking at other locations is uncertain. The level of uncertainty increases with increasing distance from nearby recording stations. The damage to structures, and therefore the economic and life loss caused by earthquakes, has been shown to correlate well with ground motion parameters that measure the peak response of simple single-degree-of-freedom (SDOF) oscillators with the same fundamental period of the real structures. For other applications, such as the prediction of liquefaction of saturated sandy soil or the response of buried pipelines, the peak responses of the ground, such as the horizontal Peak Ground Acceleration (PGA) and Velocity (PGV), are considered to be more accurate estimators of the damage severity. With improved knowledge of the probability distributions of relevant ground motion parameters at a regional level, one could more accurately estimate, for example, 1) the monetary losses associated with structures owned by a corporation or insured by an insurance company, 2) the number of injuries, casualties, and homeless people in a certain area; 3) whether the access to certain critical buildings, such as hospitals, is likely to be restricted due to likely yellow or red tagging; 4) the probability that lifeline networks for power, water, and transportation may be interrupted.

1.1 Why are ground motion intensities at different sites correlated random variables?

The probabilistic assessment of ground motion parameters at a site based on the magnitude of the event, the source-to-site distance, and the local soil conditions is a consolidated practice that started in the late 60's. Many predictive equations have been developed either from empirical data or from mathematical models when historical data are insufficient. These equations are appropriate for estimating the parameters of the ground motion at a specific site. Much less attention has been devoted, however, to estimating the statistical dependence of ground motion intensities from a single event at multiple sites. If this dependency is neglected then ground motion values that are, for example, consistently very large over a spatially extended area will have a negligible probability of being observed. As the Northridge earthquake showed clearly, these cases do occur, and can cause severe damage over a large area that cannot be predicted unless spatial ground motion correlation is adequately modeled.

In general, the values of a ground motion parameter at two sites are correlated for three reasons[1]: a) they have been generated by the same earthquake (e.g., a high stress-drop earthquake may generate ground motions that are, on average, higher than the median values generated by events of the same magnitude); b) the seismic waves travel over a similar path from source to site; and c) homogeneity of site soil conditions (only present if the median ground motion estimates for those soil conditions are biased in the

[1] 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 space sites but rather investigates the correlation of peak values of oscillator response (or of the ground motion) observed over the entire duration of the shaking.

prediction equation). Modern ground motion attenuation equations implicitly recognize the first cause of dependency via a specific inter-event error term, η_i, as follows:

$$\ln Y_{i,j} = \overline{\ln Y_{i,j}} + \eta_i \tau + \varepsilon_{i,j} \sigma \qquad (1)$$

Where $Y_{i,j}$ is the ground motion parameter of interest (e.g., $S_a(T_1)$), $\overline{\ln Y_{i,j}}$ is the median value of the log of Y predicted by the attenuation equation at site j for the magnitude-distance pair of earthquake i and local site conditions, η_i is the aforementioned inter-event standard normal error term, $\varepsilon_{i,j}$ is the site-to-site intra-event standard normal error term, and σ and τ are the corresponding standard deviations of the two error terms, or "residuals."

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

$$\ln Y_{i,j} = \overline{\ln Y_{i,j}} + \tilde{\sigma}\tilde{\varepsilon}_{i,j} \qquad (2)$$

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

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

In the context of assessing site-to-site correlation of ground motion IMs, it is convenient to use the model in Equation 2 for at least three reasons: a) there is now only one residual term for each observation ($\overline{\ln Y_{i,j}}$ and $\tilde{\sigma}$ are provided by ground motion prediction equations, and $Y_{i,j}$ is observed, so $\tilde{\varepsilon}_{i,j}$ can be computed directly); b) the residual term is easy to compute (the values of η_i, i=1,…,N, for all the N earthquakes and the frequency-dependent values of τ are usually not included by the developers of ground motion prediction equations in their publications); and c) Equation 2 is also the form commonly used in probabilistic seismic hazard analysis (PSHA) computer programs, so spatial correlation models in this format can be more easily incorporated into existing software.

Example observed $\tilde{\varepsilon}_{i,j}$ residuals from the 1999 Chi-Chi, Taiwan, earthquake are shown in Figure 1; these residuals, whose value is indicated by the color scale, include both the inter- and intra-event error terms.

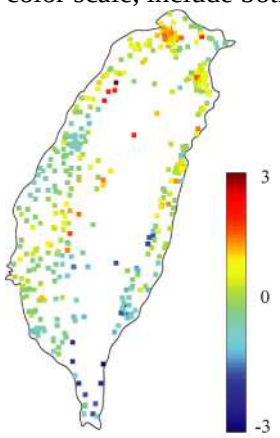


Figure 1: Observed attenuation residuals from the 1999 Chi-Chi, Taiwan, earthquake

The second source of correlation, namely the correlation between the two random variables $\varepsilon_{i,j}$ and $\varepsilon_{i,k}$ at two different sites j and k, has not yet been fully investigated. Spatial dependence can be observed in Figure 1, by noting that residuals at nearby locations take similar values.

This intra-event site-to-site correlation, which is of course not addressed in attenuation relationships for single sites, is crucial for the spatially distributed applications mentioned above. The limited research on this topic to date indicates that the correlation of peak ground acceleration or velocity values decreases with increasing spacing between two sites. This correlation can be estimated by computing empirical correlation coefficients for $\varepsilon_{i,j}$ values at a site separation distance h (plus or minus some tolerance). Because the η_i value is fixed for each i^{th} earthquake, it is effectively a constant when empirical correlation coefficients are estimated from a single earthquake. Thus, correlation coefficients obtained from $\varepsilon_{i,j}$ values or $\tilde{\varepsilon}_{i,j}$ values will be identical, but these correlation coefficients only represent the correlation in the $\varepsilon_{i,j}$ values. To obtain correlation coefficients for the $\tilde{\varepsilon}_{i,j}$ values, one must add the effect of the η_i random variable, which is perfectly correlated at all distances but cannot be detected from the previous empirical correlation coefficients. The total correlation in $\tilde{\varepsilon}_{i,j}$ values is thus

$$\tilde{\rho}(h) = \frac{\tau^2 + \sigma^2 \rho(h)}{\tilde{\sigma}^2} \tag{4}$$

where $\rho(h) = \rho_{\varepsilon_{i,j,1},\varepsilon_{i,j,2}}(h)$ is the empirical correlation coefficient calculated for intra-event $\varepsilon_{i,j}$ values separated by a distance h, and $\tilde{\rho}(h) = \rho_{\tilde{\varepsilon}_{i,j,1},\tilde{\varepsilon}_{i,j,2}}(h)$ is the correlation coefficient for the total $\tilde{\varepsilon}_{i,j}$ values defined in Equation 3. Note that for very close sites (i.e., $h \to 0$) the correlation $\tilde{\rho}(h)$ of IMs, of course, tends toward one, whereas for very distant sites (i.e., $h \to \infty$) it is simply given by the ratio of the inter-term-variance to the total variance, as expected. The variance of the difference of the same IM quantity at two sites, k and l, separated by a distance h is simply

$$VAR[X_k - X_l] = 2\sigma^2(1 - \rho(h)) \tag{5}$$

Some researchers (e.g., Boore *et al*., 2003; Kawakami and Mogi; 2003; Wang and Takada, 2005; Jeon and O'Rourke, 2005), have shown that the correlation between residuals of horizontal Peak Ground Acceleration, PGA, or velocity, PGV, at two sites decrease with increasing site spacing. Their models, however, do not agree on the amount of correlation and on how fast the correlation dies down with distance. Also the site-to-site correlation of other ground motion parameters, such as elastic spectral acceleration, $S_a(T)$ at a period T, that are better predictors of structural response than PGA and PGV received to date only limited attention (Goda and Hong, 2008). Furthermore, these models all have shortcomings with respect to the proposed applications, as will be discussed below.

This project is an attempt to fill this research gap. This study explores the site-to-site correlation of ground motion parameters in more depth. It is important to emphasize here that the correlation that we consider is between the intra-event ground motion residuals at different sites from the same earthquake, namely the $\varepsilon_{i,j}$ values generated by event i at different sites j,with j=1,..,N . The residuals are obtained by removing the median ground motion predicted for the site by a selected attenuation equation. We work with the residuals rather than with the absolute values of the ground motion parameters in order to remove the effects of the local soil conditions that, otherwise, would obscure the ground motion parameter correlation structure.

Because residuals of ground motion parameters generated at different sites by the same event are well modeled by a multivariate Gaussian distribution, the spatial dependence of a pair of parameters at two sites (e.g., $S_a(T_1)$ at the first site and $S_a(T_2)$ at the second site) is fully characterized by the correlation coefficient. In particular, we consider models for correlation as a function of the a) intensity parameters (i.e., peak values and spectral quantities); b) orientation of the recordings (e.g., fault-normal and fault-parallel components versus randomly oriented components); and c) definition of the intensity parameters (e.g., arbitrary horizontal component versus the geometric mean of the two horizontal components). We also test whether the correlation depends on other characteristics of the event and site (e.g., the magnitude of the event, the distance of the sites from the causative rupture, and Somerville *et al.*, 1997, directivity parameter) and on the relationship that is selected to predict the median motion at each site.

As described in the next section, we derived the site-to-site intra-event correlation structure using both real seismograms from well-recorded past earthquakes and simulated recordings from hypothetical events. Identification of some of the more subtle spatial dependence properties requires a very large number of data points to keep the effect from being masked by noise due to limited sample sizes. Recorded ground motion libraries will be critical for developing the basic models, but simulated ground motion fields will be used to model more subtle effects which may require many thousands of data points to identify. While the simulations cannot completely replace the studies using recorded data, they may be useful for refining correlation models.

The results of the study based on real recordings are presented in the companion report from Stanford University (Jayaram and Baker, 2008b). **Section 2** of this report presents the results of the study on spatial correlation based on synthetic accelerograms.

The two applications of these models for site-to-site correlation of ground motion parameters newly derived in this study are:

1) Computation of the mean rate of occurrence of different pairs of ground motion parameters at two different sites using a modification of the recently introduced vector-valued Probabilistic Seismic Hazard Analysis (VPSHA) and application to a small portfolio of structures; (**Section 3**)
2) Evaluation of the effects of ground motion parameter correlation on loss estimates of portfolios of properties with different spacing patterns. (**Section 4**)

1.2 Relevance of this study for seismic risk mitigation

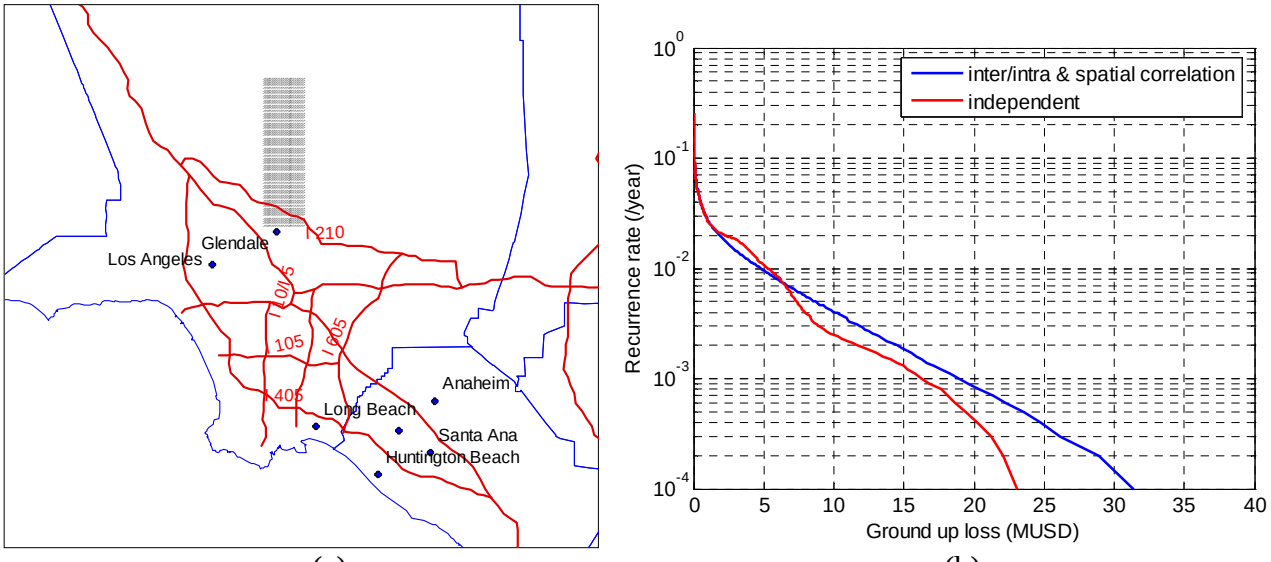
The relevance of this study to reducing losses from earthquakes in the U.S. is direct and encompasses several different aspects. First and foremost this work provides a site-to-site ground motion parameter correlation structure that is consistent with recordings from historical earthquakes. This is the crucial building block of seismic risk analyses of multi-property portfolios or spatially distributed systems, as explained below. Until now the overwhelming majority of such seismic risk analyses are performed either considering the ground motion variability for each earthquake scenario but neglecting the site-to-site ground motion correlation, or by ignoring ground motion variability altogether and assuming median ground motion everywhere in the affected region. Both approaches, and more so the latter, lead to unconservative risk estimates by failing to predict large consequences that can only be caused by larger than expected ground motions for spatially extended areas. These conditions can only be captured when the appropriate site-to-site ground motion correlation structure is adopted.

In addition to modeling ground motion correlation, we show its use in two distinct seismic risk assessment applications that can be used for risk understanding, communication, and, eventually, mitigation.

In the first application, the joint hazard at multiple sites is computed by modifying the VPSHA tool that

has been recently introduced by one of the PI's of this study (see, among others, Bazzurro, 1998; Bazzurro and Cornell, 2002; Luco *et al*., 2005a and 2005b) for estimating the joint hazard of two or more ground motion parameters at the same site. This new application demonstrates the potential of VPSHA as a tool for seismic risk analyses of spatially extended systems. The product of this research could be adapted as an interactive tool on the USGS website that would enable a user to input the coordinates, for example, of two building sites, specify the fundamental period of vibration along one of the principal axes of the two buildings (i.e., T_{B1} and T_{B2}), and obtain plots and tables with joint rates of exceedance of different pairs of values of $S_a(T_{B1})$ and $S_a(T_{B2})$. The site could also provide all $S_a(T_{B1})$ and $S_a(T_{B2})$ pairs whose joint exceedance is equal to some target value, such as 10% in 50 years. With the target exceedance rate and a pair of $S_a(T_{B1})$ and $S_a(T_{B2})$ values taken from the output, the user could also ask for a plot of the disaggregated magnitude and distance pairs of the scenarios that most contribute to the exceedance of those values at the site. Such a tool could be very valuable for computing probabilistically sound estimates of the damage and losses that the portfolio of the two buildings may suffer.

In the second application we show how important it is for the accuracy of portfolio loss estimates to properly model the ground motion spatial correlation. Figure 2 shows some results for a hypothetical portfolio of 1,200 identical woodframe houses worth $100,000 each located on a regular rectangular grid with 500m spacing between consecutive sites (Figure 2a). The curves in Figure 2b show the mean annual rate of exceedance of different level of losses for this portfolio computed first by neglecting ground motion spatial correlation (lower red curve) and then by modeling it according to Boore *et al*. (2003). The difference is most significant for the rare losses. Note that a much larger difference in loss estimates is expected for portfolios with a limited number of properties located in a smaller geographical area. It is clear from this simple example that modeling ground motion spatial correlation is crucial for accurate assessment of future seismic risk. The effects of spatial correlation of ground motion intensities on portfolio losses is investigated in this report using the correlation structure developed in this study rather than that of Boore *et al*. mentioned above.



(a) (b)

Figure 2*:* (b) Mean rate of exceedance curves for different level of losses for a hypothetical portfolio of 1,200 residential wood frame houses located in Southern California (see regular grid shown as a rectangle in Panel a).

1.3 Objectives

There are two main objectives of the proposed research study:

- **Objective 1**: Development of a correlation structure for ground motion parameters generated by the same event at different sites. Previous studies of spatial dependence of ground motions report widely varying results and are also limited to PGA and PGV. The statistical study proposed here covers a greater range of ground motion parameters, and also considers more sophisticated models that include the effects of distance from the fault rupture, magnitude of the event, and directivity parameter of the site, if such effects are supported by the data. The correlation models are first developed using recorded ground motions, but the more complex aspects of the models are investigated with the aid of large datasets from ground motion simulations.
- **Objective 2**: Integration of the correlation models developed here into two different applications: a) computation of the joint probabilistic hazard of ground motion parameters at two different sites via a modification of the VPSHA framework; c) estimation of the earthquake-induced losses to spatially distributed portfolios of many structures while accounting for site-to-site ground motion correlation.

2 SPATIAL CORRELATION OF GROUND MOTION INTENSITIES IN SIMULATED RANDOM FIELDS

This study utilizes simulated ground motions produced by Drs. Brad Aagaard and Robert Graves (Aagaard et al., 2008b, 2008a) to study the spatial correlation between ground motion intensities. Jayaram and Baker (2008b) used geostatistical tools to quantify and establish spatial correlations using ground motions recordings from past earthquakes. They also identified various factors influencing the extent of the spatial correlation, and developed a predictive model that can be used to select appropriate correlation estimates for use in risk assessment problems. While recorded ground motions are a great asset 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) which are commonly used in the development of spatial correlation models. This is on account of the scarcity of ground motion recordings for any particular earthquake.

This limitation is overcome when using synthetic ground motion random fields. The current study utilizes ground motions simulated at 35,547 locations based on source models of the 1989 Loma Prieta earthquake and the 1906 San Francisco earthquakes for verifying these commonly-used assumptions.

2.1 Modeling correlations using semivariograms

Ground-motion models that predict intensities at an individual site *i* due to an earthquake *j* take the following form of Equation 1 in Section 1, which is repeated here for convenience:

$$\ln Y_{i,j} = \overline{\ln Y_{i,J}} + \eta_i \tau + \varepsilon_{i,j} \sigma \qquad (6)$$

where Y_{ij} denotes the ground-motion parameter of interest (e.g., $S_a(T_1)$, the spectral acceleration at period T_1); $\overline{Y}_{ij}$ denotes the predicted (by the ground-motion model) median ground-motion intensity (which depends on parameters such as magnitude, distance, period and local site conditions); ε_{ij} and η_j denote, respectively, the normalized intra-event and inter-event residuals, which are both standard normal random variables. Finally, σ and τ are the corresponding standard deviations of the two types of residuals. These standard deviations are estimated as part of the ground-motion prediction model and are a function of the response period of interest, and in some models also a function of the earthquake magnitude and the distance of the site from the rupture. During an earthquake, the inter-event residual (η_j) computed at any particular period is a constant across all the sites.

Jayaram and Baker (2008a) showed that a vector of spatially-distributed intra-event residuals[2] $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_d)$ follows a multivariate normal distribution. Hence, the distribution of $\boldsymbol{\varepsilon}$ is fully defined by its mean and covariance of the residuals. The mean of ε_i equals zero, while its variance is provided by the ground-motion model. The correlation between ε_i and ε_j is, however, unknown and needs to be estimated. The current work uses geostatistical tools to empirically estimate these correlations using the simulated ground motion data sets. These tools are described briefly in this section;

[2] We dropped the subscript *j* that refers to the j^{th} earthquake to simplify the notation.

a detailed discussion can be found in, for example, Goovaerts (1997), Deutsch and Journel (1998) and in Jayaram and Baker (2008b).

Let $\boldsymbol{\varepsilon}$ denote the set of inter-event residuals distributed over space. The semivariogram of $\boldsymbol{\varepsilon}$ is a measure of the dissimilarity between the residuals and is useful in computing the spatial correlation between the residuals. While working with correlations, it is convenient to work with normalized residuals as compared to ε's since normalized residuals are homoscedastic (i.e., have a constant variance). As mentioned above, normalized residuals ($\grave{\varepsilon}_{ij}$) are obtained by dividing the ε's by the appropriate standard deviations obtained from the ground-motion model. The semivariogram of the random function (a random function is a collection of spatially-distributed random variables) of $\grave{\varepsilon}_{ij}$'s is defined as follows:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\grave{\varepsilon}_{u_\alpha} - \grave{\varepsilon}_{u_\alpha + h} \right]^2 \tag{7}$$

where $\hat{\gamma}(h)$ denotes the experimental (i.e., empirical) semivariogram of the normalized residuals; $(u_\alpha, u_\alpha + h)$ denotes the location of a pair of sites separated by *h* and *N(h)* is the number of such pairs. When empirically estimated, $\hat{\gamma}(h)$ only provides semivariogram values at discrete values of *h*, and hence, a function (e.g., an exponential function) is usually fitted to the discrete values to obtain the semivariogram values for continuous *h*. For instance, an exponential semivariogram takes the following form:

$$\hat{\gamma}(h) = a\left[1 - \exp\left(-3h/b\right)\right] \tag{8}$$

where *a* denotes the "*sill*" of the semivariogram (which equals the variance of the normalized 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 is to be noted that Equations 7 and 8 are implicitly based on the assumptions of stationarity and isotropy. This issue is addressed in detail in Goovaerts (1997), Deutsch and Journel (1998).

Park et al. (2007) and Jayaram and Baker (2008b) showed that correlation coefficients and, therefore, the semivariogram of intra-event residuals (Equation 7) can be directly obtained using total residuals rather than using intra-event residuals (which is advantageous since the empirical data directly provides the total residuals). Let $\tilde{\varepsilon}$ denote the total residual normalized by the standard deviation of the intra-event residual (as provided by the ground-motion model). It can be easily shown that:

$$\begin{aligned} \hat{\gamma}(h) &= \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\grave{\varepsilon}_{u_\alpha} - \grave{\varepsilon}_{u_\alpha + h} \right]^2 \\ &= \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\tilde{\varepsilon}_{u_\alpha} - \tilde{\varepsilon}_{u_\alpha + h} \right]^2 \end{aligned} \tag{9}$$

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

$$\hat{\rho}(h) = 1 - \hat{\gamma}(h) \tag{10}$$

Hence, it can be seen that the correlations are completely defined by the semivariogram, which in turn, is a function only of the range (The sill is known to equal 1, the variance of the normalized residuals for which the semivariogram is constructed.) Moreover, note from equations 8 and 10 that a larger range implies a smaller rate of increase in $\hat{\gamma}(h)$ with *h*, and subsequently, a smaller rate of decay of correlation with separation distance.

In the current study, ranges of semivariograms of residuals are estimated using simulated ground motions. The residuals are computed using the ground-motion model of Boore and Atkinson (2007). Further, the simulated data sets are used to test the assumptions of second-order stationarity and isotropy, used in the development of the spatial correlation model.

2.2 Data sets used

In the current spatial correlation study, we used four simulated ground-motion data sets:

- the 1989 Loma Prieta earthquake data set of Aagaard and Graves (described in Aagaard et al., 2008b);
- the 1906 San Francisco earthquake Song-Mod data set of Aagaard (Aagaard et al., 2008a);
- the 1906 San Francisco earthquake RandomHypo06 data set of Aagaard (Aagaard et al., 2008a); and
- the 1906 San Francisco earthquake RandomHypoC data set of Aagaard (Aagaard et al., 2008a).

The 1989 Loma Prieta simulations are based on the Beroza (1991) and the Wald et al. (1991) source models. The 1906 San Francisco earthquake Song-Mod simulations are based on the Song et al. (2008) source model with modified rise times to match the Boatwright and Bundock (2008) intensities. The 1906 San Francisco earthquake RandomHypo06 simulations are similar to the Song-Mod simulations, except that the slip distribution is randomized. The RandomHypoC simulations are based on a modified epicenter (in Bodega Bay rather than a few kilometers offshore San Francisco, as in the other two cases) and a randomized slip distribution. Ground motions are available at 35,547 sites for the simulations of Aagaard and at 40,000 sites for those from Graves' simulations. Sites on softer soil conditions with $V_{s30} \leq 500$m/s, however, are excluded from the computations, due to limitations in capturing nonlinear soil behavior in the simulations. Also, current limitations in the simulation procedure allowed us to investigate only the spatial correlation of spectral accelerations at periods greater than or equal to 2s. Simulated $S_a(T)$'s values for $T<2$s were considered unreliable.

2.3 Results of the Spatial Correlation Study

2.3.1 The 1989 Loma Prieta earthquake simulations by Aagaard

The total residuals, $\tilde{\varepsilon}$'s, were computed from the fault-normal components for $S_a(T)$ with T=2s, 5s, 7.5s, and 10s using the ground motion prediction equation of Boore and Atkinson (2008). The semivariograms of the residuals at discrete values of the separation distance, *h*, were computed and exponential models were fitted using the visual approach described in Deutsch and Journel (1998) and Jayaram and Baker (2008b). Special attention was paid to obtaining the best fit for values at short separation distances. Capturing the effects of the correlation at nearby locations has a much larger influence on the ground motions at a given site due to the so-called "shielding effect" (Goovaerts, 1997). Jayaram and Baker (2008b) discuss the reasons why this approach, which minimizes the error at short separation distances, provides semivariograms that are superior for practical applications to those fitted using the method of least squares, which minimizes the error over a wide range of separation distances.

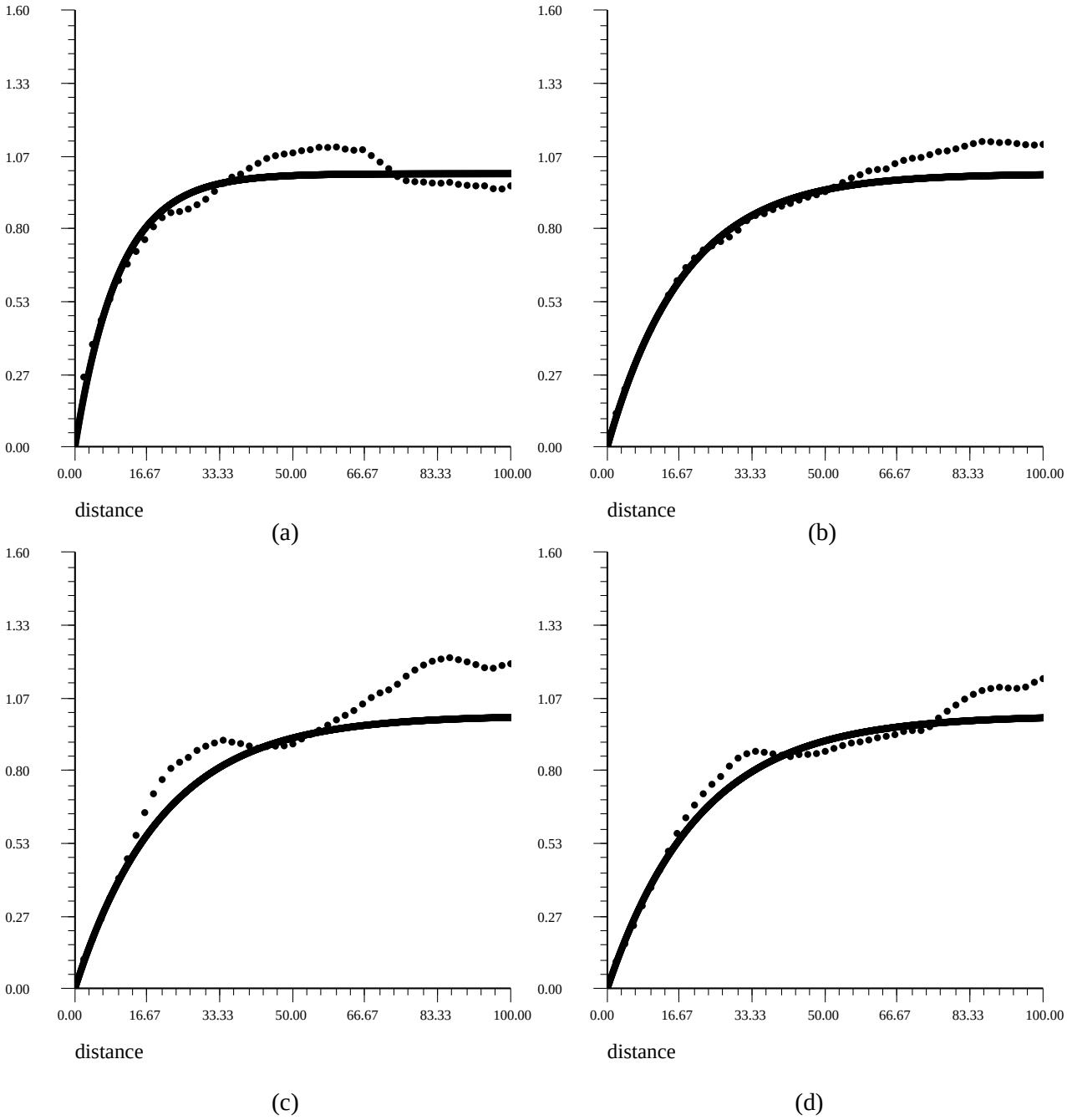
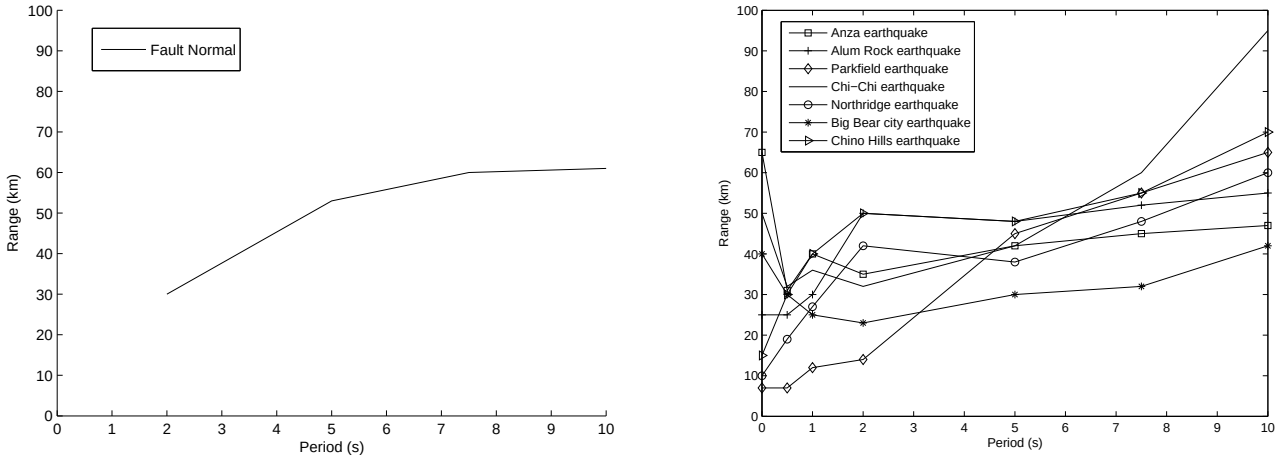


Figure 3: Semivariograms of residuals computed the $S_a(T)$'s from the fault normal component of the 1989 Loma Prieta simulations: Residuals for $S_a(T)$ at (a) T=2s; (b) T=5s; (c) T=7.5s; and (d) T=10s.

The semivariograms obtained using the residuals of the fault-normal components from the simulated data set are shown in Figure 3 for four oscillator periods while the corresponding ranges are plotted in Figure 4a. It can be seen that the values of the range and, therefore, the amount of spatial correlation increases with oscillator period. This trend is to be expected given that the coherency between the period components of the ground motion increases with period (Zerva and Zerva, 2002; 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 (2008b) (Figure 4b). There may be several reasons for the larger spatial correlation of the simulated ground

motions compared to that of the recorded ones and more research is needed to uncover them all. However, as mentioned earlier, the primary reason for using simulated ground-motion sets in this study is to utilize their large number of available time histories to investigate the assumptions of isotropy and second-order stationarity of ground motion intensities. The testing of these assumptions can be carried out irrespective of the extent of correlations observed.



(a) (b)

Figure 4: Ranges of semivariograms at different periods: Residuals computed using (a) Brad Aagaard 1989 Loma Prieta simulations and (b) recorded ground motions (Jayaram and Baker, 2008b)

Effect of ground-motion component orientation on range

In order to test if the orientation of component of the ground motion used has an influence on the estimates of spatial correlation, additional semivariograms of residuals were estimated using the fault parallel, north-south and east-west components of the simulated data set. The ranges of these semivariograms are presented in Figure 5. The range estimates are essentially identical for T=2s and do not show a significant variation on the component used for longer oscillator periods. Hence, most of the following analyses in this section are based on the fault normal components of the simulated ground motions.

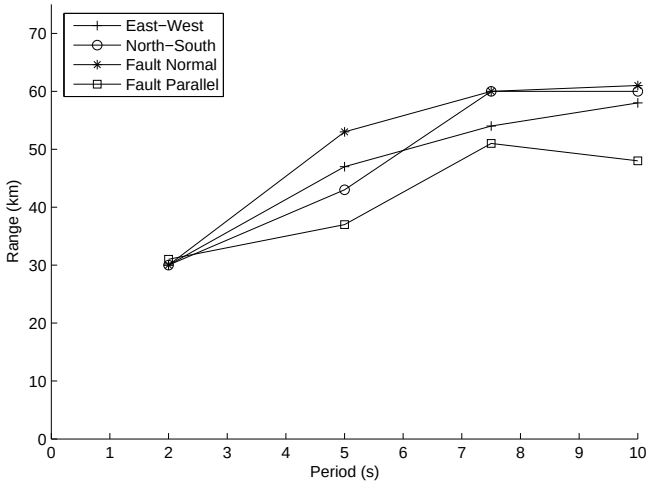


Figure 5: Ranges of semivariograms at different periods. Residuals are computed using Aagaard's Loma Prieta ground motions at different orientations

Testing the assumption of isotropy using directional semivariograms

Directional semivariograms of residuals (Deutsch and Journel, 1998, and Jayaram and Baker, 2008b) are obtained as shown in Equation 9 except that the estimates are obtained using only pairs of $(z_{u_\alpha}, z_{u_\alpha+\boldsymbol{h}})$ such that the azimuth of the vector $\boldsymbol{h}$ is identical 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 different pre-specified azimuths will differ from one another and from the omni-directional semivariogram (i.e., the semivariogram obtained using all pairs of points irrespective of the azimuth).

Figure 6a and Figure 6b compare the omni-directional semivariogram with the semivariograms obtained considering azimuths of 0°, 45° and 90° for T=2s and T=10s, respectively. All the semivariograms are almost identical for separation distances below 10 km, where capturing correlation is most important, and are reasonably close for separation distances between 10 km and 20km. As mentioned earlier, spatial correlation of ground motion intensities between sites separated by more than 20 km need not be modeled with great accuracy on account of the shielding effect. Therefore, based on this data set, it can be concluded that the correlations can be adequately represented using an isotropic model.

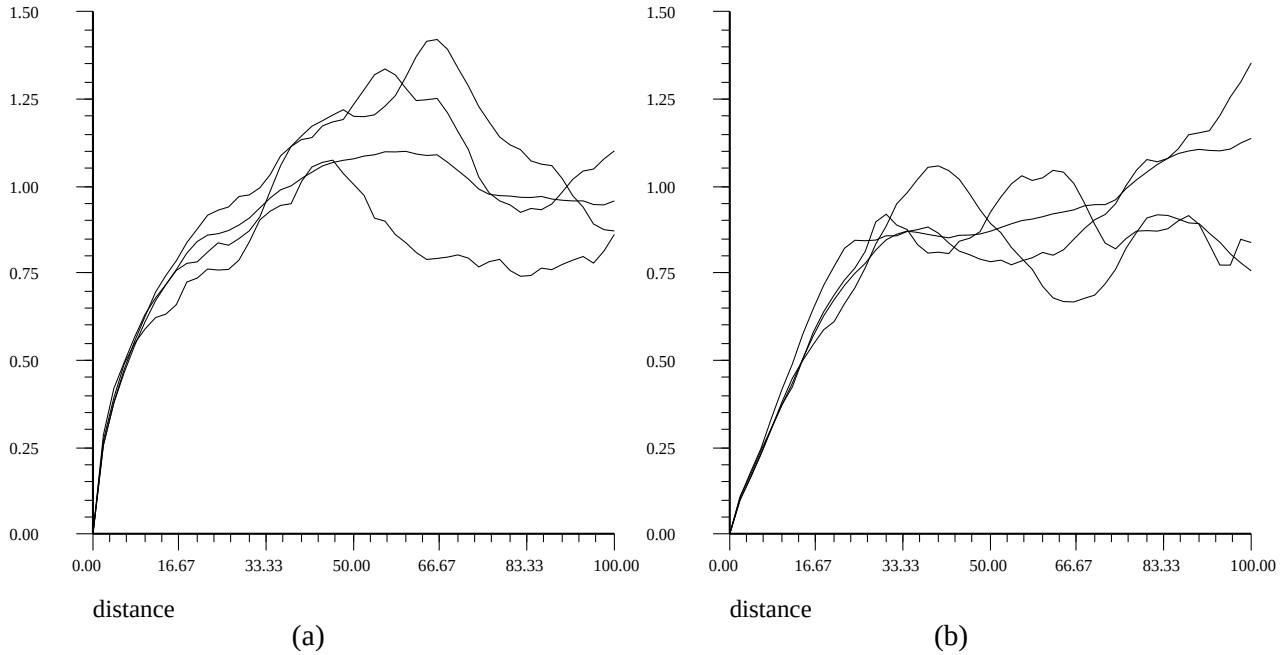


Figure 6: Semivariograms computed from components of different azimuth angles of the Aagaard 1989 Loma Prieta simulation data set: Residuals for (a) T=2s; (b) T=10s. The omni-directional semivariograms for T=2s and T=10s were also shown in Figure 3a and 3d, respectively.

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 separation distance between u and v and not on the actual locations. In other words, the covariance is the same between any two sites that are at the same distance and direction no matter which sites are chosen. The assumption of second-order stationarity is not only convenient while developing correlation models since it allows the data available over the entire region of interest to be pooled together but also simplifies considerably the application of the models.

The assumption of second-order stationarity can be verified by comparing semivariograms constructed exclusively using residuals at sites belonging to different spatial domains. If the semivariograms are similar, it will 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, if there is one. Here, the width of the domains was selected judiciously to avoid both pitfalls above.

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. Only residuals at sites that belong to a particular spatial domain are then used to compute the semivariograms for that spatial domain. The ranges of these semivariograms corresponding to the seven distance domains are reported in Figure 7, along with the ranges corresponding to the semivariogram obtained by pooling all the fault normal residuals together regardless of the distance from the rupture. It can be seen that the semivariograms computed using residuals at sites at 20-160km of the rupture are closer to the semivariogram obtained using all fault normal residuals than those from bins that are closer and farther from the fault. Semivariograms corresponding to a distance bin farther than 160 km from the rupture show significantly smaller ranges, as do the semivariograms for the distance bin within 20 km of the rupture. The ground motion values 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 may not be very critical. On the contrary, it is 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.

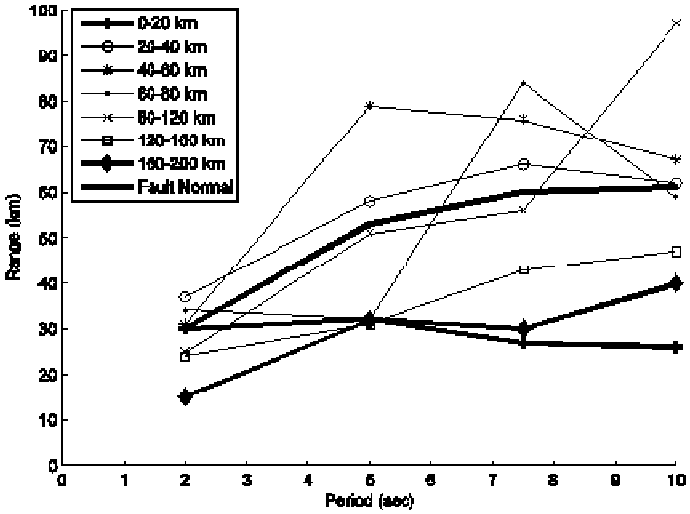


Figure 7: Ranges of semivariograms at different periods. Residuals are computed using Aagaard's 1989 Loma Prieta simulated ground motions at different distances from the fault rupture.

Effect of directivity on spatial correlation

Ground-motions at near-fault sites are typically influenced by directivity effects, resulting 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 in the ground motion prediction model. We intend to study here whether ground motions that show directivity effects are spatially correlated in a dissimilar way from those that do not. 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) was 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 were computed based on these ground motions and semivariograms of the residuals were developed. The estimated ranges (Figure 8) 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 (Figure 7). For a comparison, Figure 8 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 all practical purposes (e.g., risk assessment of building portfolios), the only information required are the ranges computed based on all near-fault ground motions (shown in Figure 7) unless sites where pulse-like ground motions will be present can be accurately predicted.

Finally, it also to be noted that the pulse-like ground motions simulated for the Loma Prieta earthquake are concentrated in a fairly small region. As a result, the estimates of spatial correlation of pulse-like ground motion may not be very robust. The results obtained from the 1906 San Francisco earthquake simulations described later will not have this limitation.

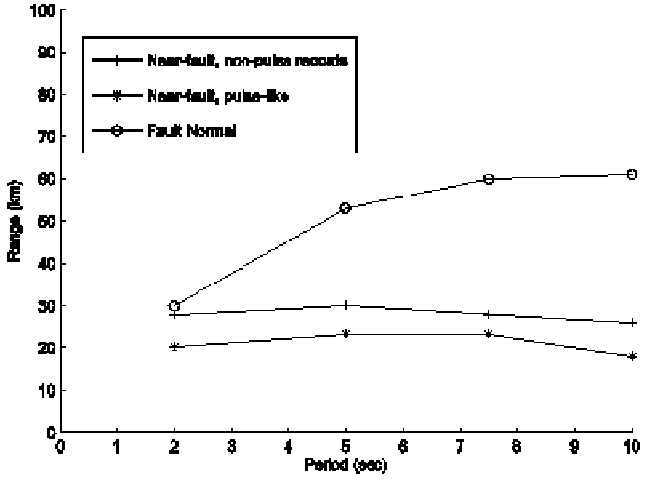


Figure 8: Ranges computed using pulse-like and non-pulse-like near fault 1989 Loma Prieta ground motions.

2.3.2 The 1989 Loma Prieta earthquake simulations by Graves

All the tests carried out on the Aagaard's simulations discussed in Subsection 2.3.1 were repeated using this data set by Graves. Again, we used as base case the fault-normal components. Figure 9 and Figure 10 show the semivariograms estimated using the residuals computed for PGA, and for $S_a(T)$ at T equal to

0.5s, 1.0s, 2.0s, 5.0s, 7.5s, and 10.0s, while Figure 11 displays the corresponding ranges. The ranges at the shorter periods are close to the values obtained in the previous subsection, but the values of the long-period ranges are slightly larger than previously observed. Investigating the reasons behind this difference is beyond the scope of this study. However, we speculate that this higher correlation may be partly due to a procedure called "bulldozing" adopted by Graves which converts the 3D surface of the earth in to a plane surface. This procedure may introduce additional systematic errors in the generated ground motions, which would then result in an increase in the range of the semivariograms. Figure 12 shows the ranges of semivariograms obtained from the residuals of the fault-normal (also shown in Figure 11), fault-parallel, north-south, and east-west components. As with the previous data set, it can be seen that the ranges are reasonably close, irrespective of the ground motion component used. Therefore, the subsequent analyses are based on only the fault normal residuals. The assumption of isotropy was verified further by comparing the directional (for selected azimuth angles) and the omni-directional semivariograms of the residuals computed at T=2.0s and T=10.0s (Figure 13). The directional semivariograms and the omni-directional semivariograms match reasonably well at short separation distance, thereby indicating that isotropy is a reasonable assumption.

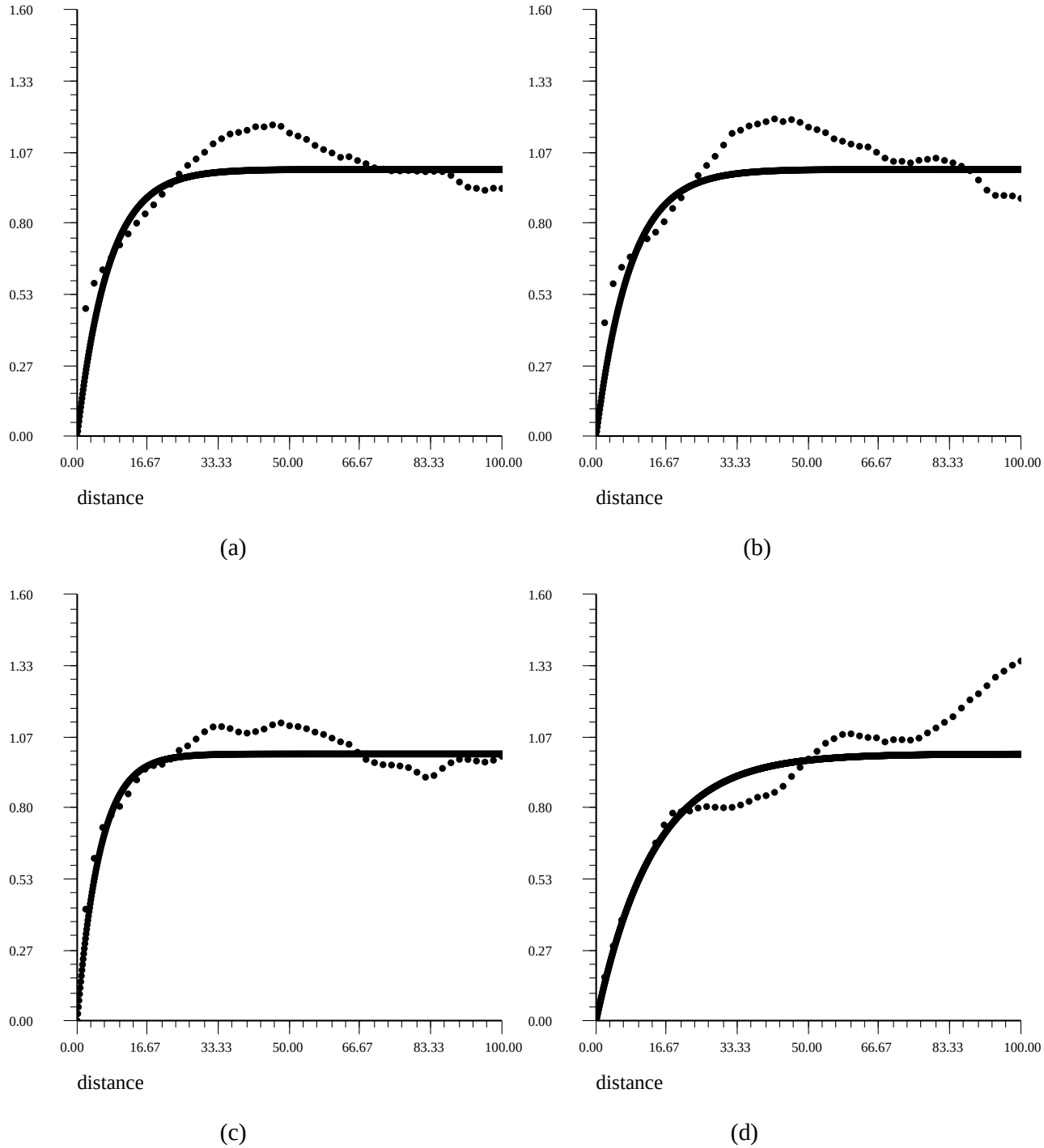


Figure 9: Semivariograms of residuals computed using the fault normal component of the simulated 1989 Loma Prieta ground motions by Graves. Residuals computed for (a) PGA; and $S_a(T)$ at (b) T=0.5s; (c) T=1.0s; and (d) T=2.0s

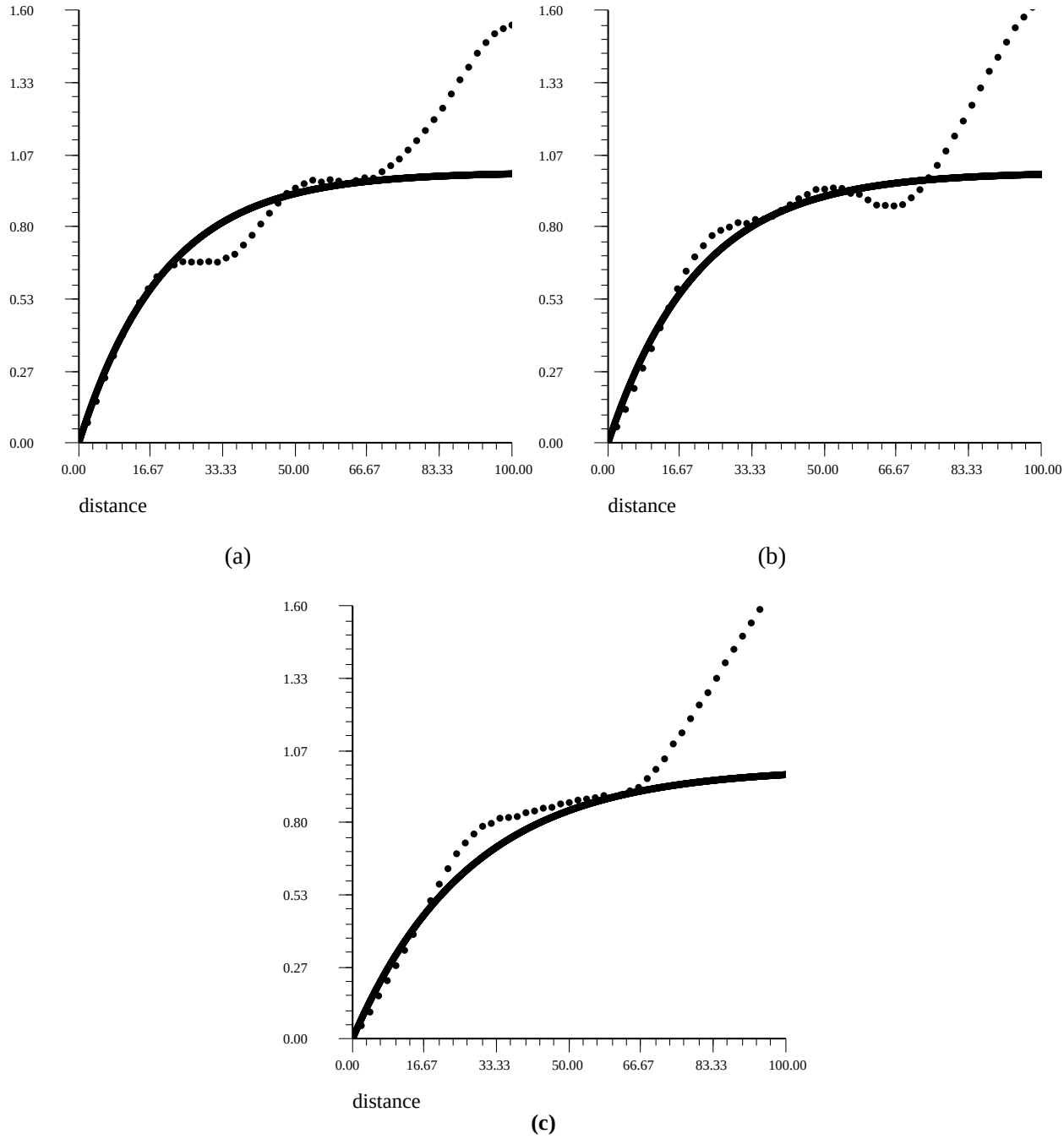


Figure 10: Semivariograms of residuals computed using the fault normal component of the simulated 1989 Loma Prieta ground motions by Graves. Residuals computed for $S_a(T)$ at (a) 5.0s; (b) T=5.0s; and (c) T=10.0s.

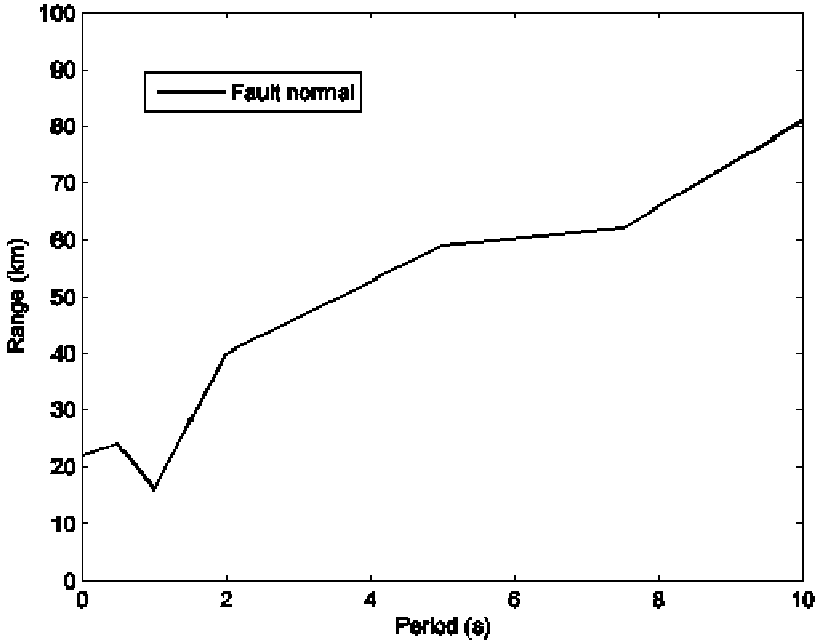


Figure 11: Ranges of semivariograms of residuals computed using the fault-normal components of the ground motions simulated by Graves for the Loma Prieta earthquake.

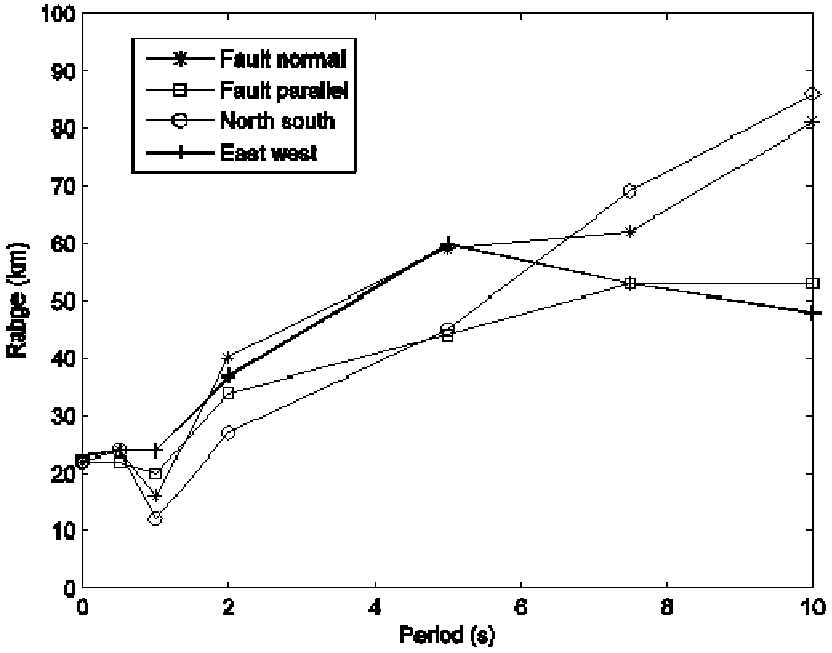


Figure 12: Ranges of semivariograms of residuals computed using different orientations of the ground motion components simulated by Graves for the Loma Prieta earthquake.

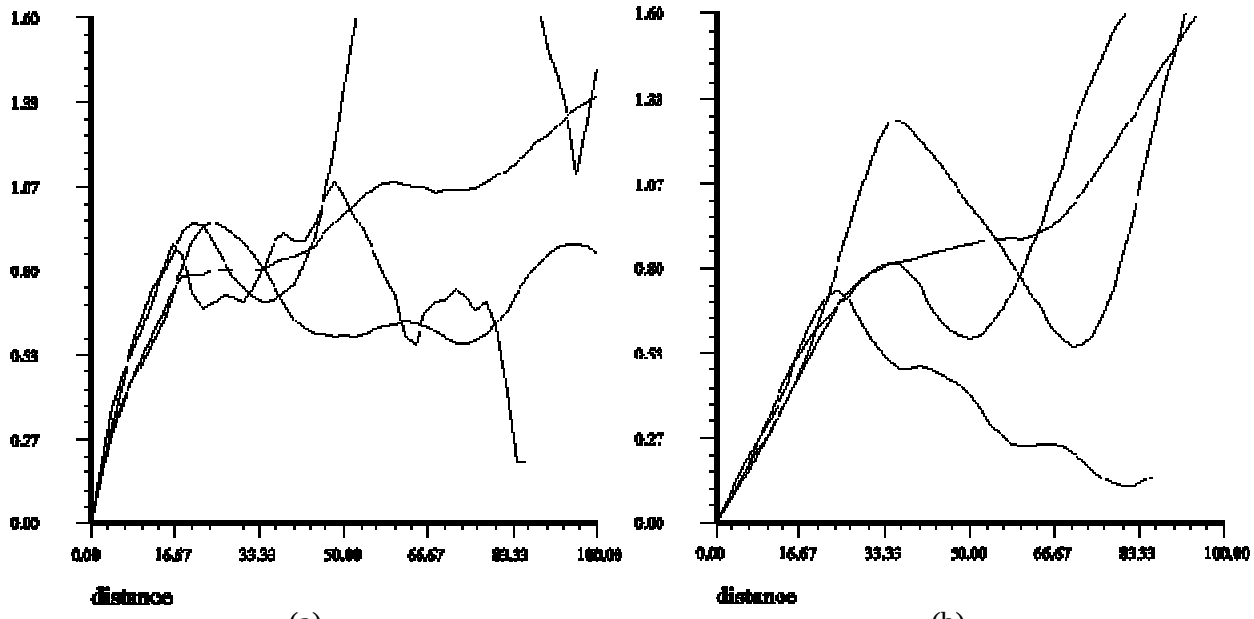


Figure 13: Semivariograms computed from components of different azimuth angles of the 1989 Loma Prieta simulation data set by Graves: Residuals for $S_a(T)$ at (a) T=2s; (b) T=10s. The omni-directional semivariograms for T=2s and T=10s were also shown in Figure 9c and Figure **10**d, respectively.

The test of second-order stationarity of the semivariograms is limited in this case to three distance domains: sites between 0-20km, 20-40km, and 40-60km from the rupture. Sites beyond 60 km were few and scattered and, hence, not considered in the study. The ranges of semivariograms corresponding to these three domains are shown in Figure 14 along with the semivariograms from all the sites. Consistently with the results from the previous data set, ground motions at sites within 20 km of the rupture show a lower spatial correlation than those at farther sites. The ranges for sites within 20km at longer periods, however, are significantly larger than those from the data set by Aagaard. The ranges for sites between 20 and 40 km of the rupture imply a larger correlation than that from the previous data set particularly at long periods. This increase, however, seems to be due to the local effects mentioned before that are generated by a cluster of similar-valued residuals in this distance domain. At shorter periods, however, this local effect is not found, and the ranges are comparable to those obtained using all the fault-normal residuals. The ranges in the 40-60km domain are similar to those obtained using all fault-normal residuals. Therefore, this limited analysis show a reasonable agreement with the conclusions derived using Aagaard's simulations for the same event.

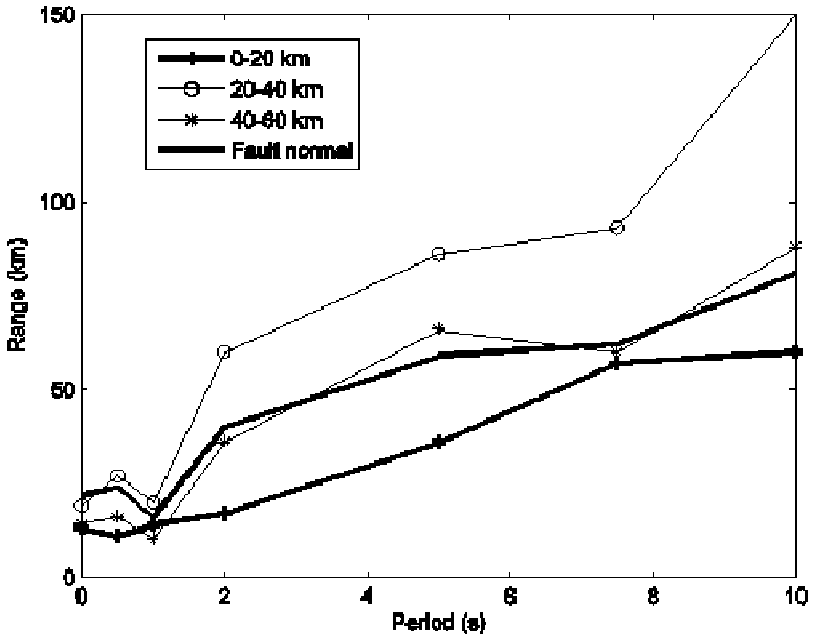


Figure 14: Ranges of semivariograms at different periods. Residuals are computed using Graves' 1989 Loma Prieta simulated ground motions at different distances from the fault rupture.

Finally, to verify the effects of directivity on spatial correlation of ground motion intensities, the same wavelet analysis technique of Baker (2007) identified 1,636 pulses in the fault normal components of this simulated data set (incidentally, 499 fault-parallel pulses were also extracted). Clearly, the much larger number of pulses observed in Graves' simulations compared to that in Aagaard's simulations is indicative of some profound differences between the two techniques. As in the previous section, the pulse-like records are fairly concentrated in a narrow area and, hence, no attempt was made to quantify their spatial correlation. In any case, the most practically-useful near-fault range estimate is that obtained using all recordings within about 20 km from the rupture discussed above.

2.3.3 The 1906 San Francisco earthquake simulations: Song-Mod dataset

In this section, all the tests performed using the 1989 Loma Prieta simulations are repeated using the Song-Mod data set of the 1906 San Francisco earthquake by Aagaard. The epicenter of the earthquake is assumed to be a few kilometers offshore from San Francisco (Long: 122.557; Lat: 37.75). Again, the base case considers the fault-normal components. The analyses are based on residuals for $S_a(T)$ at T=2.0s, T=5.0s, T=7.5s, and T=10.0s computed using the prediction equation of Boore and Atkinson (2008). The semivariograms of the residuals and the corresponding ranges are shown in Figure 15 and Figure 16, respectively. These figures noticeably shows that the estimates of the ranges are larger than those from both the simulated Loma Prieta data set and recorded data sets for the historical events considered in Jayaram and Baker (2008b). Again, it is not unexpected to observe range estimates larger than those from real recordings since limitations in simulation techniques may introduce systematic errors in the ground motions that result in artificially higher spatial correlation.

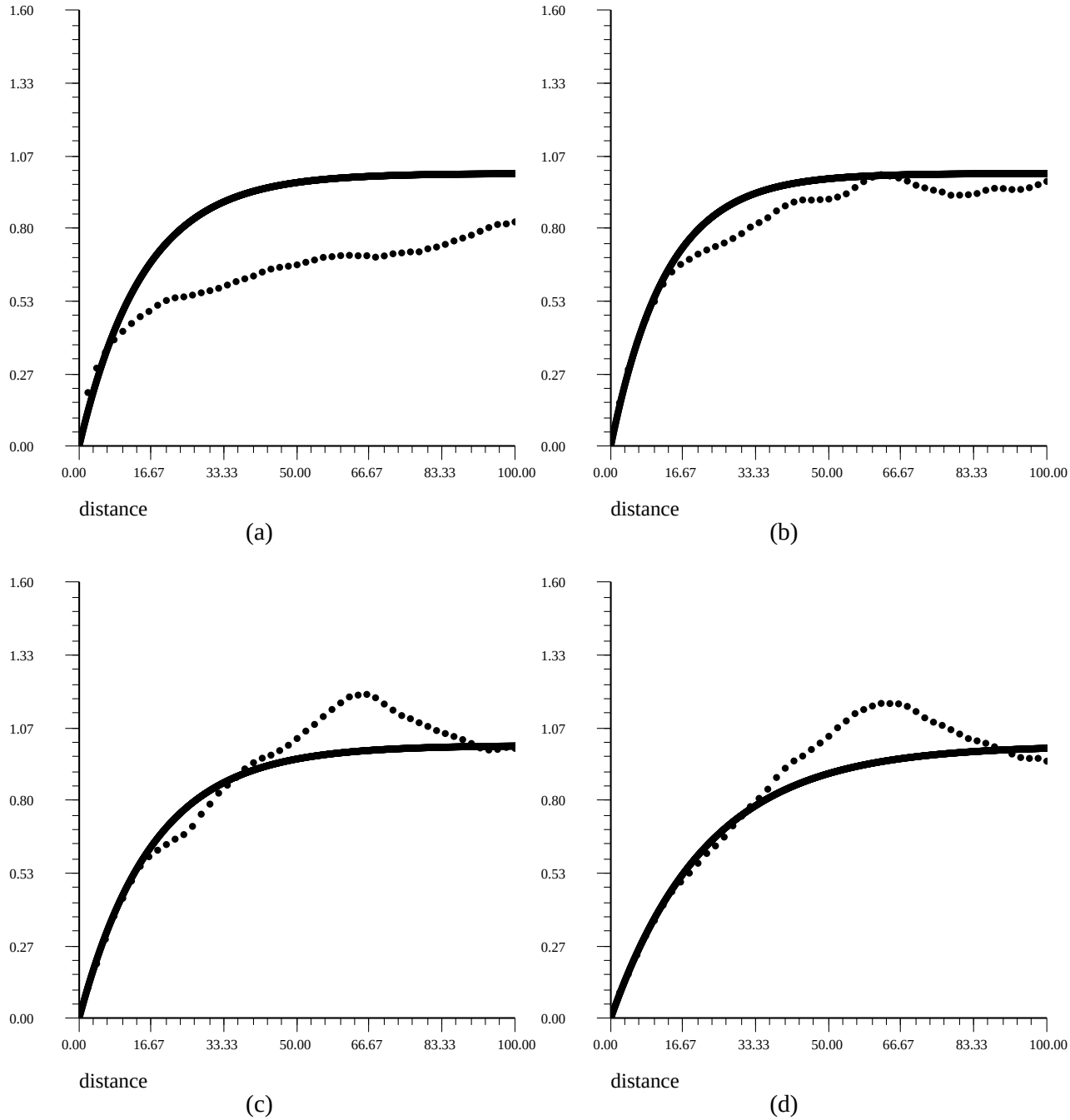


Figure 15: Semivariograms of residuals computed using the fault normal component of the simulated 1906 San Francisco earthquake Song-Mod data set. Residuals computed for $S_a(T)$ at (a) T=2.0s (b) T=5.0s; (c) T=7.5s; and (d) T=10.0s.

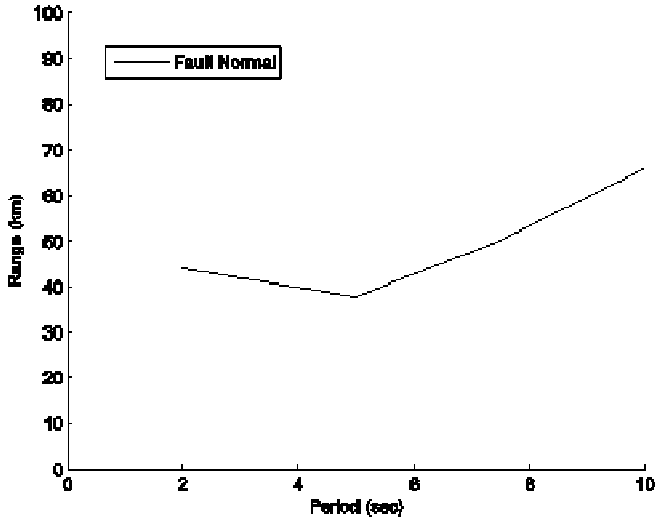


Figure 16: Ranges of semivariograms of residuals computed using the fault-normal components of the 1906 San Francisco earthquake Song-Mod ground motion data set.

Figure 17 compares the ranges of the semivariograms of the residuals computed using the fault-normal, fault-parallel, north-south, and east-west components of the simulations. As seen in the previous section, the ranges do not show a strong dependence on the ground-motion component used and, therefore, subsequent analyses in this section are based on only the fault normal component of the ground motion.

The assumption of isotropy was also verified using directional semivariograms of residuals computed at three different azimuths (0°, 45° and 90°). These directional semivariograms and the omni-directional semivariogram are shown in Figure 18. As seen from the Loma Prieta data set, the directional semivariograms obtained are all reasonably similar to the omni-directional semivariogram, thereby corroborating the isotropic assumption.

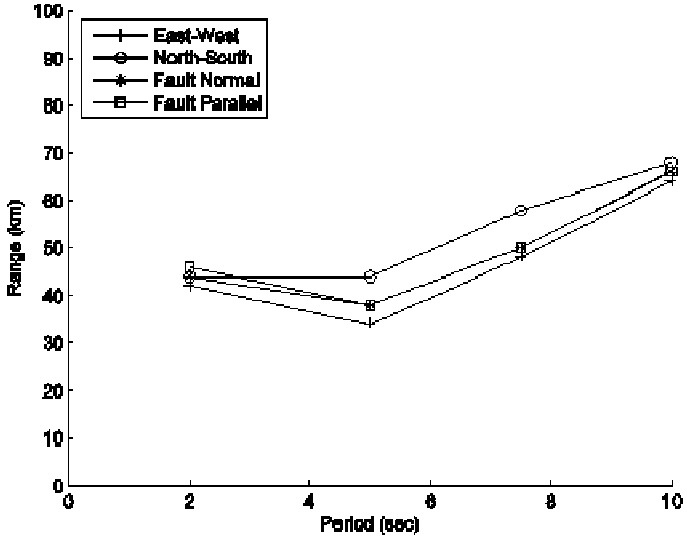


Figure 17: Ranges of semivariograms of residuals computed using components with different orientations from the 1906 San Francisco earthquake Song-Mod ground motion data set.

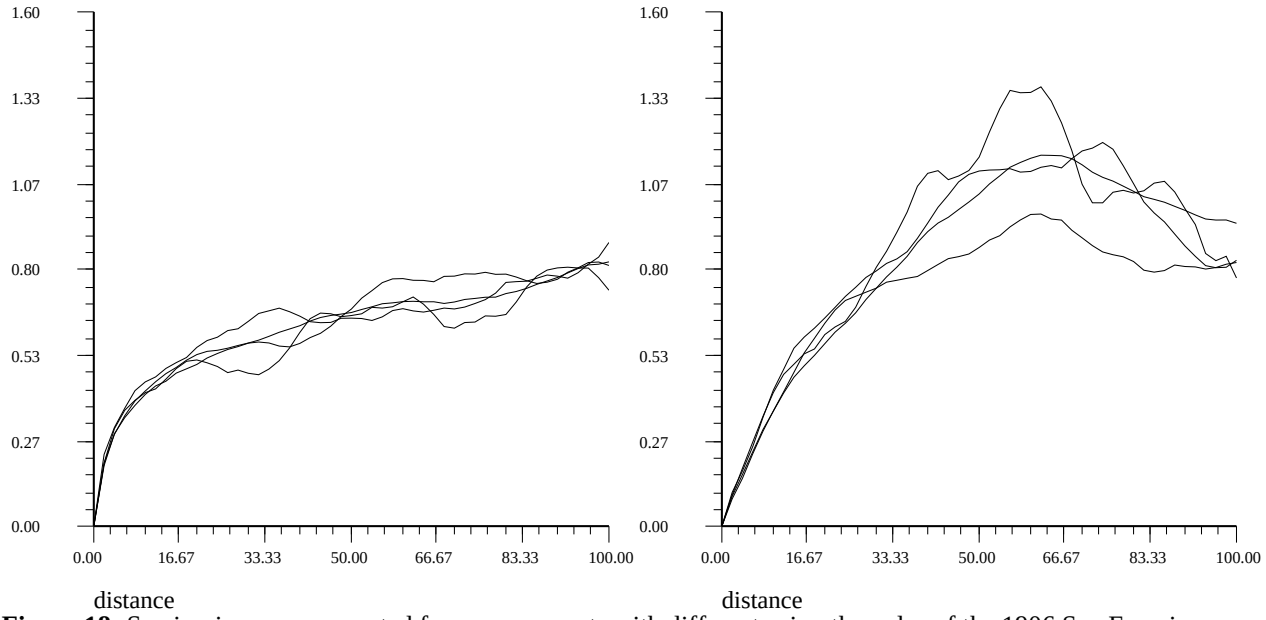


Figure 18: Semivariograms computed from components with different azimuth angles of the 1906 San Francisco earthquake Song-Mod simulation data set by Aagaard: Residuals for $S_a(T)$ at (a) T=2s; (b) T=10s. The omni-directional semivariograms for T=2s and T=10s were also shown in Figure 15a and Figure 15d, respectively.

As before, the assumption of second-order stationarity was verified by constructing semivariograms of residuals at sites belonging to different spatial domains. The test of second-order stationarity of the semivariograms considers four distance domains: sites between 0-10km, 10-30km, 30-50km and farther than 50km from the rupture. The semivariogram ranges for these cases, which are shown in Figure 19, are fairly close to those obtained assuming second-order stationarity (i.e., using all the fault normal residuals). As before, the estimates of the ranges of semivariograms obtained using residuals at sites very close to the rupture are, again, smaller than the average ranges. In this simulated data set, however, the differences are less significant than those observed in the previous cases.

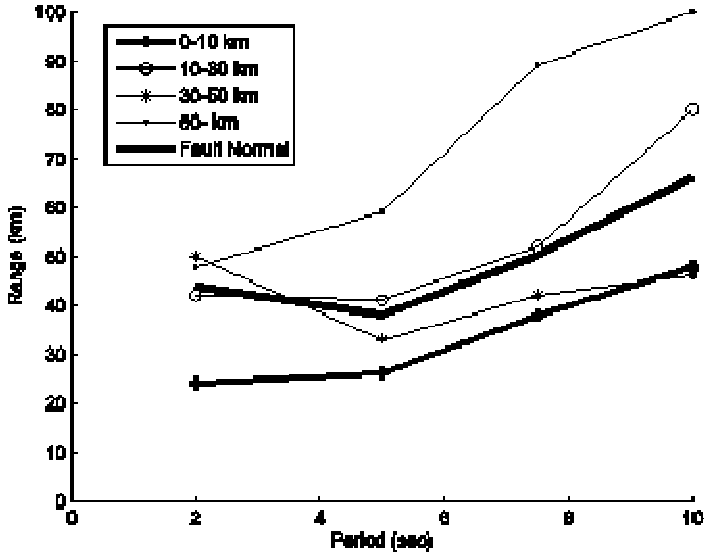


Figure 19: Ranges of semivariograms at different periods. Residuals are computed using the Song-Mod 1906 San Francisco earthquake ground motions at different distances from the fault rupture.

Finally, to verify the effects of directivity, the wavelet analysis technique of Baker (2007) identified 2,577 pulse-like ground motions in the fault normal component of the simulations (and, although not used here, also 2,315 fault-parallel pulses). The estimates of the ranges of semivariograms based on these ground-motion residuals are shown in Figure 20. Also shown in the figures are the ranges of semivariograms obtained using all the fault-normal residuals, the ranges estimated using residuals at sites without pulses but within 10 km from the rupture, and the ranges computed using residuals at all sites within 10 km of the rupture (which are also shown in Figure 19). It can be seen that the ranges obtained using pulse-like ground motions are similar to those estimated using non-pulse-like ground motions, as seen in the previous section.

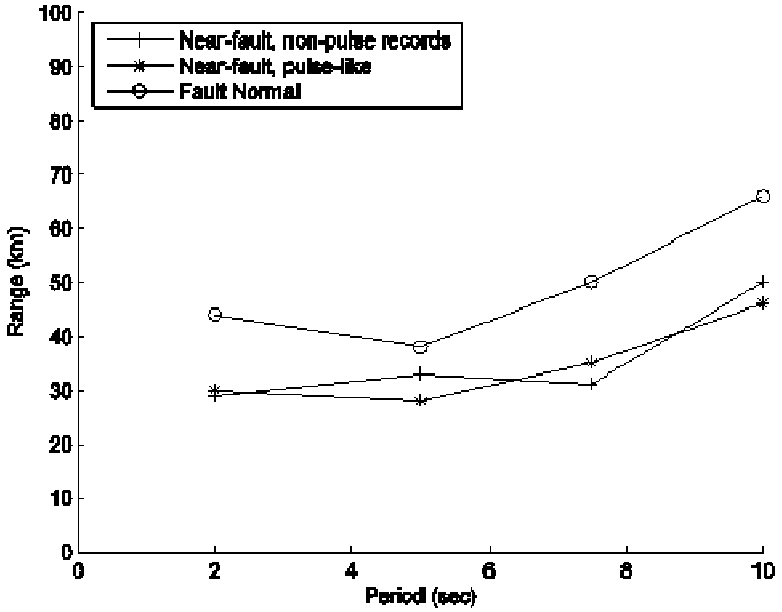


Figure 20: Ranges computed using pulse-like and non-pulse-like near fault Song-Mod ground motions for the 1906 San Francisco earthquake.

Incidentally, note that the number of fault parallel pulses extracted is comparable to the number of fault normal pulses found. This is quite unusual, and is not common in recorded ground motions. The source inversion model of Song et al. (2008), however, includes 100km of super-shear rupture just north of the hypocenter. Aagaard and Heaton (2004) analyzed long-period near-source ground motions from simulations of M7.4 events on a strike-slip fault with super-shear ruptures and concluded that the super-shear ruptures show significant fault-parallel motion before fault-normal motion. Further, they observed that in these cases the maximum horizontal displacements and velocities tend to rotate from the fault normal direction to the fault parallel orientation. Therefore, it is reasonable that the simulations produce a comparable number of fault-normal and fault-parallel pulse-like ground motion time histories.

2.3.4 The 1906 San Francisco earthquake simulations: RandomHypo06 and RandomHypoC datasets

The RandomHypo06 and the RandomHypoC data sets are similar to the Song-Mod data set, except that they are both based on different slip distributions, and the RandomHypoC data set is also based on a different hypocenter (in Bodega Bay, which is located in the northern portion of the 1906 rupture). The estimated ranges of the semivariograms are shown in Figure 21and Figure 22 for the two sets of simulations. Clearly, the ranges estimated using these two data sets are much larger than those estimated using the other two data sets described previously. This is an indication that the simulations for these two

scenarios have produced much more correlations between ground motions than seen in the recorded ground motions. One possible reason for these larger correlations is the use of a more uniform rupture speed in the RandomHypo06 and the RandomHypoC simulations, as compared to that in the Song-Mod simulations. As mentioned previously, however, the main motivation behind using simulated ground motions is to verify the assumptions used in developing the correlation models, rather than to derive the extent of the correlations between ground motions. Hence, these data sets are used for verifying the assumption of second-order stationarity despite the large ranges seen.

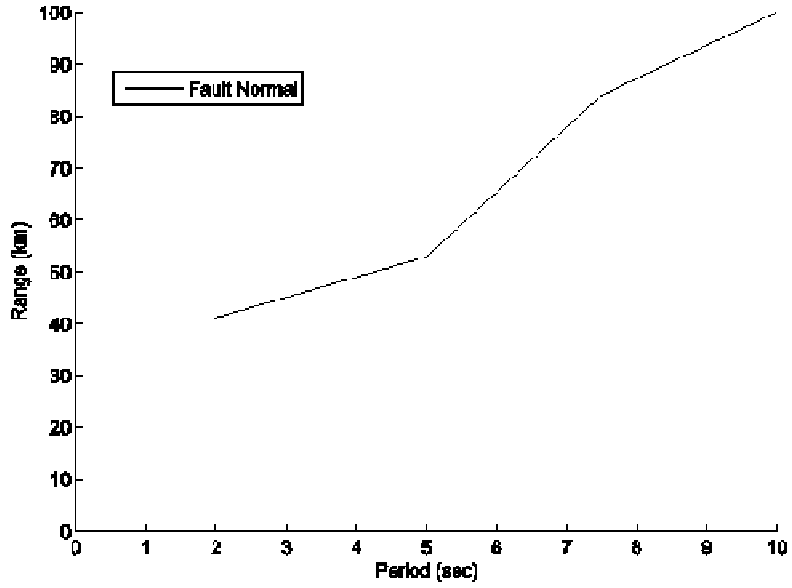


Figure 21: Ranges of semivariograms of residuals computed using the RandomHypo06 simulations of the 1906 San Francisco earthquake.

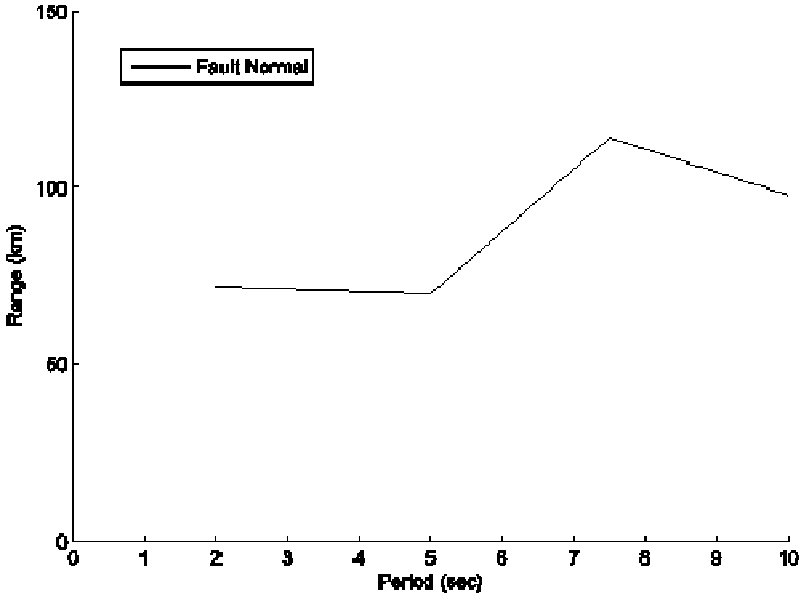


Figure 22: Ranges of semivariograms of residuals computed using the RandomHypoC simulations of the 1906 San Francisco earthquake.

The ranges of the semivariograms of residuals computed using the fault-normal, fault-parallel, north-south and east-west components of these simulated ground motion data sets are shown in Figure 23 and Figure 24. As seen in the previous sections, the ranges do not show a strong dependence on the ground-motion component used and, therefore, any further analysis in this section is based on the fault normal component of the ground motions.

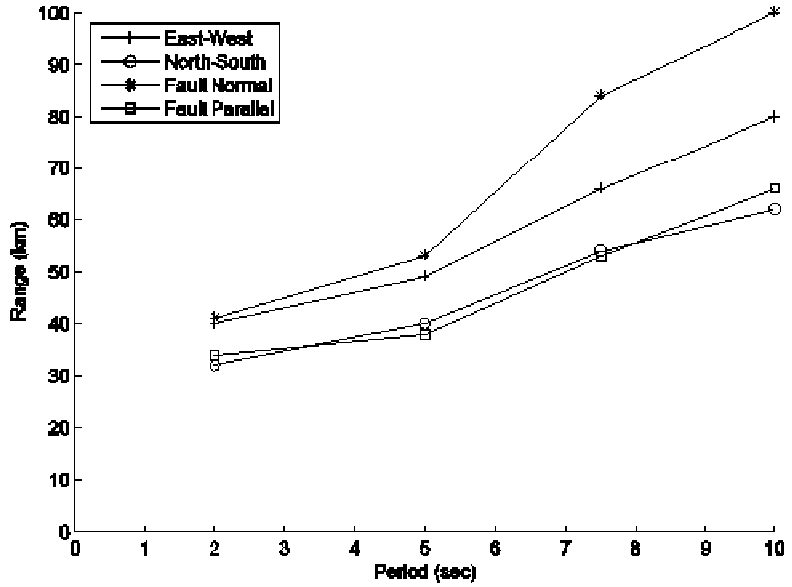


Figure 23: Ranges of semivariograms at different periods. Residuals are computed using RandomHypo06 ground motions at different orientations

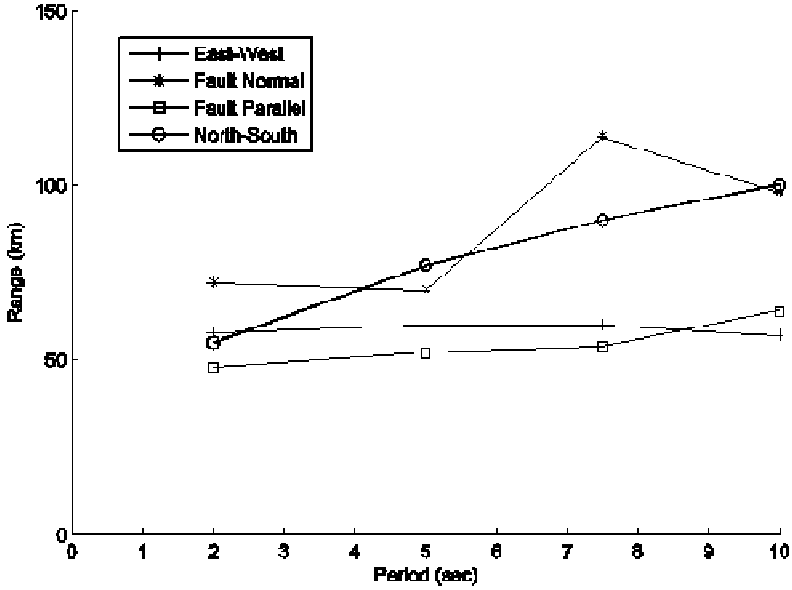


Figure 24: Ranges of semivariograms at different periods. Residuals are computed using RandomHypoC ground motions at different orientations

As before, semivariograms were computed using residuals at sites belonging to the following four distance domains: sites between 0-10km, 10-30km, 30-50km and farther than 50km from the rupture.
The corresponding estimates of the ranges along those for the semivariograms computed all the fault normal components regardless of their distance from the rupture are plotted for the two data sets in Figure 25 and Figure 26. In both cases, the assumption of second-order stationarity seems to be reasonably valid (barring a few outliers) except at sites that are within 10km of the rupture, whose spatial correlation is lower. This is consistent with observations from the other simulated ground-motion datasets.

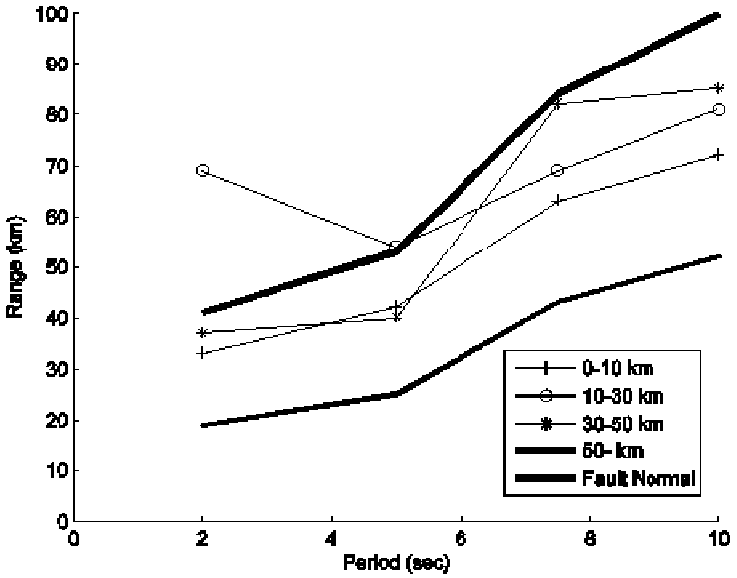


Figure 25: Ranges of semivariograms at different periods. Residuals are computed using RandomHypo06 ground motions at different orientations

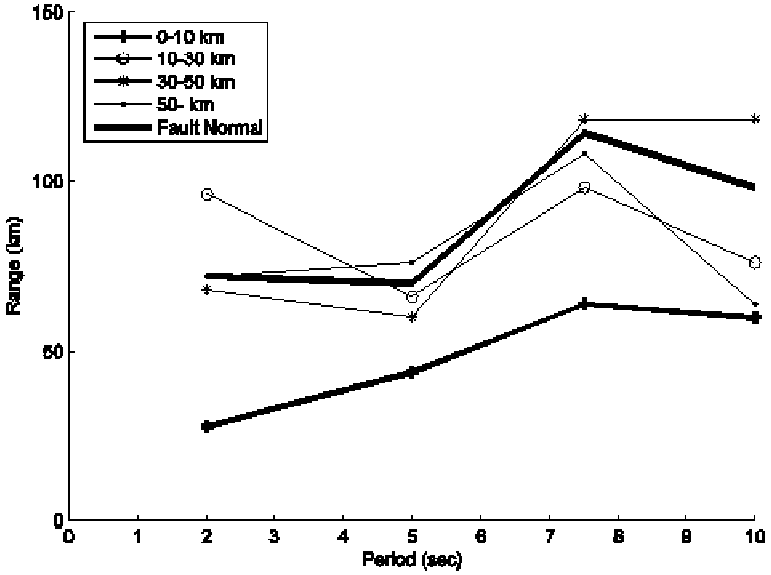


Figure 26: Ranges of semivariograms at different periods. Residuals are computed using RandomHypoC ground motions at different orientations

Finally, as before, the wavelet analysis procedure of Baker (2007) identified 2,649 and 2,698 pulses in the fault normal components of the ground motions from the RandomHypo06 and the RandomHypoC

datasets, respectively. Semivariograms were computed using these pulse-like ground motions, and the ranges of these semivariograms are shown in Figure 27 and Figure 28. Also these figures include the range estimates computed using near-fault (i.e., 0-10km) non-pulse-like ground motions and using all near-fault ground motions. Unlike in the previous data sets, the ranges computed based on pulse-like and non-pulse-like ground motions from the same spatial domain (near-fault) are drastically different. The source of these differences is not clear, but, as mentioned previously, only the average ranges calculated based on all the near-fault ground motions are, at this point in time, useful in practice. However, these near-fault ground motion ranges show a similar trend to those seen from other simulated datasets and this is comforting from the practical viewpoint of developing standardized correlation models.

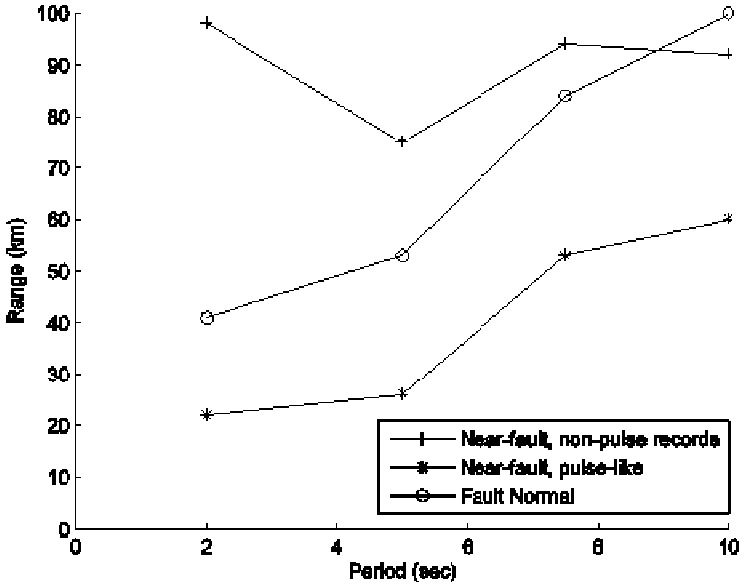


Figure 27: Ranges computed using pulse-like and non-pulse-like near fault RandomHypo06 ground motions

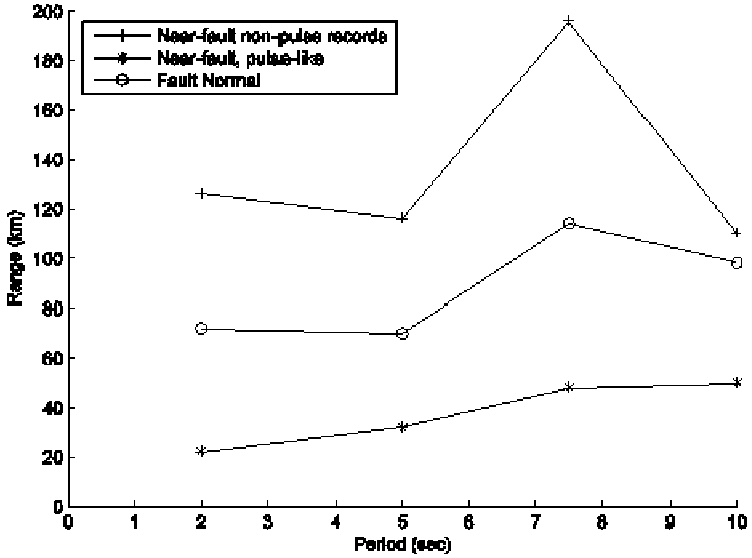


Figure 28: Ranges computed using pulse-like and non-pulse-like near fault RandomHypoC ground motions

3 VECTOR PROBABILISTIC SEISMIC HAZARD ANALYSIS FOR A SMALL PORTFOLIO OF STRUCTURES

In this section, we illustrate how site-to-site ground motion parameter correlation affects the joint performance of a small portfolio of key structures of interest to a stakeholder (i.e., a government agency, a corporation, or an insurance company). The extension to large portfolios follows in Section 4. Possible measures of performance include the expected losses for the entire portfolio, or the probability that some fraction of the structures in the portfolio will collapse, given the occurrence of a specific event or within some period of time. For example, a stakeholder may be interested in the probability that his/her primary and backup facilities both fail in the same event, or in the probability that total repair costs for the two facilities will exceed $10M (regardless of how the loss is partitioned between the facilities). Estimates of losses and collapse probabilities are essential when making decisions to mitigate future seismic risk.

The first step in performing these loss analyses is to assess the joint probability of occurrence of multiple ground motion parameters at the building sites. It has been shown that for most structures a meaningful link can be established between the intensity of a ground motion parameter (e.g., spectral acceleration, S_a, at the fundamental period of the structure, T), and structural response. Therefore, it is extremely useful to have a tool that computes the annual rate of occurrence of any pairs of ground motion parameter values at the building sites. If the sites are very far from each other (e.g., 100km) they can be treated separately as customarily done via Probabilistic Seismic Hazard Analysis (PSHA). However, sites that are relatively close to each other (e.g., 10km apart or less) during an earthquake are affected by ground motion whose peak intensity parameters are correlated (see Section 2 and companion report by Jayaram and Baker, 2008b).

When the number of ground motion parameters and the number of sites is small (say, one parameter per site for up to five sites) then the Vector-valued PSHA (VPSHA) tool (Bazzurro, 1998; Bazzurro and Cornell, 2002) originally developed for assessing joint hazard of multiple ground motion parameters at a single site is an appropriate tool for this objective. Only some relatively minor adjustments to the original methodology are needed, as explained below. A case with more parameters and more sites is, however, too computational intensive to be treated within the VPSHA framework with the current computer resources and it is better addressed using the Monte Carlo approach described in the next section. How VPSHA can be used for computing the joint hazard at multiple sites is discussed in the next subsection.

3.1 Methodology

The VPSHA approach implemented here is based on the tenable assumption of the joint lognormality of the correlated ground motion parameters (Baker and Jayaram, 2008) conditional on the characteristics of the causative event. . Recall that the same assumption has been exploited for essentially any ground motion prediction equation in existence and proven for the distribution of different peak parameters at the same site (Jayaram and Baker, 2008a). The joint Mean Rate Density, MRD (for definition and details, see Bazzurro and Cornell, 2002) or, alternatively and equivalently, the Mean Annual Rate (MAR) of occurrence of a pool of ground motion parameters can be computed with the knowledge of the following input:

- *Site-specific seismic hazard curves for the ground motion parameters at the sites*.
  The vector of ground motion parameter is denoted here as $\boldsymbol{S}$, where the bold character indicates that the quantity is a vector. In the application at hand, this vector could include, for example, the horizontal spectral acceleration at a given period at two sites or the spectral acceleration at two different periods at two sites. The periods could correspond to the first mode of vibration of each structure. These two hazard curves can be obtained with any standard PSHA code.

- *The pair-wise correlation matrix of all the ground motion parameters*).
  In the companion report Jayaram and Baker (2008b) have empirically derived the correlation structure for spectral accelerations with the same period at two sites regardless of the component orientation. The correlation matrix for spectral acceleration at different periods at two sites or of other ground motion parameters (e.g., Arias intensity) at two sites has not yet been developed but it can be derived using the same approach as in Jayaram and Baker (2008b).

- *The disaggregation results from scalar PSHA.*
  The joint distributions of all the basic variables, $\boldsymbol{X}$, used in the ground motion prediction equation of choice (i.e., M, R, and all the other variables – such as the style of faulting, the directivity parameters, the distance to the top of the co-seismic rupture, and dip angle, that are needed to compute the level of ground motion for every earthquake rupture) at each site conditional on the value of the selected ground motion parameter is a straightforward extension of the disaggregation results (e.g., based on M and R only) routinely available from standard scalar PSHA codes. The necessary modifications are conceptually simple and involve only disaggregation of the site hazard in terms of additional RVs beyond the magnitude, M, the source-to-site-distance, R, etc. as done in the past.

The input and the output of the VPSHA methodology will be illustrated for an example of two sites in San Francisco in the nest subsection.

Again, one of the distinct appealing qualities of this methodology is that it can be written as a standalone post-processor routine of a standard PSHA code. The accuracy of the results, however, could potentially be jeopardized by the selection of too wide bins during the discretization of the domain of each ground motion parameter (e.g., M and R).

The adopted VPSHA methodology described below has been successfully applied to compute the joint hazard of multiple parameters at the same site in Bazzurro *et al*. (2008). We have modified it here to be applicable to multiple sites.

To be concise but without losing generality, we present the details of the procedure for the case of three sites: Site 1, Site 2, and Site 3. The parameter of choice is a spectral acceleration at a given period (perhaps different at different sites). This approach, which requires some basic matrix algebra, is scalable to a larger number of sites and can include any other ground motion parameters (e.g., Peak Ground Velocity, Peak Ground Acceleration, Arias Intensity, etc.) for which spatial correlation structure and prediction equation are available. For simplicity, here we will also treat the RVs representing the ground motion parameter at each one of the three sites (i.e., S_{a1};S_{a2};S_{a3}) as discrete rather than continuous quantities.

Let us denote with $\boldsymbol{S}$=[S_{a1};S_{a2};S_{a3}] the vector of RVs for which we seek to obtain the joint hazard and with MAR[S_{a1};S_{a2};S_{a3}] = MAR$_{Sa1;Sa2;Sa3}$(a_1;a_2;a_3) the mean annual rate of three spectral acceleration quantities S_{a1}, S_{a2}, and S_{a3} in the neighborhood of any combination of three values a_1, a_2, and a_3, respectively. Strictly speaking, note that S_{a1};S_{a2};S_{a3} represent the natural logarithm of the spectral accelerations but the logarithm has been dropped here to avoid heavy notations. MAR[S_{a1};S_{a2};S_{a3}] could denote, for example, the MAR that spectral accelerations at the fundamental mode of each building assume values in the neighborhood of, say, 1.0g at each one of the three sites. In an application, these spectral acceleration values could be related to the onset of an important structural limit state (e.g., collapse) found using a statistical analysis of the response of the structure to many ground motion records.

Using simple probability theory and the theorem of total probability, one can write the following:

$$MAR[S_{a1};S_{a2};S_{a3}] = P[S_{a1} \mid S_{a2};S_{a3}] \cdot P[S_{a2} \mid S_{a3}] \cdot MAR[S_{a3}] \quad (6)$$

where

- $P[S_{a1} \mid S_{a2};S_{a3}] = \sum_{\boldsymbol{X}} P[S_{a1} \mid S_{a2};S_{a3};\boldsymbol{X}] \cdot P[\boldsymbol{X} \mid S_{a2};S_{a3}]$ (7)

  is the conditional distribution of S_{a1} given S_{a2} and S_{a3}. This term can be numerically computed by conditioning it on the pool of variables, $\boldsymbol{X}$, in PSHA that appear in the selected ground motion prediction equation and integrating over all possible values of $\boldsymbol{X}$. Given the assumption of joint lognormality of $\boldsymbol{S}$ mentioned before, for every possible value of $\boldsymbol{X}$ the quantity $P[S_{a1} \mid S_{a2};S_{a3};\boldsymbol{X}]$ can be computed simply with the knowledge of the variance-covariance matrix of S_{a1}, S_{a2}, and S_{a3} (e.g., Baker and Cornell, 2006) and the ground motion prediction equation of choice. More mathematical details are provided below. $P[\boldsymbol{X} \mid S_{a2};S_{a3}]$, which is the probability of $\boldsymbol{X}$ conditional on the values of S_{a2} and S_{a3}, can instead be obtained via disaggregation and Bayes theorem as follows:

  $$P[\boldsymbol{X} \mid S_{a2};S_{a3}] = \frac{P[\boldsymbol{X},S_{a2} \mid S_{a3}]}{\sum_{\boldsymbol{X}} P[S_{a2} \mid S_{a3};\boldsymbol{X}] \cdot P[\boldsymbol{X} \mid S_{a3}]} = \frac{P[S_{a2} \mid S_{a3};\boldsymbol{X}] \cdot P[\boldsymbol{X} \mid S_{a3}]}{P[S_{a2} \mid S_{a3}]} \quad (8)$$

  where

  - $P[\boldsymbol{X} \mid S_{a3}]$ can be derived using conventional scalar PSHA disaggregation.
  - $P[S_{a2} \mid S_{a3};\boldsymbol{X}]$, as for a similar term above, can be computed with only the knowledge of the variance-covariance matrix of S_{a1}, S_{a2}, and S_{a3}, and the attenuation relationship of choice.
- $P[S_{a2} \mid S_{a3}] = \sum_{\boldsymbol{X}} P[S_{a2} \mid S_{a3};\boldsymbol{X}] \cdot P[\boldsymbol{X} \mid S_{a3}]$ can be evaluated as explained above.
- $MAR[S_{a3}]$ is the absolute value of the differential of the conventional seismic hazard curve for the scalar quantity S_{a3} at Site 3.

In more detail, the above conditional terms (i.e., $P[S_{a1} \mid S_{a2};S_{a3};\boldsymbol{X}]$ and $P[S_{a2} \mid S_{a3};\boldsymbol{X}]$) can be obtained using the multivariate normal distribution theorem. More in general, if we call $\boldsymbol{S} = [S_{a1}, S_{a2}, ..., S_{an}]^T$ the vector of the natural logarithm of the random variables for which the joint hazard is sought, then $\boldsymbol{S}$ is jointly normally distributed with a mean vector ($\boldsymbol{\mu}$) and variance-covariance matrix ($\boldsymbol{\Sigma}$), i.e., in mathematical terms $\boldsymbol{S} =\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})$. By partitioning $\boldsymbol{S}$ into 2 vectors $\boldsymbol{S_1} = [S_{a1}, S_{a2}, ..., S_{ak}]^T$ and $\boldsymbol{S_2} = [S_{ak+1}, S_{ak+2}, ..., S_{an}]^T$ where $\boldsymbol{S_2}$ comprises the conditioning variables (in the example above $\boldsymbol{S_1}$=[S_{a1}] and $\boldsymbol{S_2}$=[S_{a2}, S_{a3}]), one can write the following:

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{S_1} \\ \hline \boldsymbol{S_2} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu_1} \\ \hline \boldsymbol{\mu_2} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} \right) \quad (9)$$

For jointly normal distribution, the conditional mean and conditional variance can be determined as

$$\boldsymbol{S_2} \mid \{\boldsymbol{S_1} = \boldsymbol{s_1}\} \sim N(\boldsymbol{\mu_{2|1}}, \boldsymbol{\Sigma_{2|1}}) \quad (10)$$

$$\boldsymbol{\mu}_{2|1} = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\left(\boldsymbol{s}_1 - \boldsymbol{\mu}_1\right); \quad \boldsymbol{\Sigma}_{2|1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} \tag{11}$$

The values of these two parameters derived as shown above define the conditional distributions mentioned above (i.e., $P[S_{a1} \mid S_{a2}; S_{a3}; \boldsymbol{X}]$ and $P\left[S_{a2} \mid S_{a3}; \boldsymbol{X}\right]$).

The joint hazard can also be disaggregated on the vector parameters $\boldsymbol{X}$ (for example, M and R). The same concept can also be applied to other ground motion prediction equations (e.g., Chiou and Youngs, 2008). Note that the conditional probability distribution $P\left[S_{a2} \mid S_{a1}; \boldsymbol{X}\right]$ in the VPSHA calculation (see Eqns. 7 and 8) should be also carried out with the same ground motion prediction equation used in scalar PSHA calculation.

Again using the assumption that $\varepsilon_{i,j}$ terms in Equation (1) are jointly normally distributed, the only extension of traditional PSHA needed when implementing this approach is a covariance matrix describing the spatial correlation of the parameters of interest. This covariance matrix can be computed from the correlation coefficients $\varepsilon_{i,j}$ at multiple sites. These correlation coefficients are precisely what have been developed in the companion report (Jayaram and Baker, 2008b). The study in Section 2 of this report showed that an isotropic correlation model appear to be satisfactory for most applications.

3.2 Application

To illustrate the effects of spatial correlation of ground motion intensities on the joint hazard, we consider two sites in San Francisco, which are 3.8km apart (Figure 29), where two low-rise ductile concrete buildings are located. We intend to estimate the joint hazard for 5%-damped elastic spectral acceleration at 0.3s, which is the fundamental period of these buildings. For ground motion computation purposes we have used the Abrahamson and Silva (1997) prediction equation for stiff soil sites.

Figure 29: Location of the two sites in San Francisco (Site 1: Long: -122.3925 Lat: 37.7905; Site 2: Long: -122.4330 Lat: 37.8022). The two sites are 3.8km apart.

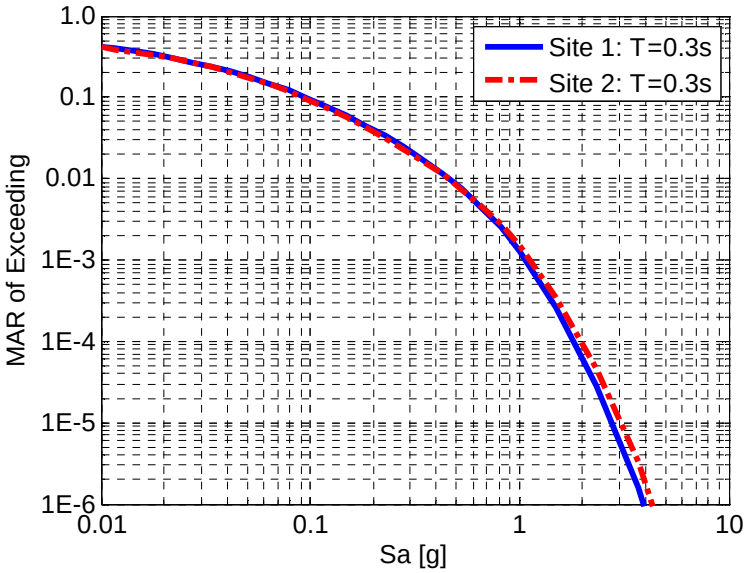
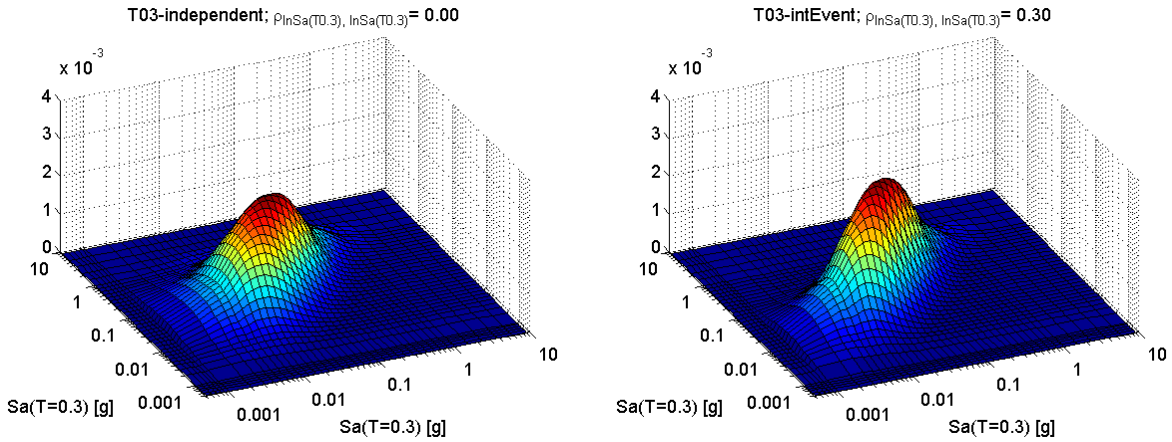


Figure 30: Seismic hazard curves for S_a(0.3s) at both sites

Figure 30 shows the conventional hazard curves for S_a(0.3s) at these two sites. These hazard curves, however, are expressed in terms of mean annual rates of exceedance rather than annual probability of exceedance. If the values of S_a(0.3s) generated at these two sites by the same earthquake were to be considered as independent RVs, as is conventionally done (see Equation 2), then the joint MRD of S_a(0.3s) would be the one shown in Figure 31a. In passing, note that even when the ground motion parameters at both sites are computed independently they are still (marginally) dependent due to the commonality of the same scenario events in the PSHA calculations (for details, see Bazzurro and Cornell, 2002). In other words the contours of the joint MRD in Figure 31a. are ellipses and not circles.

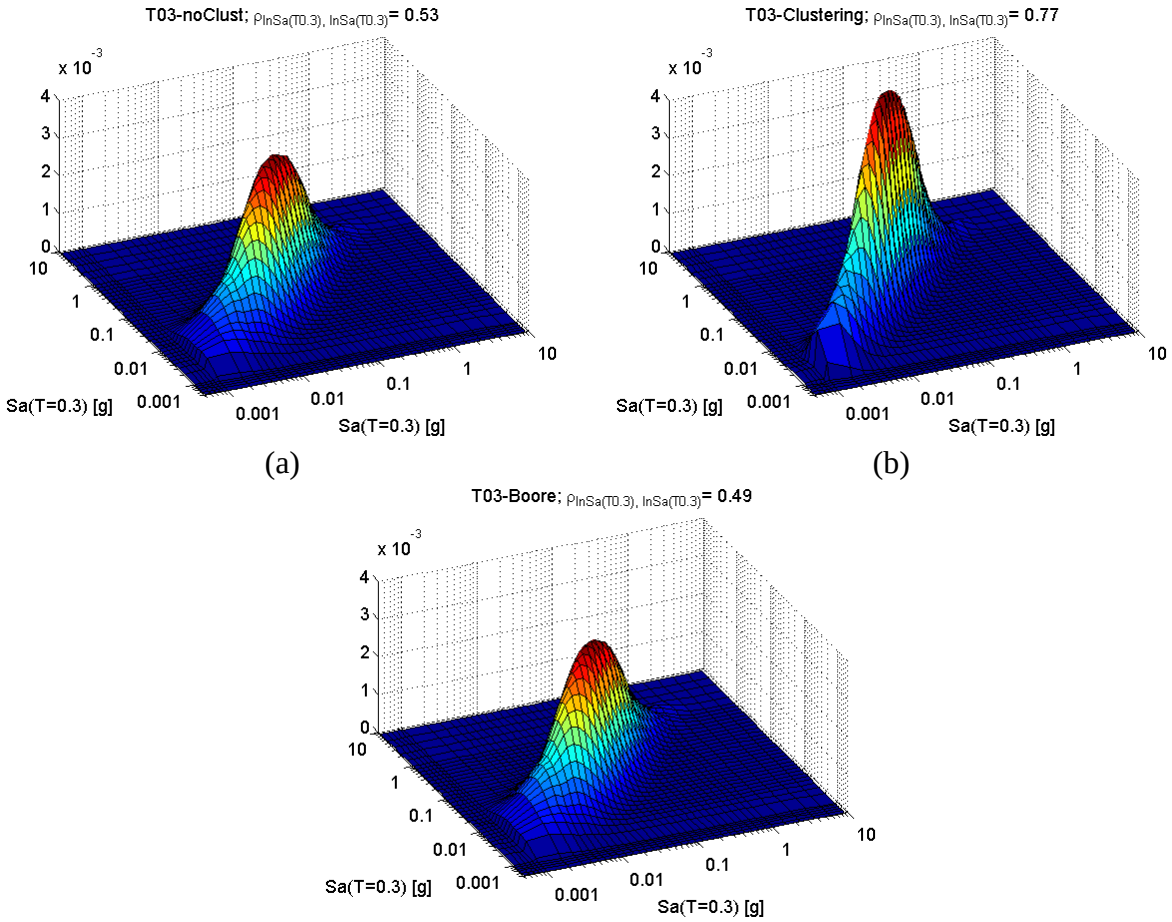
If, however, the joint hazard is computed according to Equation 1, the common inter-event error term introduces a mild spatial correlation between the ground motion parameters at both sites. This case is shown in Figure 31b and the correlation coefficient between (the logarithm of) S_a(0.3s) at Site 1 and (the logarithm of) S_a(0.3s) at Site 2 is 0.30. (Note that when the spatial correlation is modeled only by the inter event term the correlation coefficient, $\tilde{\rho}(h)$, which in this case is equal to $\rho(h)$**,** is identical between the ground motion parameters at any two sites at any distance from one another). Of course, the two marginal distributions of the joint distribution shown in these figures are consistent with the hazard curves in Figure 30. It is clear by comparing both panels of Figure 31 that accounting for the inter-event term makes the chance of observing high (or low) ground motions at both sites significantly higher than in the independent case.

When the spatial correlation is appropriately accounted for the ridge introduced by this additional source of correlation in the joint MRD of S_a(0.3s) at the two sites becomes considerably more evident (Figure 32). The joint MRD in Figure 32a was obtained by assuming that the geologic conditions vary significantly over the region while the Figure 32b assumes that considerable clusters of sites exist with similar soil conditions. Both correlation models were developed by Jayaram and Baker (2008b) in the companion report. In the latter case the empirical model at the two sites predicts higher correlation ($\tilde{\rho}(h)$ =0.77) for the correlation of S_a(0.3s) than in the former case ($\tilde{\rho}(h)$ =0.53). Finally Figure 32c shows for comparison purposes the joint MRD for S_a(0.3s) computed using the correlation model by Boore *et al.* (2003), which for this site-to-site distance produces a correlation coefficient of $\tilde{\rho}(h)$ =0.49 which is lower than those from the previous two cases.



(a) (b)

Figure 31: Joint MRD of S_a(0.3s) at the two sites shown in Figure 29 when the values of S_a(0.3s) generated by the same event at the two sites are considered to be independent (left) or correlated via the inter-event term .



(a) (b)

(c)

Figure 32: Joint MRD of S_a(0.3s) at the two sites shown in Figure 29 when the values of S_a(0.3s) generated by the same event at the two sites are considered to be spatially correlated according to the model by a) Jayaram and Baker, (2008b) with no soil clustering, b) Jayaram and Baker (2008b) with soil clustering, and c) Boore *et al.* (2003) in addition to the correlation contributed by the inter-event term.

The effects of including or neglecting the ground motion spatial correlation when estimating the likelihood of future losses for these two buildings can be detected by inspecting the mean rate of exceedance loss curves showed in Figure 33. For this illustrative example we assumed that each site hosts a low-rise reinforced concrete frame building of modern construction worth 10M USD. The differences among these curves produced by the four of the five modeling approaches considered in Figure 32 and Figure 33 are more evident Figure 34. As expected, considering the ground motion correlation increases the likelihood of observing very low and very high losses. Hence, the MRE loss curves that consider correlation cross the independent case one at some intermediate loss level. It is emphasized here that neglecting the ground motion correlation overestimates the likelihood of all losses in exceedance of about 300,000 USD and underestimates the likelihood of observing losses that are larger than that amount. An analysis of the results from another perspective (Figure 34b) shows that all the losses with mean return period longer than about 12 years are underestimated if the ground motion correlation is not considered, as routinely done. Finally note in Table 1 that, at least in this case, the estimates of the average annual losses are fairly stable regardless of whether the ground motion spatial correlation is modeled or not. This is due to the compensating over- and under-estimation errors introduced by neglecting the spatial correlation. The lack of effect on the AAL is, however, not generally valid for other portfolios.

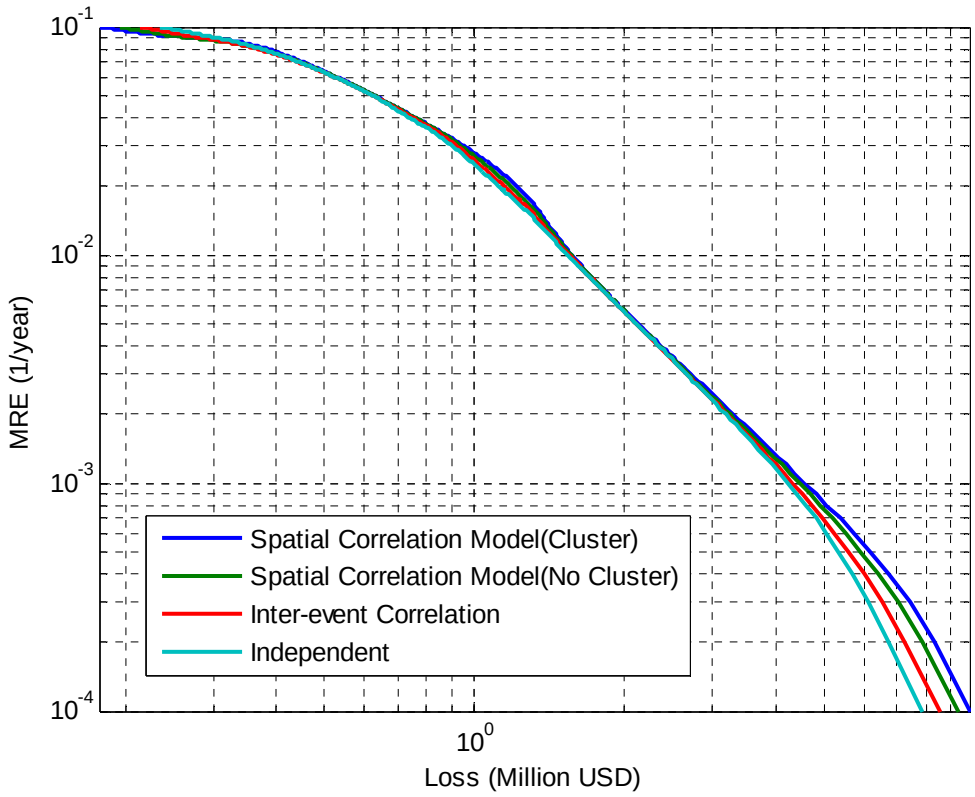


Figure 33: Mean rate of exceedance curves for the portfolio of two buildings.

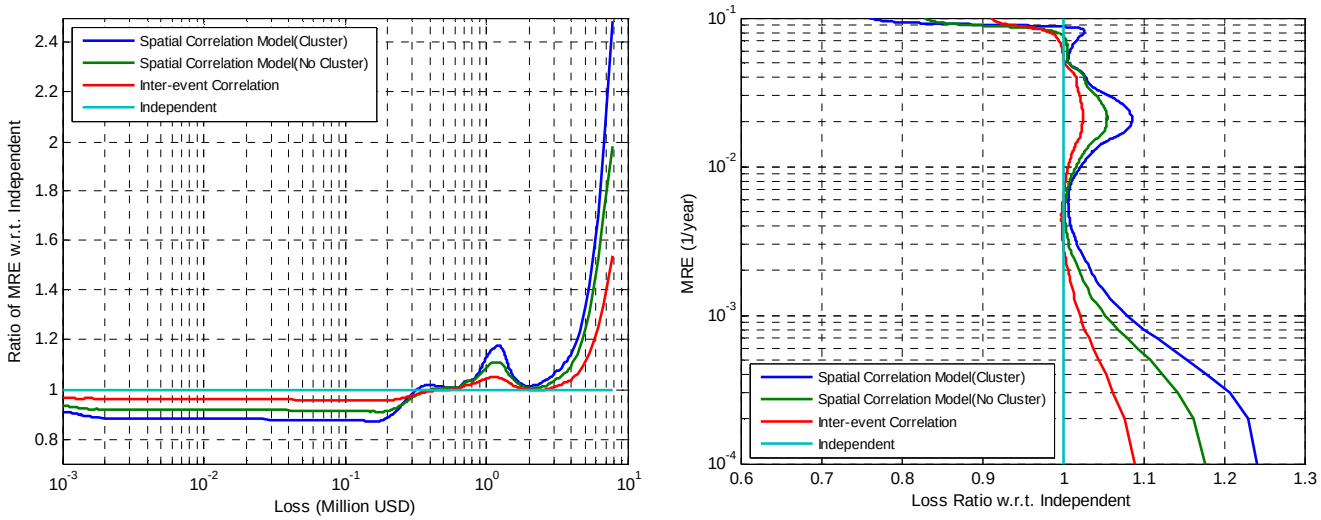


Figure 34: Ratio of loss mean exceedance curves with respect to the independent ground motion case plotted versus loss (left) and versus mean rate of exceedance (right).

Independent	Inter-event Correlation	Spatial Correlation w/o cluster	Spatial Correlation w/ cluster
$88,722	$88,160	$88,163	$88,465

Table 1: Average Annual Loss (AAL) values for the portfolio of buildings estimated using four representations of the spatial correlation of ground motion intensities: a) independent ground motion and the two sites; b) correlated ground motion via inter-event error term only; c) spatially correlated ground motion without soil clustering effects; and d) with clustering effects.

4 EFFECTS OF SPATIAL CORRELATION OF GROUND MOTION ON LOSSES TO LARGE PORTFOLIOS OF STRUCTURES

To study the effects of spatial correlation of ground motion intensities on portfolio losses we considered two hypothetical portfolios:

- Large Portfolio of 41,400 buildings located in the six counties around the San Francisco Bay Area, with 200 properties in each postal code (Figure 35).
- Small Portfolio of 200 buildings concentrated in the 94111 postal code downtown San Francisco (Figure 36). In this portfolio all buildings are within 2km from one another.

In both cases we assumed that these buildings are low-rise modern, ductile low rise concrete moment-resisting frame structures with a replacement value of $100,000 per property. To simplify ground motion calculation we assumed NEHRP B-C soil type conditions at all locations.

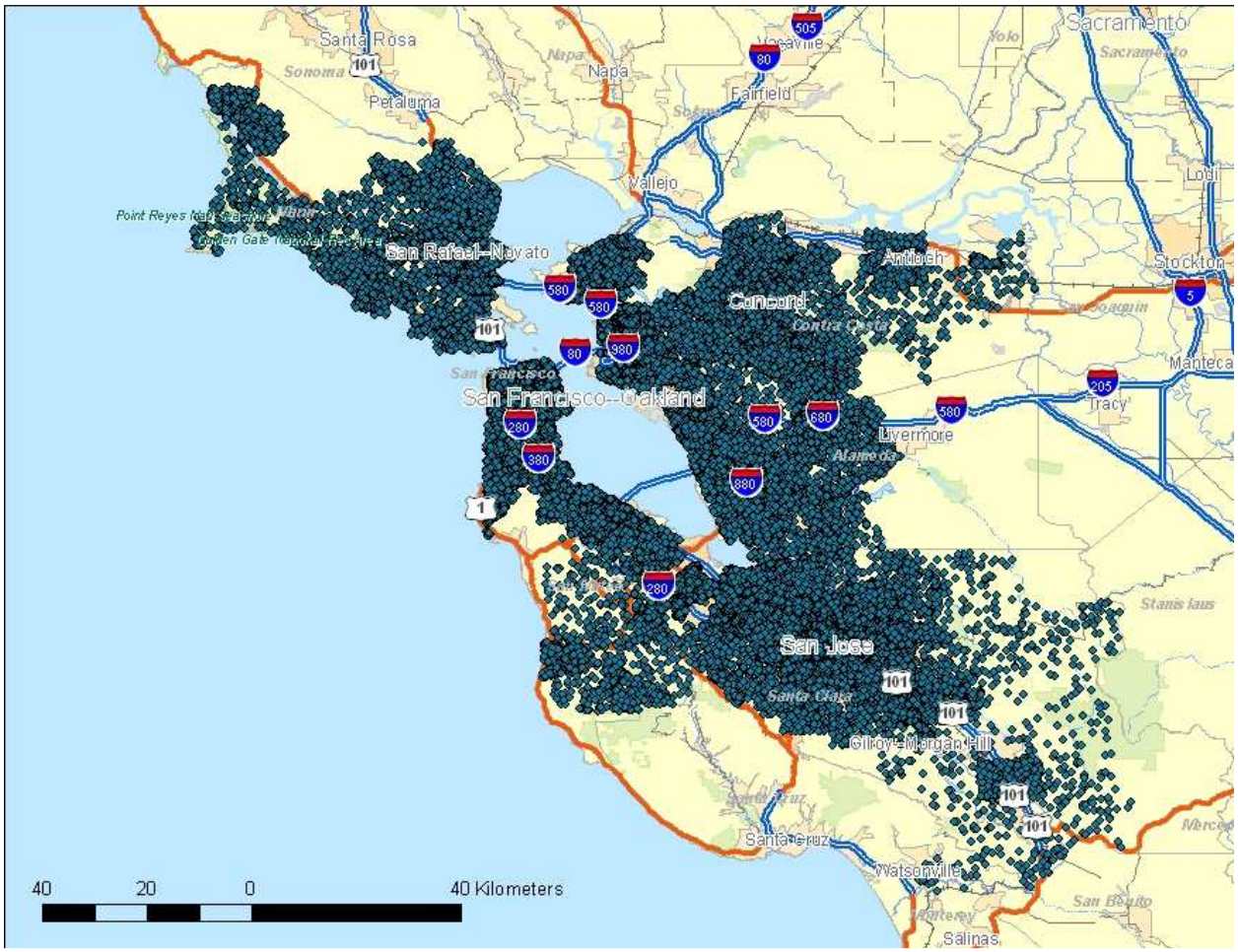


Figure 35: Hypothetical large portfolio of 41,400 buildings.

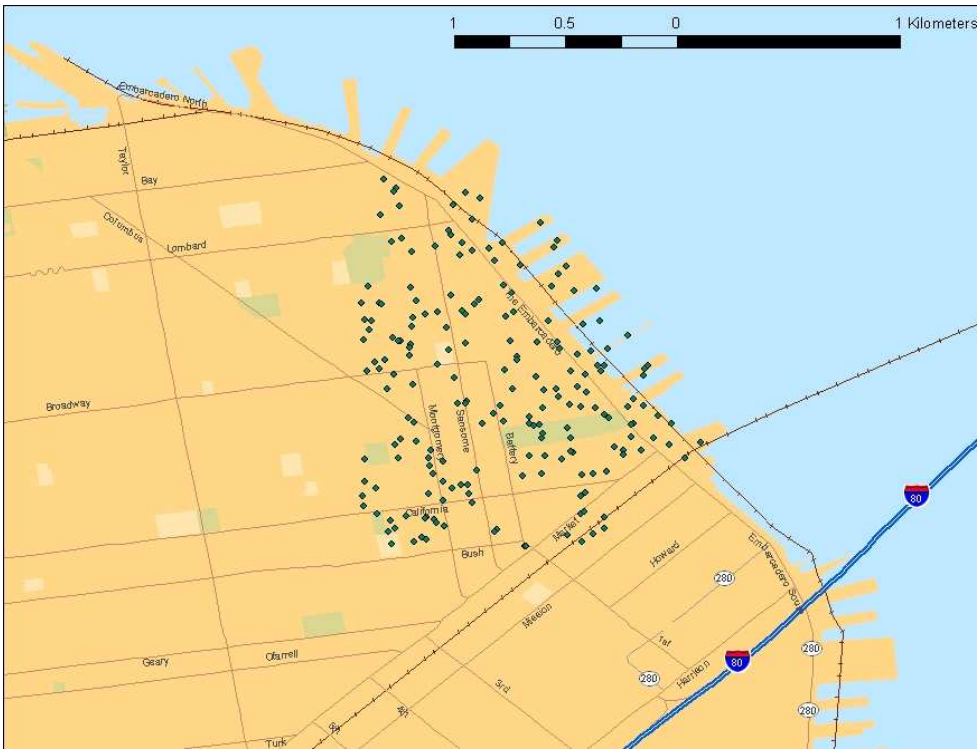


Figure 36: Hypothetical small portfolio of 200 buildings.

As mentioned in the previous section, for portfolios with a large number of structures the convolution of the joint hazard computed via vector PSHA with the damage functions of each structure is not feasible. In this case we selected a Monte Carlo simulation approach that uses a catalog of earthquakes representative of 10,000 realizations of next year seismic activity. The location and the magnitude of the earthquakes in the catalog were selected according to their occurrence rates and magnitude distributions. The steps of the Monte Carlo approach are listed below:

Step 1. For each earthquake rupture we simulate the ground motion at each building location using the Abrahamson and Silva (1997) equation according to one of the four modeling schemes:

(a) Independent ground motion at each site (see Equation 2)
(b) Constant correlation of ground motion intensity at each site. In this case the correlation is introduced by the constant inter-event error term in Equation 1.
(c) Site-to-site correlation of ground motion intensities assuming the model by Jayaram and Baker (2008b) without soil clustering effects.
(d) Site-to-site correlation of ground motion intensities assuming the model by Jayaram and Baker (2008b) with soil clustering effects. For the problem at hand, this is the most appropriate modeling approach and, therefore, the results from it are considered as benchmark for the results from the other three.

The result of this first step is a ground motion random field that is independent in Case a and correlated with increasing level of correlation in Cases b, c, and d.

Step 2. Given the ground motion at each site the building losses are simulated from the damage function for the type of concrete buildings considered. The damage function is simply a relationship that for a given level of ground motion intensity (here S_a(0.3s)) provides the expected damage ratio (i.e., the repair cost divided by the replacement value of the building) and associated variability. No building-to-building loss statistical correlation is considered, only the functional correlation stemming from the use of the same damage functions for all buildings is.

Step 3. The losses at all sites are then summed to compute the losses for the considered event.

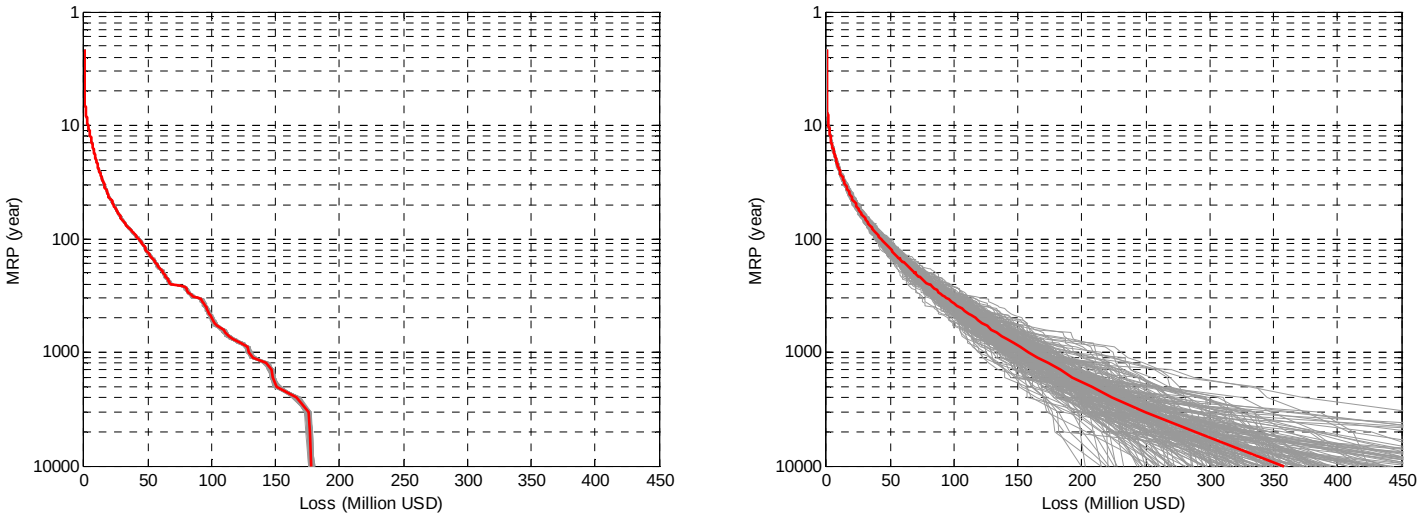
Step 4. Steps 1-3 are repeated for all the earthquakes in the catalog, the losses for all the events are then ranked from the highest to the lowest to compute the loss Mean Rate of Exceedance (MRE) curve for the portfolio for this realization. The highest loss is assigned a Mean Return Period (MRP) of 10,000 years (or, equivalently, a mean rate of exceedance of 1×10^{-4}), the second highest a MRP of 10,000/2=5,000 years (or a MRE of 2×10^{-4}), and so on. This process produces one of the curves presented in light gray in Figure 37.

Step 5. For each of the four ground motion modeling schemes, the simulation process is repeated 500 times to produce the 500 loss MRE curves shown in Figure 37a-d. The losses for each MRE (or MRP) are then averaged to produce the mean curve shown in red in Figure 37.

A comparison of the loss mean rate of exceedance curves is facilitated in Figure 38, which shows the ratio of the loss mean rate curves for all four cases to the loss mean rate curve of the benchmark case (Case d, spatial correlation with soil clustering effects). It is evident that assuming independent ground motion as routinely done in most such applications provides a poor representation of the risk. For this portfolio, the frequent losses are overestimated by as much as 200% and the large losses are underestimated up to 60%. The inaccuracy in the loss estimates is confined to ±20% when the spatial correlation is modeled via the inter-event term (Case b) or via the spatial correlation scheme with no soil clustering effects. Note that the levels of accuracy achieved by Cases a-c when compared with the benchmark results are portfolio-dependent, it tends to increase for larger portfolios and to decrease for portfolios of buildings, especially if they are tightly clustered. Note that the mean exceedance rate curves for Cases a-c cross the benchmark one at a MRP of about 50years in this case. This is not a general result since the crossing point varies from portfolio to portfolio and tends to occur at longer MRP values for smaller portfolios, as will be shown later. Similarly, in this case the approach with constant correlation (Case b) and the approach with spatial correlation without soil clustering effects (which provides correlation coefficients that decay very rapidly to the plateau of constant correlation modeled in Case b) provide very similar MRE curves. This finding holds with portfolios with a large number of buildings located in an extended area and it is not a general finding. Portfolios with fewer properties more clustered together will exhibit MRE curves for Cases b and c that significantly differ, as will be shown below.

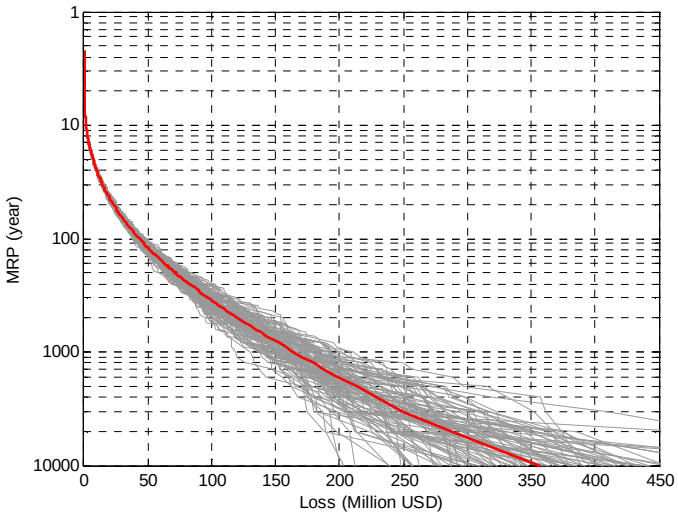
Finally, an inspection of Table 2 shows that the impact on the Average Annual Loss (AAL) of different ground motion spatial correlation modeling techniques is rather limited for this portfolio. This is because neglecting or reducing correlation lowers both tails of the loss distribution with counteracting effects on the AAL estimates. It is emphasized, however, that although this consideration about counteracting effects is general the limited difference between

AAL estimates from different modeling techniques is not. AAL estimates may be materially different for portfolios with different characteristics, especially in cases with a few buildings clustered together.

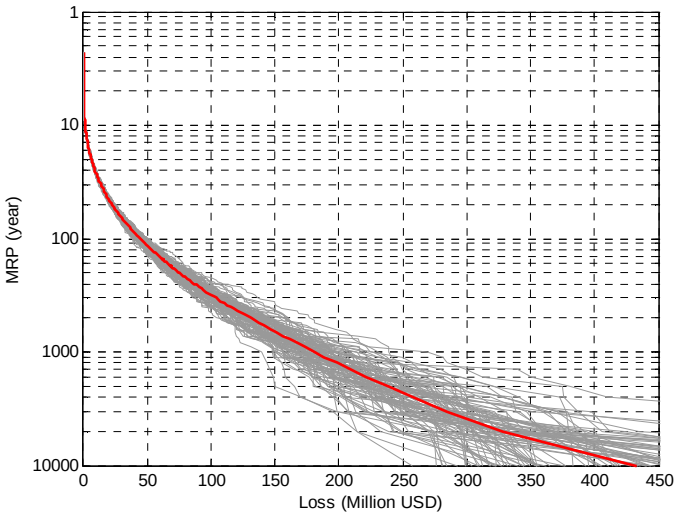


(a)

(b)



(c)



(d)

Figure 37 Loss exceedance rate curves for the large portfolio assuming (a) independent ground motion intensities at each site; (b) constant spatial correlation via the inter-event error term only; (c) spatially correlated ground motion intensities at each site without soil clustering effects and (d) spatially correlated ground motion intensities at each site with soil clustering effects. The red line is the mean exceedance curve.

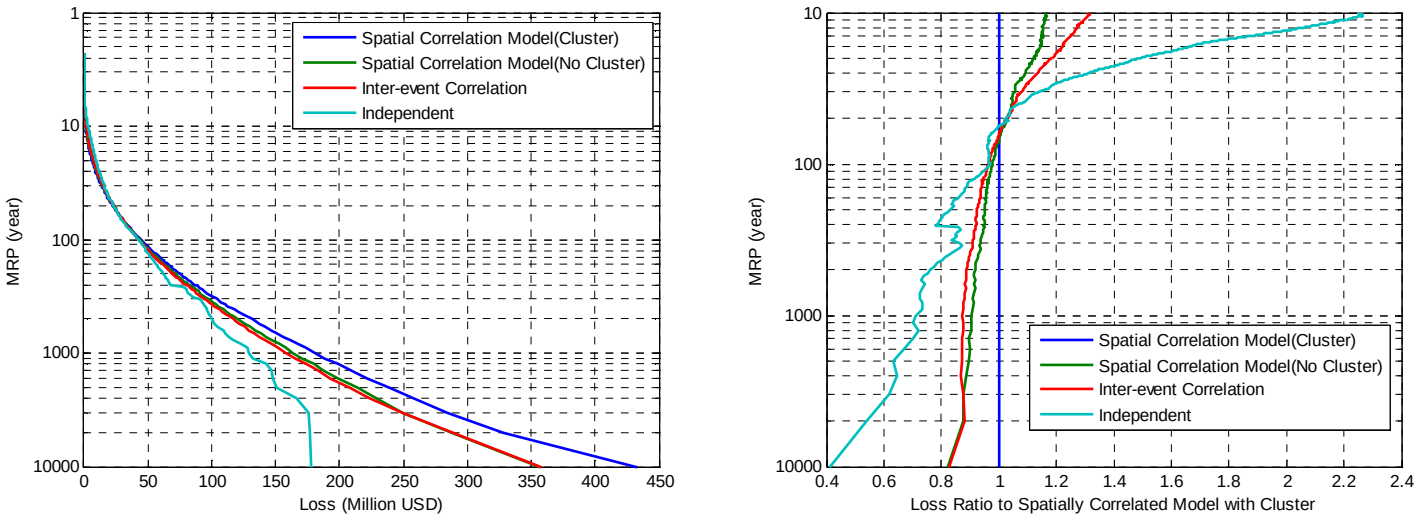


Figure 38: (a) Comparison among the mean loss exceedance probability curves obtained for the large portfolio for the four cases. (b) Ratio of mean loss exceedance probability curves to the one obtained with spatial correlation with clustering effects.

Table 2: Mean Loss for specific mean return periods and average annual loss for different ground motion correlation modeling techniques. (Million USD)

MRP (year)	Spatial Correlation Model (Cluster)	Spatial Correlation Model (No Cluster)	Inter-event Correlation	Independent
10,000	$ 433.00	$ 356.77	$ 358.14	$ 177.89
5,000	$ 328.01	$ 287.31	$ 288.73	$ 176.65
2,500	$ 257.74	$ 228.78	$ 223.78	$ 166.27
2,000	$ 238.85	$ 213.73	$ 207.93	$ 150.74
1,000	$ 181.33	$ 163.80	$ 158.50	$ 128.98
500	$ 130.91	$ 120.27	$ 116.27	$ 99.89
250	$ 86.70	$ 82.27	$ 79.84	$ 67.95
200	$ 75.43	$ 71.57	$ 69.55	$ 62.54
100	$ 44.18	$ 43.29	$ 42.58	$ 42.83
50	$ 22.47	$ 23.15	$ 23.14	$ 23.00
AAL	$ 1.86	$ 1.85	$ 1.87	$ 1.92

To support some of the statements above regarding the portfolio-dependency of some of the findings discussed for large portfolio of properties, we consider the smaller portfolio displayed in Figure 36. The loss exceedance rate curves for this portfolio are presented in Figure 39 while the ratios of the mean exceedance rate curves are shown in Figure 40. When the portfolio is comprised of fewer buildings concentrated in a smaller geographical area, then the likelihood of observing much larger (or smaller) ground motion than expected for the given event at all the building sites are much higher. Such instances are adequately modeled by the benchmark Case d and by Case c, which both captures spatial correlation, and increasingly less appropriately modeled by Cases b and a. With this consideration in mind, it appears intuitive that the MRE

curves for Cases a and b differ significantly from those obtained in the benchmark Case d. In this case the results for Case c are very similar to the target results because all the buildings in this portfolio are very closely spaced and the two correlation models do not significantly differ at very short distances. If the geographical area had been, say, a county rather a postal code, the buildings would have been, on average, farther apart from one another and in that case Case c would have generated less accurate MRE curves. Finally, by inspecting Table 3 it is clear that in this case as well the compensating effects of the inaccurate estimates of the likelihood of observing very large and very small losses provided by Cases a and b cause AAL estimates that are fairly precise. The AAL estimates from the independent ground motion case and the constant spatial correlation case differ more markedly from the target results of Case d for portfolios of only a few buildings clustered together.

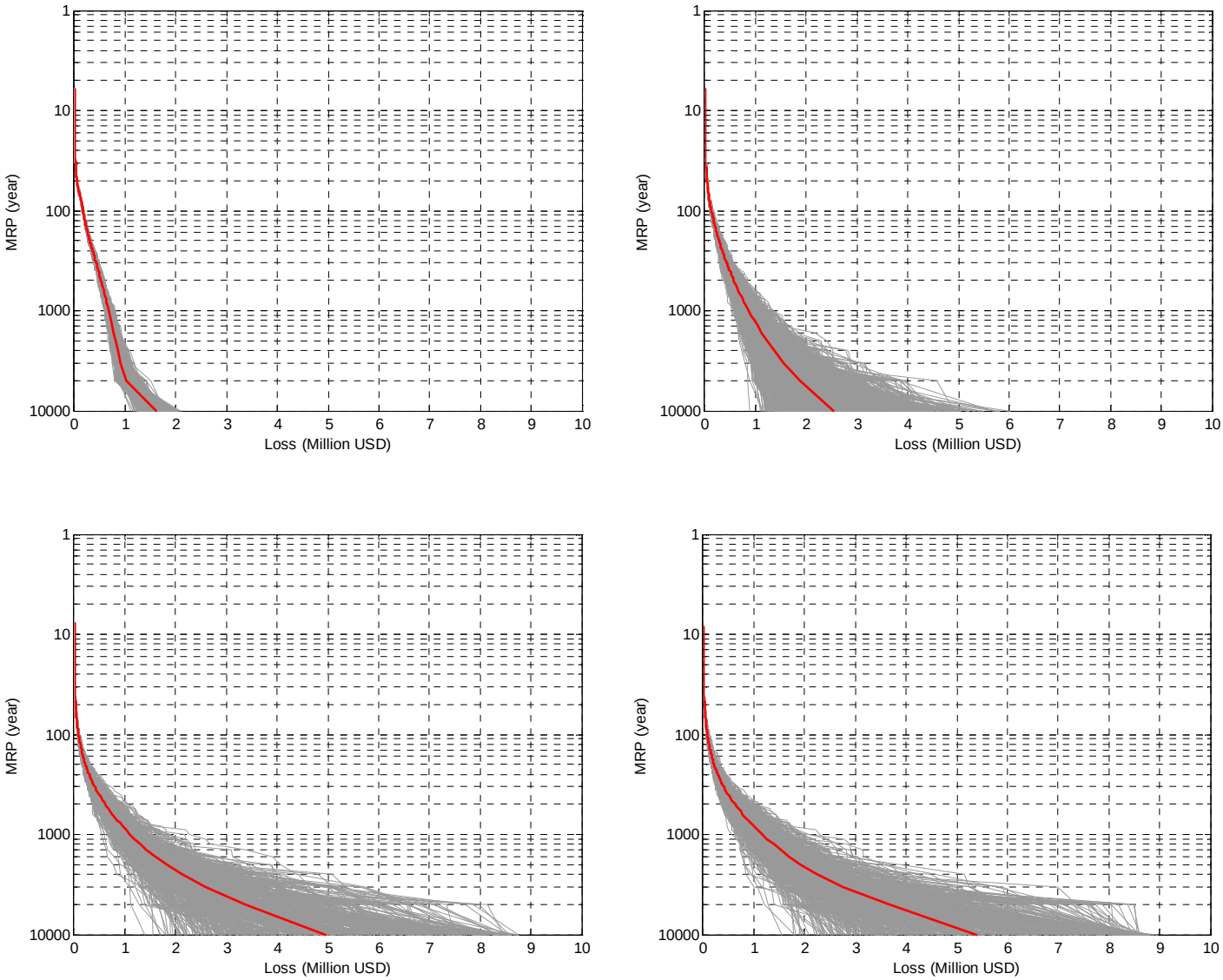


Figure 39 Loss exceedance probability curves for the small portfolio assuming (a) independent ground motion intensities at each site; (b) constant spatial correlation via the inter-event error term only; (c) spatially correlated ground motion intensities at each site without soil clustering effects and (d) spatially correlated ground motion intensities at each site with soil clustering effects. The red line is the mean exceedance curve.

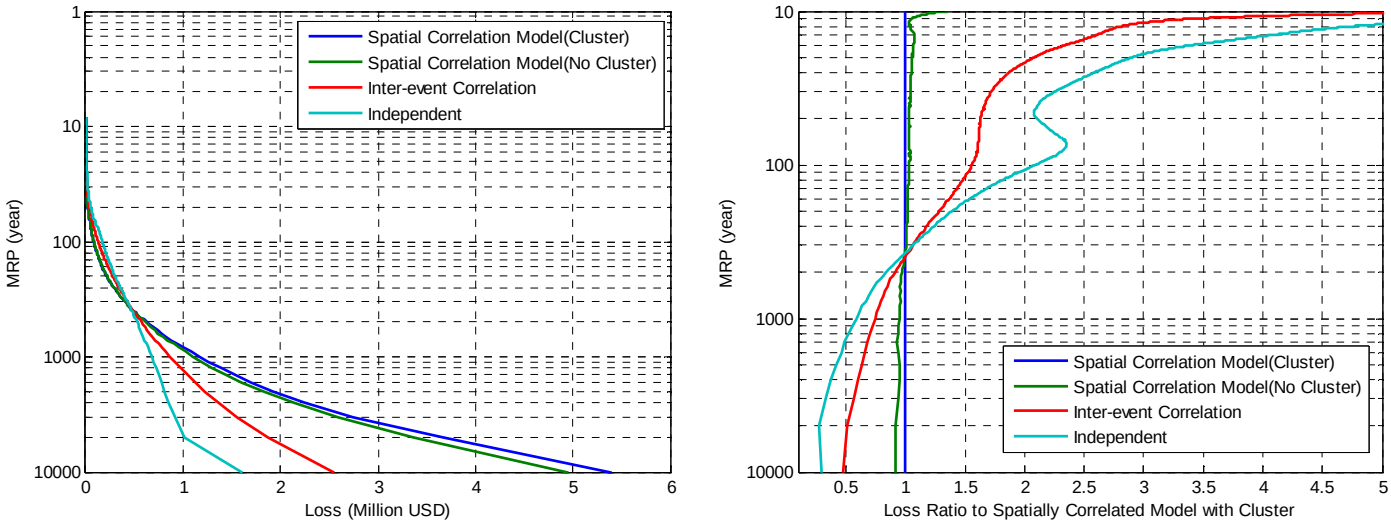


Figure 40: (a) Comparison among the mean loss exceedance probability curves obtained for the small portfolio for the four cases. (b) Ratio of mean loss exceedance probability curves to the one obtained with spatial correlation with clustering effects.

Table 3: Table 4 Mean Loss for specific mean return period and average annual loss for different simulation methodologies. (Thousand USD)

MRP (year)	Spatial Correlation Model (Cluster)	Spatial Correlation Model (No Cluster)	Inter-event Correlation	Independent
10,000	$ 5,395.49	$ 4,957.59	$ 2,550.22	$ 1,613.80
5,000	$ 3,664.73	$ 3,368.57	$ 1,868.93	$ 1,021.10
2,500	$ 2,241.57	$ 2,130.28	$ 1,361.94	$ 855.30
2,000	$ 1,922.82	$ 1,826.88	$ 1,227.86	$ 813.73
1,000	$ 1,161.59	$ 1,105.42	$ 866.60	$ 681.38
500	$ 626.72	$ 610.61	$ 572.69	$ 528.05
250	$ 288.44	$ 291.45	$ 340.90	$ 358.61
200	$ 217.83	$ 222.00	$ 279.76	$ 302.36
100	$ 82.01	$ 84.74	$ 127.80	$ 171.07
50	$ 24.69	$ 25.43	$ 40.30	$ 51.89
AAL	$ 5.64	$ 5.51	$ 5.53	$ 5.73

5 CONCLUSIONS AND RECOMMENDATIONS

This study has addressed the issue of spatial correlation of ground motion intensities generated by a single earthquake and has shown how the spatial correlation can be incorporated into assessing both the joint seismic hazard at multiple sites and also portfolio losses. The results of this study have been presented in this report and in the companion report by Jack Baker and Nirmal Jayaram of Stanford University.

The study discussed here consists of three main parts. The first part deals with ground motion correlation in synthetic datasets and complements the study performed at Stanford University using real ground motion recordings. Synthetic data were used here to investigate statistical properties of the spatial correlation function that could only be addressed with a wealth of spatially distributed ground motions generated by the same earthquake that is simply unavailable from historical events. This preliminary study has shown that an isotropic model for the spatial correlation is, in general, supported by the data. Moreover, the spatial correlation between intensities at two sites has been found to be dependent on the site-to-site distance but, in most cases, independent of where the two sites were located with respect to the fault rupture. The only exception is for near-fault sites within 10km to 20km from the rupture whose spatial correlation of intensity measures has been observed to be lower than that of sites at the same distance but farther from the rupture. Note that the statements above could gain strength from more corroborating evidence that may come from analyzing additional synthetic datasets generated by a larger pool of researchers for a larger set of earthquakes. Although beyond the scope of this study, note that the results of the correlation study of synthetic data can be used as a guidance in assessing whether the ground motion simulation techniques adopted to produce these time histories have, statistically speaking, the same spatial signature as those generated by real earthquakes.

The second part of this study showed how the spatial correlation can be incorporated into the computation of the joint seismic hazard at multiple sites using a direct numerical integration approach. This has been achieved by modifying the Vector Probabilistic Seismic Hazard Analysis (VPSHA) tool that was originally developed for computing the joint hazard of different ground motion parameters at the same site. The application included here also shows how the joint hazard computation can be included in the loss estimation of a portfolio of two sites located in San Francisco 4km from each other. We also show the errors introduced in the hazard and loss estimates when the spatial correlation is either neglected or coarsely modeled.

With the current computer resources, the approach based on VPSHA and, therefore, on numerical integration can produced accurate results for portfolios of about five sites. Beyond this threshold the accuracy in the estimates of extreme joint ground motion hazard and losses may go down because of the need of using coarser bins during the numerical integration. To successfully estimating earthquake losses for large portfolio of properties in the presence of spatially correlated ground motion intensities we adopted a Monte Carlo simulation approach that we apply here to two large portfolios of structures in the San Francisco Bay Area. Again, the effects of neglecting or crudely modeling spatial correlation are outlined.

The results show that, as expected, an appropriate modeling of spatial correlation of ground motion is essential when a portfolio of structures is clustered in a rather small region (e.g., a postal code) of and becomes less important when the portfolio is spread out in much larger geographical area (e.g., the six counties in the San Francisco Bay Area). Note that the latter statement above holds true only when one is interested in the loss statistics for the entire (large) portfolio. If one drills down on the results from the analyses to extract the loss statistics for a smaller area of the portfolio (e.g., the subset of structures within a city boundary or a postal code), then those loss estimates losses will not be accurate unless ground motion spatial correlation is adequately modeled.

6 ACKNOWLEDGEMENTS

We are very thankful to Drs. Brad Aagaard and Robert Graves for providing the simulated ground motion data sets that form the basis of the study discussed in Section 2. 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.

7 REFERENCES

B. T. Aagaard and T. H. Heaton (2004). Near-source ground motions from simulations of sustained intersonic and supersonic fault ruptures. *Bulletin of the Seismological Society of America*, 94(6):2064–2078.

B. T. Aagaard, T. M. Brocher, D. Dolenc, D. Dreger, R. W. Graves, S. Harmsen, S. Hartzell, S. Larsen, K. McCandless, S. Nilsson, N. A. Petersson, A. Rodgers, B. Sjogreen, and M. L. Zoback (2008a). Ground-motion modeling of the 1906 San Francisco earthquake, part II: Ground-motion estimates for the 1906 earthquake and scenario events. *Bulletin of the Seismological Society of America*, 98(2):1012–1046.

B. T. Aagaard, T. M. Brocher, D. Dolenc, D. Dreger, R. W. Graves, S. Harmsen, S. Hartzell, S. Larsen, and M. L. Zoback (2008b). 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.

Abrahamson, N. and W. Silva (1997). "Equations for estimating horizontal response spectra and peak acceleration from West North American Earthquakes: A summary of recent work", *Seismological Research Letters*, **68**, 94-127.

Baker, J.W. and C.A. Cornell (2006). "Correlation of Response Spectral Values for Multicomponent Ground Motions", *Bulletin of Seismological Society of America*, February 2006; Vol. 96; No. 1; pp. 215-227; DOI: 10.1785/0120050060.

W. Baker and Jayaram, N. (2008). Correlation of spectral acceleration values from NGA ground motion models, *Earthquake Spectra*, 24 (1), 299-317

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.

Bazzurro, P. (1998). ``Probabilistic Seismic Demand Analysis". Ph.D. Dissertation, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, CA, August.

Bazzurro, P., and C.A. Cornell (2002). "Vector-valued Probabilistic Seismic Hazard Analysis (VPSHA)". *Proceedings of 7th U.S. National Conference on Earthquake Engineering*, Boston, MA, July 21-25, Paper No. 61.

Bazzurro, P., Tothong, P., and J. Park (2008). "Efficient Approach to Vector-Valued Probabilistic Seismic Hazard Analysis of Multiple Correlated Ground Motion Parameters", Southern California Earthquake Center (SCEC) Report, Award No. 123956, January.

G. C. Beroza (1991). Near-source modeling of the Loma Prieta earthquake: evidence for heterogeneous slip and implications for earthquake hazard. *Bulletin of the Seismological Society of America*, 81(5): 1603–1621.

J. Boatwright and H. Bundock (2008). Modified Mercalli intensity maps for the 1906 San Francisco earthquake plotted in ShakeMap format. Technical report, U. S. Geological Survey, Open-File Report 2005-1135.

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.

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.

Chiou, B., and Youngs, R., (2008). Chiou-Youngs NGA ground-motion relations for the geometric mean horizontal component of peak and spectral ground-motion parameters, *EarthquakeSpectra* **24**, 173–215.

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.

K. Goda and H. P. Hong. Spatial correlation of peak ground motions and response spectra. *Bulletin of the Seismological Society of America*, 98(1):354–365, 2008.

P. Goovaerts (1997). Geostatistics for Natural Resources Evaluation. Oxford University Press, Oxford, New York.

N. Jayaram and J. W. Baker (2008a). Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of the Seismological Society of America*, 98(5), 2231-2243.

N. Jayaram and J. W. Baker (2008b). Correlation model for spatially-distributed ground-motion intensities. *Earthquake engineering and structural dynamics* (in review).

Jeon S.-S. and O'Rourke T.D., (2005). Northridge earthquake effects on pipelines and residential buildings, *Bulletin of the Seismological Society of America*, **95** (1), 294-318.

Kawakami, H., and Mogi, H., (2003). "Analyzing Spatial Intraevent Variability of Peak Ground Accelerations as a Function of Separation Distance." *Bulletin of the Seismological Society of America*, 93(3), 1079-1090.

Luco, L., Manuel, L., Baldava, S., and P. Bazzurro (2005a). "Correlation of Damage of Steel Moment-Resisting Frames to a Vector-valued Ground Motion Parameter Set that includes Energy Demands", Report prepared for U.S. Geological Survey (USGS) Award No. 03HQGR0057, February.

Luco, L., Manuel, L., Baldava, S., and P. Bazzurro (2005b). “Correlation of Damage of Steel Moment-Resisting Frames to a Vector-Valued Set of Ground Motion Parameters”, Accepted for publications in *Proceedings of the 9th ICOSSAR05*, June 19-22, Rome, Italy.

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.

Somerville, P.G., Smith, N.F., Graves, R.W., Abrahamson, N.A. (1997), "Modification of empirical strong ground motion attenuation relations to include the amplitude and duration effect of rupture directivity," *Seismological Research Letters*, 68(1): 199-222.

S. Song, G. C. Beroza, and P. Segall (2008). A unified source model for the 1906 San Francisco earthquake. *Bulletin of the Seismological Society of America*, 98(2):823–831.

D. J. Wald, D. V. Helmberger, and T. H. Heaton (1991). Rupture model of the 1989 Loma Prieta earthquake from the inversion of strong-motion and broadband teleseismic data. *Bulletin of the Seismological Society of America*, 81(5):1540–1572.

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

A. Zerva and V. Zerva (2002). Spatial variation of seismic ground motions. *Appl. Mech. Rev*., 55(3):271–297.