

Correlation model for spatially distributed ground-motion intensities

Nirmal Jayaram^{*,†} and Jack W. Baker

Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, U.S.A.

SUMMARY

Risk assessment of spatially distributed building portfolios or infrastructure systems requires quantification of the joint occurrence of ground-motion intensities at several sites, during the same earthquake. The ground-motion models that are used for site-specific hazard analysis do not provide information on the spatial correlation between ground-motion intensities, which is required for the joint prediction of intensities at multiple sites. Moreover, researchers who have previously computed these correlations using observed ground-motion recordings differ in their estimates of spatial correlation. In this paper, ground motions observed during seven past earthquakes are used to estimate correlations between spatially distributed spectral accelerations at various spectral periods. Geostatistical tools are used to quantify and express the observed correlations in a standard format. The estimated correlation model is also compared with previously published results, and apparent discrepancies among the previous results are explained.

The analysis shows that the spatial correlation reduces with increasing separation between the sites of interest. The rate of decay of correlation typically decreases with increasing spectral acceleration period. At periods longer than 2 s, the correlations were similar for all the earthquake ground motions considered. At shorter periods, however, the correlations were found to be related to the local-site conditions (as indicated by site V_s30 values) at the ground-motion recording stations. The research work also investigates the assumption of isotropy used in developing the spatial correlation models. It is seen using the Northridge and Chi-Chi earthquake time histories that the isotropy assumption is reasonable at both long and short periods. Based on the factors identified as influencing the spatial correlation, a model is developed that can be used to select appropriate correlation estimates for use in practical risk assessment problems. Copyright © 2009 John Wiley & Sons, Ltd.

Received 6 October 2008; Revised 11 March 2009; Accepted 17 March 2009

KEY WORDS: spatial correlation; spectral accelerations; coherency; risk assessment; infrastructure systems

*Correspondence to: Nirmal Jayaram, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, U.S.A.

†E-mail: nirmalj@stanford.edu

Contract/grant sponsor: U.S. Geological Survey (USGS); contract/grant number: 07HQGR0031

1. INTRODUCTION

The probabilistic assessment of ground-motion intensity measures (such as spectral acceleration) at an individual site is a well researched topic. Several ground-motion models have been developed to predict median ground-motion intensities as well as dispersion about the median values e.g. [1–4]. Site-specific hazard analysis does not suffice, however, in many applications that require knowledge about the joint occurrence of ground-motion intensities at several sites, during the same earthquake. For instance, the risk assessment of portfolios of buildings or spatially distributed infrastructure systems (such as transportation networks, oil and water pipeline networks and power systems) requires the prediction of ground-motion intensities at multiple sites. Such joint predictions are possible, however, only if the correlation between ground-motion intensities at different sites are known e.g. [5, 6]. The correlation is known to be large when the sites are close to one another, and decays with increase in separation between the sites. Park *et al.* [7] report that ignoring or underestimating these correlations overestimates frequent losses and underestimates rare ones, and hence, it is important that accurate ground-motion correlation models be developed for loss assessment purposes. The current work analyzes correlations between the ground-motion intensities observed in recorded ground motions, in order to identify factors that affect these correlations and to select a correlation model that can be used for the joint prediction of spatially distributed ground-motion intensities in future earthquakes.

Ground-motion models that predict intensities at an individual site i due to an earthquake j take the following form:

$$\ln(Y_{ij}) = \ln(\bar{Y}_{ij}) + \varepsilon_{ij} + \eta_j \quad (1)$$

where Y_{ij} denotes the ground-motion parameter of interest (e.g. $S_a(T)$, the spectral acceleration at period T); \bar{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} denotes the intra-event residual, which is a random variable with zero mean and standard deviation σ_{ij} ; and η_j denotes the inter-event residual, which is a random variable with zero mean and standard deviation τ_j . The standard deviations, σ_{ij} and τ_j , are estimated as a part of the ground-motion model and are a function of the spectral period of interest, and in some models also as 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 [8] showed that a vector of spatially distributed intra-event residuals $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{dj})$ follows a multivariate normal distribution. Hence, the distribution of $\boldsymbol{\varepsilon}_j$ can be completely defined using the first two moments of the distribution, namely, the mean and variance of $\boldsymbol{\varepsilon}_j$, and the correlation between all ε_{i_1j} and ε_{i_2j} pairs. (Alternately, the distribution can be defined using the mean and the covariance of $\boldsymbol{\varepsilon}_j$, since the covariance completely specifies the variance and correlations.) Since the intra-event residuals are zero-mean random variables, the mean of $\boldsymbol{\varepsilon}_j$ is the zero vector of dimension d . The covariance, however, is not entirely known from the ground-motion models since the models only provide the variances of the residuals and not the correlation between residuals at two different sites.

Researchers, in the past, have computed these correlations using ground-motion time histories recorded during earthquakes [9–11]. Boore *et al.* [11] used the observations of peak ground acceleration (PGA , which equals $S_a(0)$) from the 1994 Northridge earthquake to compute the spatial correlations. Wang and Takada [10] computed the correlations using the observations of peak ground velocities (PGV) from several earthquakes in Japan and the 1999 Chi-Chi earthquake.

Goda and Hong [9] used the Northridge and Chi-Chi earthquake ground-motion records to compute the correlation between PGA residuals, as well as the correlation between residuals computed from spectral accelerations at three periods between 0.3 and 3 s. The results reported by these research works, however, differ in terms of the rate of decay of correlation with separation distance. For instance, while Boore *et al.* [11] report that the correlation drops to zero at a site separation distance of approximately 10 km, the non-zero correlations observed by Wang and Takada [10] extend past 100 km. Further, Goda and Hong [9] observe the differences between the correlation decay rate estimated using the Northridge earthquake records and the correlation decay rate based on the Chi-Chi earthquake records. To date, no explanation for these differences has been identified.

The current work uses observed ground motions to estimate correlations between ground-motion intensities (in particular, spectral accelerations). Factors that affect the rate of decay in the correlation with separation distance are identified. The work also provides probable explanations for the differing results reported in the literature. In this study, an emphasis is placed on developing a standard correlation model that can be used for predicting spatially distributed ground-motion intensities for risk assessment purposes.

2. MODELING CORRELATIONS USING SEMIVARIOGRAMS

Geostatistical tools are widely used in several fields for modeling spatially distributed random vectors (also called random functions) [12, 13]. The current research work takes advantage of this well-developed approach to model the correlation between spatially distributed ground-motion intensities. The needed tools are briefly described in this section.

Let $\mathbf{Z} = (Z_{u_1}, Z_{u_2}, \dots, Z_{u_d})$ denote a spatially distributed random function, where u_i denotes the location of the site i ; Z_{u_i} is the random variable of interest (in this case, $\varepsilon_{u_i j}$ from Equation (1)) at site location u_i and d denotes the total number of sites. The correlation structure of the random function \mathbf{Z} can be represented by a semivariogram, which is a measure of the average dissimilarity between the data [13]. Let u and u' denote the two sites separated by \mathbf{h} . The semivariogram ($\gamma(u, u')$) is computed as half the expected squared difference between Z_u and $Z_{u'}$.

$$\gamma(u, u') = \frac{1}{2} [E\{Z_u - Z_{u'}\}^2] \quad (2)$$

The semivariogram defined in Equation (2) is location-dependent and its inference requires repetitive realizations of \mathbf{Z} at locations u and u' . Such repetitive measurements of $\{Z_u, Z_{u'}\}$ are, however, never available in practice (e.g. in the current application, one would need repeated observations of ground motions at every pair of sites of interest). Hence, it is typically assumed that the semivariogram does not depend on site locations u and u' , but only on their separation \mathbf{h} . The stationary semivariogram ($\gamma(\mathbf{h})$) can then be obtained as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2} [E\{Z_u - Z_{u+\mathbf{h}}\}^2] \quad (3)$$

Equation (2) can be replaced with Equation (3) if the random function (\mathbf{Z}) is second-order stationary. Second-order stationarity implies that (i) the expected value of the random variable Z_u is a constant across space and (ii) the two-point statistics (measures that depend on Z_u and $Z_{u'}$) depend only on the separation between u and u' , and not on the actual locations (i.e. the statistics depend on the separation vector \mathbf{h} between u and u' and not on u and u' as such). A stationary

semivariogram can be estimated from a data set as follows:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} \{z_{u_\alpha} - z_{u_\alpha+\mathbf{h}}\}^2 \tag{4}$$

where $\hat{\gamma}(\mathbf{h})$ is the experimental stationary semivariogram (estimated from a data set); z_u denotes the data value at location u ; $N(\mathbf{h})$ denotes the number of pairs of sites separated by \mathbf{h} ; and $\{z_{u_\alpha}, z_{u_\alpha+\mathbf{h}}\}$ denotes the α th such pair. A stationary semivariogram is said to be isotropic if it is a function of the separation distance ($h = \|\mathbf{h}\|$) rather than the separation vector \mathbf{h} .

The function $\hat{\gamma}(\mathbf{h})$ provides a set of experimental values for a finite number of separation vectors \mathbf{h} . A continuous function must be fitted based on these experimental values in order to deduce semivariogram values for any possible separation \mathbf{h} . A valid (permissible) semivariogram function needs to be negative definite so that the variances and conditional variances corresponding to this semivariogram are non-negative. In order to satisfy this condition, the semivariogram functions are usually chosen to be the linear combinations of basic models that are known to be permissible. These include the exponential model, the Gaussian model, the spherical model and the nugget effect model.

The exponential model, in an isotropic case (i.e. the vector distance \mathbf{h} is replaced by a scalar separation length $\|\mathbf{h}\|$, also denoted as h), is expressed as follows:

$$\gamma(h) = a[1 - \exp(-3h/b)] \tag{5}$$

where a and b are the sill and the range of the semivariogram function, respectively (Figure 1(a)). The sill of a semivariogram equals the variance of Z_u , while the range is defined as the separation distance h at which $\gamma(h)$ equals 0.95 times the sill of the exponential semivariogram.

The Gaussian model is as follows:

$$\gamma(h) = a[1 - \exp(-3h^2/b^2)] \tag{6}$$

The sill and the range of a Gaussian semivariogram are as defined for an exponential semivariogram.

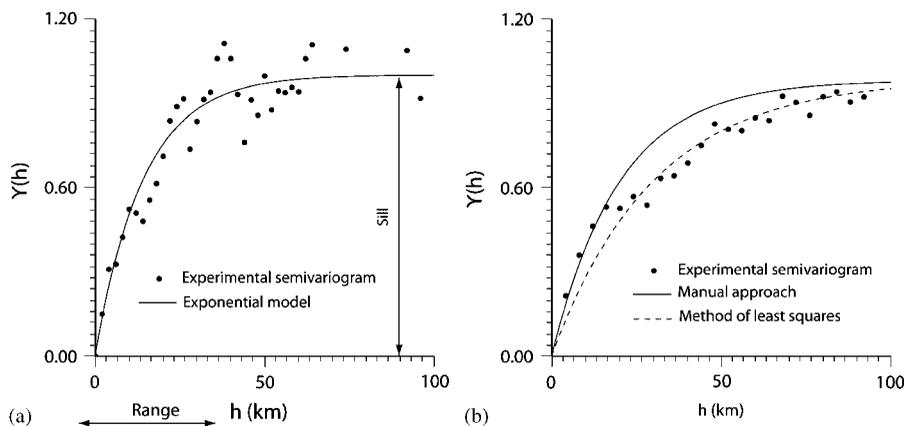


Figure 1. (a) Parameters of a semivariogram and (b) Semivariograms fitted to the same data set using the manual approach and the method of least squares.

The spherical model is as follows:

$$\begin{aligned}\gamma(h) &= a \left[\frac{3}{2} \left(\frac{h}{b} \right) - \frac{1}{2} \left(\frac{h}{b} \right)^3 \right] \quad \text{if } h \leq b \\ &= a \quad \text{otherwise}\end{aligned}\quad (7)$$

where a and b are again the sill and range of the semivariogram, respectively. The range of a spherical semivariogram is the separation distance at which $\gamma(h)$ equals a .

The nugget effect model can be described as:

$$\gamma(h) = a[I(h > 0)] \quad (8)$$

where $I(h > 0)$ is an indicator variable that equals 1 when $h > 0$ and equals 0 otherwise.

The covariance structure of Z is completely specified by the semivariogram function and the sill and the range of the semivariogram. It can be theoretically shown that the following relationship holds [13]:

$$\gamma(h) = a(1 - \rho(h)) \quad (9)$$

where $\rho(h)$ denotes the correlation coefficient between Z_u and Z_{u+h} . It can also be shown that the sill of the semivariogram equals the variance of Z_u . Therefore, it would suffice to estimate the semivariogram of a random function in order to determine its covariance structure. Moreover, based on Equations (5) (for instance) and (9), it can be seen that a large range implies a small rate of increase in $\gamma(h)$ and therefore, large correlations between Z_u and Z_{u+h} . Further, it can be seen from Equation (8) that the nugget effect model specifies zero correlation for all non-zero separation distances.

In the current work, correlations between ground-motion intensities at different sites are represented using semivariograms. Ground-motion recordings from past earthquakes are used to estimate ranges of semivariograms and to identify the factors that could affect the estimates. Throughout this work, the semivariograms are assumed to be second-order stationary. Second-order stationarity is assumed so that the data available over the entire region of interest can be pooled and used for estimating semivariogram sills and ranges. In the current work, like many other works involving spatial correlation estimation, the semivariograms are also assumed to be isotropic. The assumptions of stationarity and isotropy are investigated in more detail subsequently in this paper.

3. COMPUTATION OF SEMIVARIOGRAM RANGES FOR INTRA-EVENT RESIDUALS USING EMPIRICAL DATA

As mentioned earlier, the covariance of intra-event residuals can be represented using a semivariogram, whose functional form (e.g. exponential model), sill and range need to be determined. This section discusses the semivariograms estimated based on observed ground-motion time histories.

For a given earthquake, it can be seen from Equation (1) that

$$\varepsilon_i + \eta = \ln(Y_i) - \ln(\bar{Y}_i) \quad (10)$$

Let $\hat{\varepsilon}_i$ denote the normalized intra-event residual at site i . (The subscript j in Equation (1) is no longer used since the residuals used in these calculations are observed during a single earthquake.) $\hat{\varepsilon}_i$ is computed as follows:

$$\hat{\varepsilon}_i = \frac{\varepsilon_i}{\sigma_i} \quad (11)$$

where σ_i denotes the standard deviation of the intra-event residual at site i . Further, let $\tilde{\varepsilon}_i$ denote the sum of the intra-event residual (ε_i) and inter-event residual (η) normalized by the standard deviation of the intra-event residual (σ_i). $\tilde{\varepsilon}_i$ can be computed as follows:

$$\tilde{\varepsilon}_i = \frac{\varepsilon_i + \eta}{\sigma_i} = \frac{\ln(Y_i) - \ln(\bar{Y}_i)}{\sigma_i} \quad (12)$$

While assessing covariances, it is convenient to work with $\tilde{\varepsilon}$'s rather than ε 's, since $\tilde{\varepsilon}$'s are homoscedastic (i.e. constant variance) with unit variance unlike the ε 's.

Since the inter-event residual (η), computed at any particular period, is a constant across all the sites during a given earthquake, the experimental semivariogram function of $\tilde{\varepsilon}$ can be obtained as follows (based on Equation (4)):

$$\begin{aligned} \hat{\gamma}(h) &= \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [\hat{\varepsilon}_{u_\alpha} - \hat{\varepsilon}_{u_\alpha+h}]^2 \\ &= \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\frac{\ln(Y_{u_\alpha}) - \ln(\bar{Y}_{u_\alpha}) - \eta}{\sigma_{u_\alpha}} - \frac{\ln(Y_{u_\alpha+h}) - \ln(\bar{Y}_{u_\alpha+h}) - \eta}{\sigma_{u_\alpha+h}} \right]^2 \\ &\approx \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} \left[\frac{\ln(Y_{u_\alpha}) - \ln(\bar{Y}_{u_\alpha})}{\sigma_{u_\alpha}} - \frac{\ln(Y_{u_\alpha+h}) - \ln(\bar{Y}_{u_\alpha+h})}{\sigma_{u_\alpha+h}} \right]^2 \\ &= \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [\tilde{\varepsilon}_{u_\alpha} - \tilde{\varepsilon}_{u_\alpha+h}]^2 \end{aligned} \quad (13)$$

where $\tilde{\varepsilon}$ is defined by Equation (12); $(u_\alpha, u_\alpha+h)$ denotes the location of a pair of sites separated by h ; $N(h)$ denotes the number of such pairs; Y_{u_α} denotes the ground-motion intensity at location u_α ; and σ_{u_α} is the standard deviation of the intra-event residual at location u_α . The sill of the semivariogram of $\tilde{\varepsilon}$ (i.e. the sill of $\hat{\gamma}(h)$) should equal 1 since the $\tilde{\varepsilon}$'s have a unit variance. Hence, based on Equation (9), it can be concluded that

$$\hat{\gamma}(h) = 1 - \hat{\rho}(h) \quad (14)$$

where $\hat{\rho}(h)$ is the estimate of $\rho(h)$.

Incidentally, Equation (13) shows that the covariances of intra-event residuals can be estimated without having to account for the inter-event residual η . As indicated, Equation (13) involves an approximation due to the mild assumption that $\eta/\sigma_{u_\alpha} = \eta/\sigma_{u_\alpha+h}$. The Boore and Atkinson [1] model, which is used in the current work, suggests that the standard deviation of the intra-event residuals depends only on the period at which the residuals are computed, and hence, it can be inferred that this approximation is reasonable. Incidentally, though the current work only uses the Boore and Atkinson [1] ground-motion model, the results obtained were found to be similar when an alternate model, namely, the Chiou and Youngs [3] model, was used.

The ground-motion databases typically report recordings in two orthogonal horizontal directions. For instance, the PEER NGA database [14] provides the fault-normal and the fault-parallel components of the ground motions for each earthquake. In the current work, it was found that the correlations computed using both the fault-normal and the fault-parallel time histories were similar. Hence, only results corresponding to the fault-normal orientation are reported here. In fact, Baker and Jayaram [15] and Bazzurro *et al.* [16] used several sets of recorded and simulated ground motions to show that the estimated correlations are independent of the ground-motion component used.

3.1. Construction of experimental semivariograms using empirical data

Figure 1(a) shows a sample semivariogram constructed from empirical data. The first step in obtaining such a semivariogram is to compute site-to-site distances for all pairs of sites and place them in different bins based on the separation distances. For example, the bins could be centered at multiples of h km with bin widths of δh km ($\delta h \leq h$). All pairs of sites that fall in the bin centered at h km (i.e. the sites that are separated by a distance $\in (h - \delta h/2, h + \delta h/2)$) are used to compute $\hat{\gamma}(h)$ (based on Equation (4)). If δh is chosen to be very small, it can result in few pairs of sites in the bins, which will affect the robustness of the results obtained. On the other hand, a large value of δh will mix site pairs with differing distances reducing the resolution of the experimental semivariograms. In the current work, experimental semivariograms are obtained using $\delta h = 2$ km (unless stated otherwise), since this was seen to be the smallest value that results in a reasonable number of site pairs in the bins.

The semivariogram shown in Figure 1(a) has an exponential form with a sill of 1 and a range of 40 km. This model can be expressed as follows (based on Equation (5)):

$$\gamma(h) = 1 - \exp(-3h/40) \quad (15)$$

The correlation function corresponding to this model equals $\rho(h) = 1 - \gamma(h) = \exp(-3h/40)$ (based on Equation (14)).

An easy and transparent method to determine the model and the model parameters is to fit the experimental semivariogram values obtained at discrete separation distances manually. Suppose that $\gamma(h)$ can be expressed as follows:

$$\gamma(\mathbf{h}) = c_0 \gamma_0(\mathbf{h}) + \sum_{n=1}^N c_n \gamma_n(\mathbf{h}) \quad (16)$$

where $\gamma_0(\mathbf{h})$ is a pure nugget effect and $\gamma_n(\mathbf{h})$ is a spherical, exponential or Gaussian model (as defined in Equations (5)–(8)); c_n is the contribution of the model n to the semivariogram; and N is the total number of models used (excluding the nugget effect). The ranges and the contributions of the models can be systematically varied to obtain the best fit to the experimental semivariogram values.

In the following sections, priority is placed on building models that fit the empirical data well at short distances, even if this requires some misfit with empirical data at large separation distances, because it is more important to model the semivariogram structure well at short separation distances. This is because the large separation distances are associated with low correlations, which thus have relatively a little effect on joint distributions of ground-motion intensities. In addition to having low correlation, widely separated sites also have a little impact on each other due to an effective ‘shielding’ of their influence by more closely-located sites [13]. Figure 1(b) shows the sample

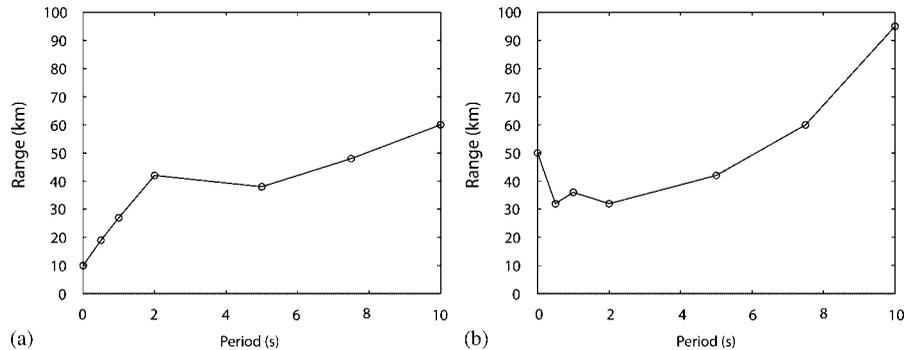


Figure 2. Range of semivariograms of \tilde{z} , as a function of the period at which \tilde{z} values are computed: (a) the residuals are obtained using the Northridge earthquake data and (b) the residuals are obtained using the Chi-Chi earthquake data.

semivariograms fitted to a data set using the manual approach and the method of least squares. It can be seen that, at small separations, the manually fitted semivariogram is a better model than the one fitted using the method of least squares. More detailed discussion on the advantages of using manual-fitting rather than least-squares fitting follows in Section 5, where the proposed approach is also compared with approaches used in previous research on this topic.

3.2. 1994 Northridge earthquake recordings

This section discusses the ranges of semivariograms estimated using observed Northridge earthquake ground motions. The manual fitting approach described previously is used to compute ranges of the semivariograms of \tilde{z} 's (obtained based on the Northridge earthquake time histories) computed at seven periods ranging between 0 and 10 s. Of the three functional forms considered (Equations (5)–(7)), the exponential model is found to provide the ‘best fit’ (particularly at small separations) for experimental semivariograms obtained using \tilde{z} 's computed at several different periods, based on recordings from different earthquakes. The constancy of the semivariogram function across periods makes it simpler to specify a standard correlation model for the \tilde{z} 's. Moreover, the use of a single model enables a direct comparison of the correlations between residuals computed at different periods, using only the ranges of the semivariograms. The ranges of these estimated semivariograms are plotted against period in Figure 2(a). The semivariogram fits corresponding to all the periods considered can be found in Baker and Jayaram [15].

It can be observed from Figure 2(a) that the estimated range of the semivariogram tends to increase with period. As described earlier, it can be inferred that the \tilde{z} values at long periods show larger correlations than those at short periods. This is consistent with comparable past studies of ground-motion coherency, which has been widely researched in the past. Coherency can be thought of as a measure of similarity in two spatially separated ground-motion time histories. Der Kiureghian [17] reports that coherency is reduced by the scattering of waves during propagation, and that this reduction is greater for high-frequency waves. High-frequency waves, which have short wavelengths, tend to be more affected by small-scale heterogeneities in the propagation path, and as a result tend to be less coherent than long period ground waves [18]. It is reasonable to expect highly coherent ground motions to exhibit correlated peak amplitudes (i.e. spectral accelerations) as well. Since the \tilde{z} 's studied here, which quantify these peak amplitudes, tend to show the same

correlation trend with period as previous coherency studies, it may be that a similar wave scattering mechanism is partially responsible for the correlation trends observed here.

The Northridge earthquake data used for the above analysis are obtained from the NGA database. In order to exclude records whose characteristics differ from those used by the ground-motion modelers for data analysis, in most cases, only records used by the authors of the Boore and Atkinson [1] ground-motion model are considered. For the purposes of this paper, these records are denoted as 'usable records'. The semivariograms of residuals computed at periods of 5, 7.5 and 10 s, however, are obtained using all available Northridge records in the NGA database. This is on account of the limited number of Northridge earthquake recordings at extremely long periods. At 5 s the residuals can be computed using 158 total available records, while 66 of these are used by the ground-motion model authors. Since there is a reasonable number of records available in both cases, a semivariogram constructed using all 158 records (denoted SV_1) can be compared with that estimated from the usable 66 records (in this case, the bin size was increased to 4 km to compensate for the lack of available records) (denoted SV_2). The ranges of the two semivariograms, SV_1 and SV_2 , are 40 and 30 km, respectively. This shows that there is a slight difference in the estimated ranges, which could be due to the additional correlated systematic errors introduced by the extra records.

As mentioned in Section 1, correlation between intensities estimated using the fault-normal components are discussed in this paper. This is because the correlations obtained using the fault-normal and the fault-parallel ground motions were found to be similar. For example, the semivariogram of $\tilde{\epsilon}$'s computed at 2 s, based on the fault-parallel ground motions recorded during the Northridge earthquake was found to be reasonably modeled using an exponential function with a unit sill and a range of 36 km. The corresponding range for the semivariogram based on the fault-normal ground motions equals 42 km. Similar results were observed when the residuals were computed at other periods, and using other earthquake recordings.

3.3. 1999 Chi-Chi earthquake

In this section, the semivariogram ranges of $\tilde{\epsilon}$'s from the Chi-Chi earthquake recordings are presented. The Chi-Chi earthquake ground motions came from the NGA database. Only records used by Boore and Atkinson [1] ground-motion model are considered. The summary plot of the estimated ranges is shown in Figure 2(b). The following can be observed from the figures:

- (a) As seen with the Northridge earthquake data, the range of the semivariogram typically increases with period (An exception is observed when the PGA's are considered, and this is explored further subsequently in this paper.).
- (b) The ranges are higher, in general, than those observed based on the Northridge earthquake data (Figure 2(a)). This is consistent with observations made by other researchers considering Northridge and Chi-Chi earthquake data e.g. [9].

The large ranges obtained here, relative to the comparable results from Northridge, can be explained using the V_s30 values (average shear-wave velocities in the top 30 m of the soil) at the recording stations. The V_s30 values are commonly used in ground-motion models as indicators of the effects of local-site conditions on the ground motion. $\tilde{\epsilon}$'s are affected if the predicted ground-motion intensities are affected by inaccurate V_s30 values, or if the V_s30 's are inadequate to capture the local-site effects adequately (i.e. the ground-motion models do not entirely capture the local-site effects using V_s30 values).

Close to 70% of the Taiwan site V_s30 values are inferred from Geomatrix site classes, while the rest of the V_s30 values are measured (NGA database). Since closely spaced sites are likely to belong to the same site class and possess similar (and unknown) V_s30 values, errors in the inferred V_s30 values are likely to be correlated among sites that are close to each other. Such correlated V_s30 measurement errors will result in correlated prediction errors at all these closely spaced sites, which will increase the range of the semivariograms.

The larger ranges of semivariograms estimated using the Chi-Chi earthquake ground motions may also be due to possible correlation between the true V_s30 values (and not just the correlation between the V_s30 errors). Larger correlation between the V_s30 's indicate a more homogeneous soil (homogeneous in terms of properties that affect site effects but not accounted by the ground-motion models). In such cases, if a ground-motion model does not accurately capture the local-site effect at one site, it is likely to produce similar prediction errors in a cluster of closely spaced sites (on account of the homogeneity). Castellaro *et al.* [19] compared the site-dependent seismic amplification factors (F_a , the site amplification factor is defined as the amplification of the ground-motion spectral level at a site with respect to that at a reference ground condition [20]) observed during the 1989 Loma Prieta earthquake to the corresponding site V_s30 values. They found substantial scatter in the plot of F_a versus V_s30 , and also found that this scatter was more pronounced at short periods (below 0.5 s) than at longer periods. This suggests that ground-motion intensity predictions based on V_s30 will have errors, particularly at periods below 0.5 s.

Figures 3(a) and (b) show semivariograms of the normalized V_s30 values (the V_s30 semivariogram is not to be confused with the $\tilde{\epsilon}$ semivariogram) at the Northridge earthquake recording stations and the Chi-Chi earthquake recording stations, respectively (Normalization involves scaling the V_s30 values so that the normalized V_s30 values have a unit variance to enable a direct comparison of the semivariograms.). Figure 3(a) shows significant scatter at all separation distances indicating zero correlation at all separations. In contrast, Figure 3(b) indicates that the Taiwan V_s30 values have significant spatial correlation. This suggests that $\tilde{\epsilon}$'s may have additional spatial correlation in Taiwan, due to homogeneous site effects that cause correlated prediction errors.

As mentioned previously, one notable aberration in the plot of range versus period (Figure 2(b)) is the large range observed when the residuals are computed at 0 s as compared with some of the longer periods. This is not consistent with the coherency argument of the previous section. It can, however, be explained using the relationship between the range and the V_s30 's described in the above paragraphs. The inaccuracies in ground-motion prediction based on V_s30 's will reflect in increased correlation between the residuals computed at nearby sites. These inaccuracies are larger at short periods (below 0.5 s) [19], which explains the larger correlation between the residuals (which ultimately results in the larger range observed) computed using PGAs.

One final test that was considered here was whether spatial correlations differed for near-fault ground motions experiencing directivity. Baker [21] identified pulse-like ground motions from the NGA database based on wavelet analysis. Thirty such pulses were identified in the fault-normal components of the Chi-Chi earthquake recordings. Experimental semivariograms of residuals were computed using these pulse-like ground motions, and their ranges were estimated. It was seen that the ranges were reasonably similar to those obtained using all usable ground motions (i.e. pulse-like and non-pulse-like). Since the available pulse-like ground-motion data set is very small, however, the results obtained were not considered to be sufficiently reliable, and hence not considered further in this paper. A more detailed analysis can be found in Baker and Jayaram [15] and Bazzurro *et al.* [16].

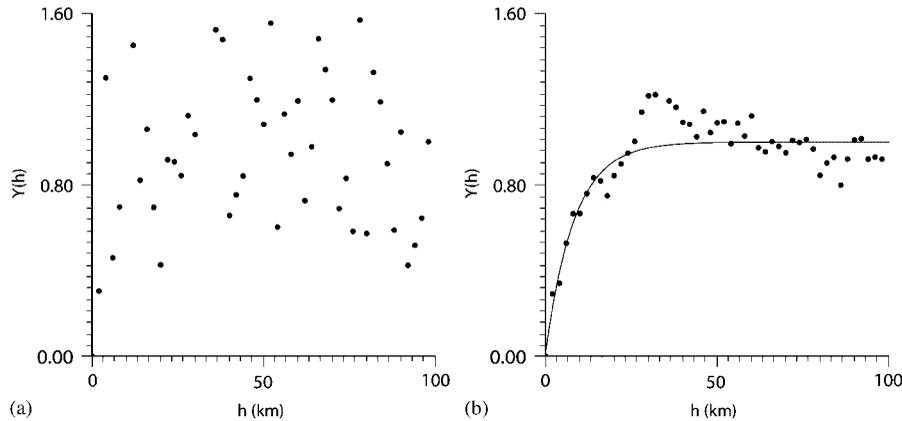


Figure 3. (a) Experimental semivariogram obtained using normalized V_s30 's at the recording stations of the Northridge earthquake. No semivariogram is fitted on account of the extreme scatter and (b) experimental semivariogram obtained using normalized V_s30 's at the recording stations of the Chi-Chi earthquake. The range of the fitted exponential semivariogram equals 25 km.

Based on the discussion in this section, it can be seen that the correlated V_s30 values and the correlated V_s30 measurement errors are the possible reasons for the larger ranges estimated in Section 3.3 than in Section 3.2. Other factors, such as the size of the rupture areas, may also affect the correlations. These factors could not, however, be investigated with the limited data set available.

3.4. Other earthquakes

The correlations computed using data from the 2003 M5.4 Big Bear City earthquake, the 2004 M6.0 Parkfield earthquake, the 2005 M5.1 Anza earthquake, the 2007 M5.6 Alum Rock earthquake and the 2008 M5.4 Chino Hills earthquake are presented in this section. The time histories for these earthquakes were obtained from the CESMD database [22]. The V_s30 data used for these computations came from the CESMD database [22] (for the Parkfield earthquake) and the U.S. Geological Survey V_s30 maps (for the other earthquakes) [23].

Exponential models are fitted to experimental semivariograms of $\tilde{\epsilon}$'s computed using the time histories from the above-mentioned earthquakes, at periods ranging from 0–10 s. Figure 4 shows plots of range versus period for the Big Bear City, Parkfield, Alum Rock, Anza and Chino Hills earthquake residuals. The ranges of the semivariograms are generally seen to increase with period, which is consistent with findings from the Chi-Chi and the Northridge earthquake data. It can also be seen from the figure that, at short periods, the ranges obtained from the Anza earthquake data are larger than those from the other earthquakes considered. On the other hand, the ranges computed using the Parkfield earthquake data are fairly small at short periods. Semivariograms of the V_s30 's at the recording stations for all five earthquakes of interest were computed. The semivariogram range computed using the Anza earthquake V_s30 's was found to be the largest at 40 km, while the ranges computed from the Chino Hills, Big Bear City, Alum Rock and Parkfield earthquake data were smaller at 35, 30, 18 and approximately 0 km, respectively. The estimated ranges of the semivariograms of the residuals and of the V_s30 's reinforce the argument made previously that

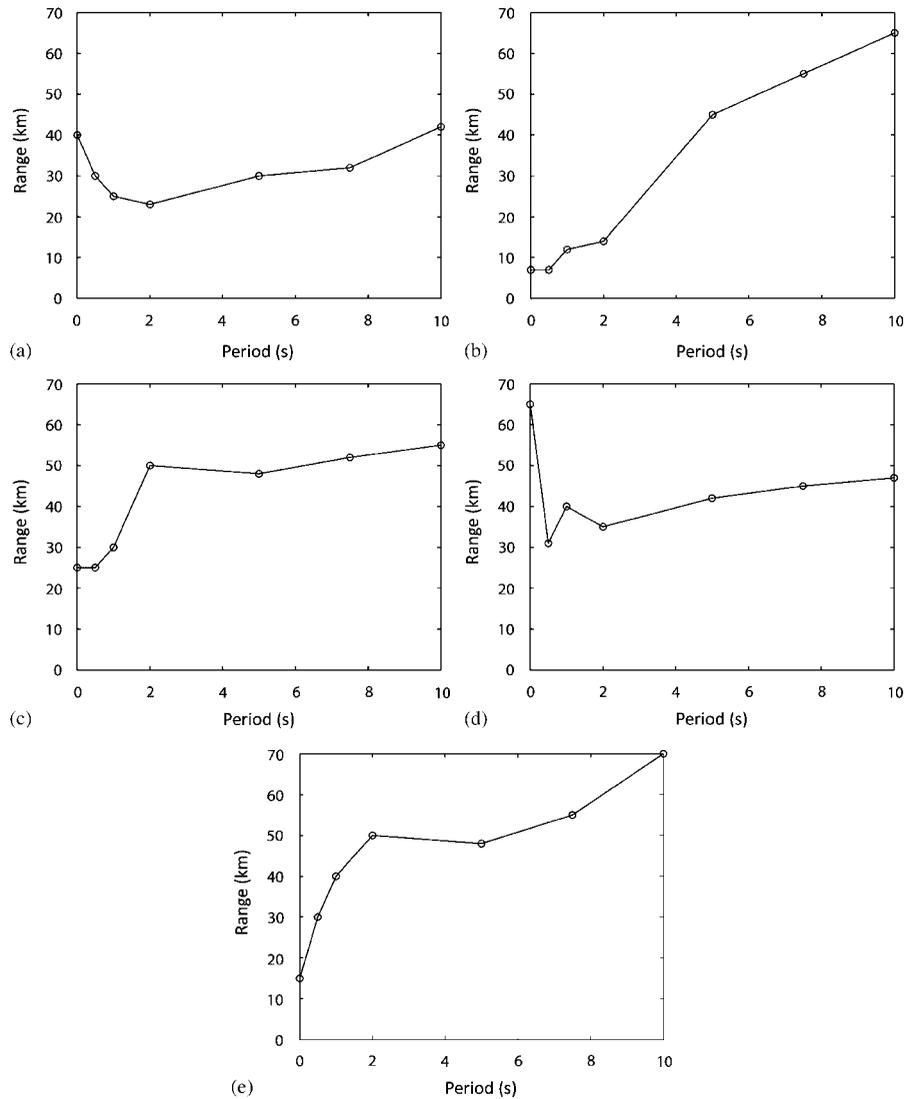


Figure 4. Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the: (a) Big Bear City earthquake data; (b) Parkfield earthquake data; (c) Alum Rock earthquake data; (d) Anza earthquake data; and (e) Chino Hills earthquake data.

clustering in the V_s30 values (as indicated by a large range of the V_s30 semivariogram) results in increased correlation among the residuals (the low PGA-based range estimated using the Chino hills earthquake data seems to be an exception, however). This trend is seen in Figure 5, which shows the range of PGA-based residuals plotted against the range of the V_s30 's, for the earthquakes considered in this work. This dependence on the V_s30 range seems to be lesser at longer periods, which is in line with the observations of Castellaro *et al.* [19] that the scatter in the plot of F_d versus

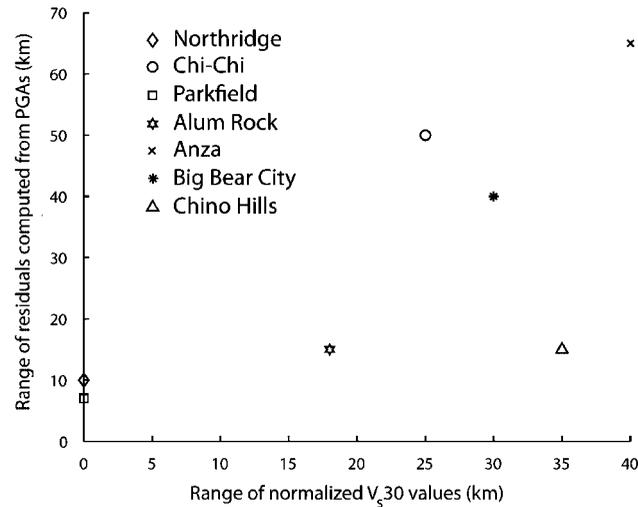


Figure 5. Ranges of residuals computed using PGAs versus ranges of normalized V_{s30} values.

V_{s30} is greater at short periods than at long periods. The authors hypothesize that the reduced dependence of range on V_{s30} 's at long periods could also be because the long-period ranges are considerably influenced by factors other than V_{s30} values, such as coherency as explained in Section 3.2 and prediction errors unrelated to V_{s30} 's (which are likely since the ground-motion models are fitted using much fewer data points at long periods). Finally, an additional advantage of considering these five additional events is that earthquakes covering a range of magnitudes have been studied. No trends of range with magnitude were detected.

A few research works studying spatial correlations use ground-motion recordings from earthquakes in Japan, based on the data provided in the KiK Net [24]. In this work, data from the 2004 Mid Niigata Prefecture earthquake and the 2005 Miyagi-Oki earthquake were explored. Though the number of sites at which the ground-motion recordings are available is fairly large, most recording stations are far away from each another. The KiK Net [24] consists of 681 recording stations, of which only 19 pairs of stations are within 10 km of one another. As explained in Section 3.2, it is important to accurately model the semivariogram at short separation distances, particularly at separation distances below 10 km. Hence, the recordings from the KiK Net [24] were not considered further for studying the ranges of semivariograms.

3.5. A predictive model for spatial correlations

The above sections presented the spatial correlations computed using recorded ground motions from several past earthquakes. In this section, these correlation estimates are used to develop a model that can be used to select appropriate correlation estimates for risk assessment purposes.

Figure 6(a) shows the ranges computed using various earthquake data as a function of period. From a practical perspective, despite the wide differences in the characteristics of the earthquakes considered, the ranges computed are quite similar, particularly at periods longer than 2 s. At short periods (below 2 s), however, there are considerable differences in the estimated ranges depending on the ground-motion time histories used. The previous sections suggested empirically that the

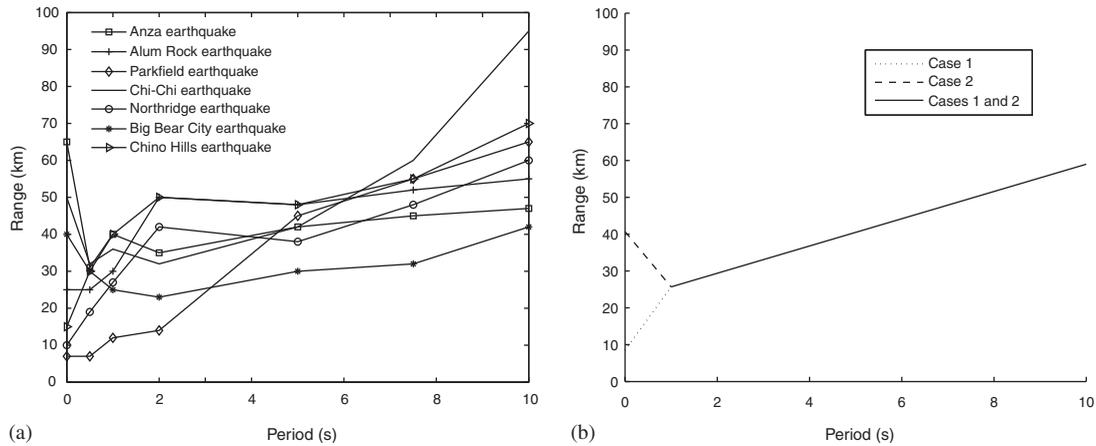


Figure 6. (a) Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained from six different sets of time histories as shown in the figure and (b) range of semivariograms of $\tilde{\varepsilon}$ predicted by the proposed model as a function of the period.

differences in correlation of $\tilde{\varepsilon}$'s is in large part explained by the V_s30 values at the recording stations for these earthquakes. Hence, the following cases can be considered for decision making:

Case 1: If the V_s30 values do not show or are not expected to show clustering (i.e. the geologic condition of the soil varies widely over the region (this can also be quantified by constructing the semivariogram of the V_s30 's as explained previously)), the smaller ranges reported in Figure 6(a) will be appropriate.

Case 2: If the V_s30 values show or are expected to show clustering (i.e. there are clusters of sites in which the geologic conditions of the soil are similar), the larger ranges reported in Figure 6(a) should be chosen.

Based on these conclusions, the following model was developed to predict a suitable range based on the period of interest:

At short periods ($T < 1$ s), for case 1:

$$b = 8.5 + 17.2T \quad (17)$$

At short periods ($T < 1$ s), for case 2:

$$b = 40.7 - 15.0T \quad (18)$$

At long periods ($T \geq 1$ s), for both cases 1 and 2:

$$b = 22.0 + 3.7T \quad (19)$$

where b denotes the range of the exponential semivariograms (Equation (5)), and T denotes the period. Based on this model, the correlation between normalized intra-event residuals separated by h km is obtained as follows (follows from Equations (5) and (14)):

$$\rho(h) = \exp(-3h/b) \quad (20)$$

It is to be noted that the correlations between intra-event residuals will exactly equal the correlations between normalized intra-event residuals defined above.

The plot of the predicted range versus period is shown in Figure 6(b). The model has been developed based on only seven earthquakes, but since the trends exhibited were found to be similar for these seven, it can be expected that the model will predict reasonable ranges for future earthquakes.

The predictive model can be used for simulating correlated ground-motion fields for a particular earthquake as follows:

Step 1: Obtain median ground-motion values (denoted \bar{Y}_{ij} in Equation (1)) at the sites of interest using a ground-motion model.

Step 2: Probabilistically generate (simulate) the inter-event residual term (η_j in Equation (1)), which follows a univariate normal distribution. The mean of the inter-event residual is zero, and its standard deviation can be obtained using ground-motion models.

Step 3: Simulate the intra-event residuals (ε_{ij} in Equation (1)) using the standard deviations from the ground-motion models and the correlations from Equations (17)–(20).

Step 4: Combine the three terms generated in Steps 1–3 using Equation (1) to obtain simulated ground-motion intensities at the sites of interest.

4. ISOTROPY OF SEMIVARIOGRAMS

This section examines the assumption of isotropy of semivariograms using the ground motions discussed previously.

4.1. Isotropy of intra-event residuals

A stationary semivariogram ($\gamma(\mathbf{h})$) is said to be isotropic if it depends only on the separation distance $h = \|\mathbf{h}\|$, rather than the separation vector \mathbf{h} . Anisotropy is said to be present when the semivariogram is also influenced by the orientation of the data locations. The presence of anisotropy can be studied using directional semivariograms [13]. Directional semivariograms are obtained as shown in Equation (4) except that the estimate is obtained using only pairs of $(z_{u_x}, z_{u_x+\mathbf{h}})$ such that the azimuth of the vector \mathbf{h} are identical and as specified for all the pairs. Since an isotropic semivariogram is independent of data orientation, the directional semivariograms obtained considering any specific azimuth will be identical to the isotropic semivariogram if the data are in fact isotropic. Differences between the directional semivariograms indicate one of two different forms of anisotropy, namely, geometric anisotropy and zonal anisotropy. Geometric anisotropy is said to be present if directional semivariograms with differing azimuths have differing ranges. Zonal anisotropy is indicated by a variation in the sill with azimuth.

4.2. Construction of a directional semivariogram

A directional semivariogram is specified by several parameters, as illustrated in Figure 7(a). The parameters include the azimuth of the direction vector (the azimuth angle (ϕ) is measured from the North), the azimuth tolerance ($\delta\phi$), the bin separation (h) and the bin width (δh). A semivariogram obtained using all pairs of points irrespective of the azimuth is known as the omni-directional semivariogram, and is an accurate measure of spatial correlation in the presence of isotropy. (The semivariograms that have been described in the previous sections are omni-directional semivariograms.) In determining the experimental semivariogram in any bin, only pairs of sites separated by distance ranging between $[h - \delta h/2, h + \delta h/2]$, and with azimuths ranging

between $[\phi - \delta\phi, \phi + \delta\phi]$ are considered. For example, let α be a site located in a two-dimensional region, as shown in Figure 7(a). It is intended to construct a directional semivariogram with an azimuth of ϕ (as marked in the figure). The computation of the experimental semivariogram value ($\hat{\gamma}(\mathbf{h})$) involves pairing up the data values at all sites falling within the hatched region (the region that satisfies the conditions on the separation distance and the azimuth, as mentioned above) with the data values at site α (i.e. u_α). The area of the hatched region is defined by the azimuth tolerance used and can be seen to increase with increase in separation distance (h) (Figure 7(a)). For large values of h , the area of the hatched region will be undesirably large and hence, in addition to placing constraints on the azimuth tolerance, a constraint is explicitly specified on the bandwidth of the region of interest, as marked in the figure.

It is usually difficult to compute experimental directional semivariograms on account of the need to obtain pairs of sites oriented along pre-specified directions. Hence, it is required that the bin width, the azimuth tolerance and the bandwidth be specified liberally while constructing directional semivariograms. The results reported in this paper are obtained by considering a bin separation of 4 km, a bin width of 4 km, an azimuth tolerance of 10° and a bandwidth of 10 km. Directional semivariograms are plotted for azimuths of 0, 45 and 90° in order to capture the effects of anisotropy, if any.

4.3. Test for anisotropy using Northridge ground-motion data

Figure 7(b–d) shows the omni-directional and the three directional experimental semivariograms of the 2 s $\tilde{\epsilon}$'s from the Northridge earthquake data. The semivariogram function shown in the figures is the exponential model with a unit sill and a range of 42 km. This exponential model (obtained assuming isotropy in Section 3.2) fits all the experimental directional semivariograms reasonably well (at short separations, which are of interest). This is a good indication that the semivariogram is isotropic. Similar results were obtained at other periods and for other earthquakes [15, 16].

5. COMPARISON WITH PREVIOUS RESEARCH

Researchers have previously computed the correlation between ground-motion intensities using observed PGA, PGV and spectral accelerations. These works, however, differ widely in the estimated rate of decay of correlation with separation distance. This section compares the results observed in the current work to those in the literature and also discusses possible reasons for the apparent inconsistencies in the previous estimates.

Wang and Takada [10] used the ground-motion relationship of Annaka *et al.* [25] to compute the normalized auto-covariance function of residuals computed using the Chi-Chi earthquake PGV. They used an exponential model to fit the discrete experimental covariance values and reported a result that is equivalent to the following semivariogram:

$$\gamma(h) = 1 - \exp(-3h/83.4) \quad (21)$$

This semivariogram has a unit sill and a range of 83.4 km (from Equation (5)). The current work does not consider the spatial correlation between PGV-based residuals. The PGVs, however, are comparable to spectral accelerations computed at moderate periods (0.5 to 1 s), and hence, the semivariogram ranges of residuals computed from PGVs can be qualitatively compared with the

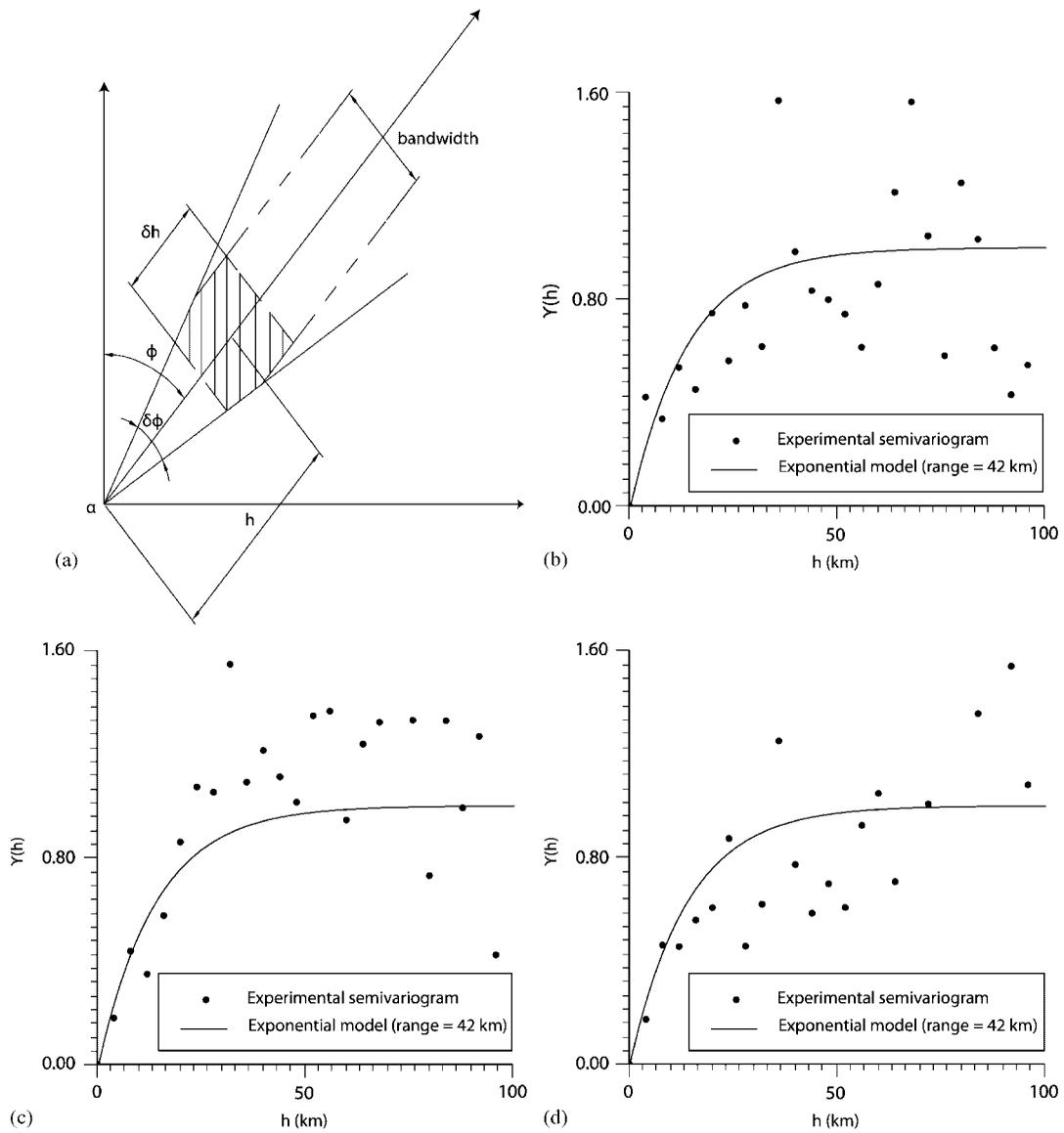


Figure 7. (a) Parameters of a directional semivariogram. Subfigures (b), (c) and (d) show experimental directional semivariograms at discrete separations obtained using the Northridge earthquake $\hat{\epsilon}$ values computed at 2 s. Also shown in the figures is the best fit to the omni-directional semivariogram: (b) Azimuth = 0°; (c) Azimuth = 45°; (d) Azimuth = 90°.

corresponding ranges estimated in this work (Figure 6(a)). It can be seen that the range reported by Wang and Takada [10] is substantially higher than the ranges observed in the current work.

In order to explain this inconsistency, the correlations computed by Wang and Takada [10] are recomputed in the current work using the Chi-Chi earthquake time histories available in the NGA database and the ground-motion model of Annaka *et al.* [25]. The Annaka *et al.* [25] ground-motion model does not explicitly capture the effect of local-site conditions. To account for the local-site effects, Wang and Takada [10] amplified the predicted PGV at all sites by a factor of 2.0 and the same amplification is carried out here for consistency. The observed and the predicted PGVs are used to compute residuals, and the experimental semivariograms (at discrete separations) of these residuals are estimated (considering a bin size of 4 km) using the procedures discussed previously in this report. Figure 8(a) shows the experimental semivariogram obtained, along with an exponential semivariogram function having a unit sill and a range of 83.4 km (there are slight differences between this experimental semivariogram and the one shown in Wang and Takada [10] possibly due to the differences in processing carried out on the raw data or the specific recordings used). It is clear from Figure 8(a), as well as the results presented in [10], that the exponential model with a range of 83.4 km does not provide an accurate fit to the experimental semivariogram values at small separation distances. This is because Wang and Takada [10] minimized the fitting error over all distances to obtain their model.

In the literature, several research works use the method of least squares (or visual methods that attempt to minimize the fitting error over all distances, which in effect, produce fits similar to the least-squares fit), to fit a model to an experimental semivariogram [9, 10, 26]. There are three major drawbacks in using the method of least squares to fit an experimental semivariogram:

- (a) As explained in Section 3.2, it is more important to model the semivariogram structure well at short separation distances than at long separation distances. This is because of the low correlation between intensities at well-separated sites and the shielding of a far-away site by more closely located sites [13]. It is, therefore, inefficient if a fit is obtained by assigning equal weights to the data points at all separation distances, as done in the method of least squares.
- (b) The results provided by the method of least squares are highly sensitive to the presence of outliers (because differences between the observed and predicted $\gamma(h)$'s are squared, any observed $\gamma(h)$ lying away from the general trend will have a disproportionate influence on the fit).
- (c) The least-squares fit results can be sensitive to the maximum separation distance considered. This is of particular significance if the method of least squares is used to determine the sill of the semivariogram in addition to its range.

Some of these drawbacks can be corrected within the framework of the least-squares method. Drawback (a) can be partly overcome by assigning large weights to the data points at short separation distances. The presence of outliers can be checked rigorously using standard statistical techniques [27] and the least-squares fit can be obtained after eliminating the outliers in order to overcome the second drawback mentioned above. These procedures, however, add to the complexity of the approach. For this reason, experimental semivariograms are fitted manually rather than using the method of least squares in the current work (as recommended by Deutsch and Journel [12]). This approach allows one to overlook outliers and also to focus on the semivariogram model at distances that are of practical interest. Though this method is more subjective than the method of least squares, experience shows that the results obtained are reasonably robust.

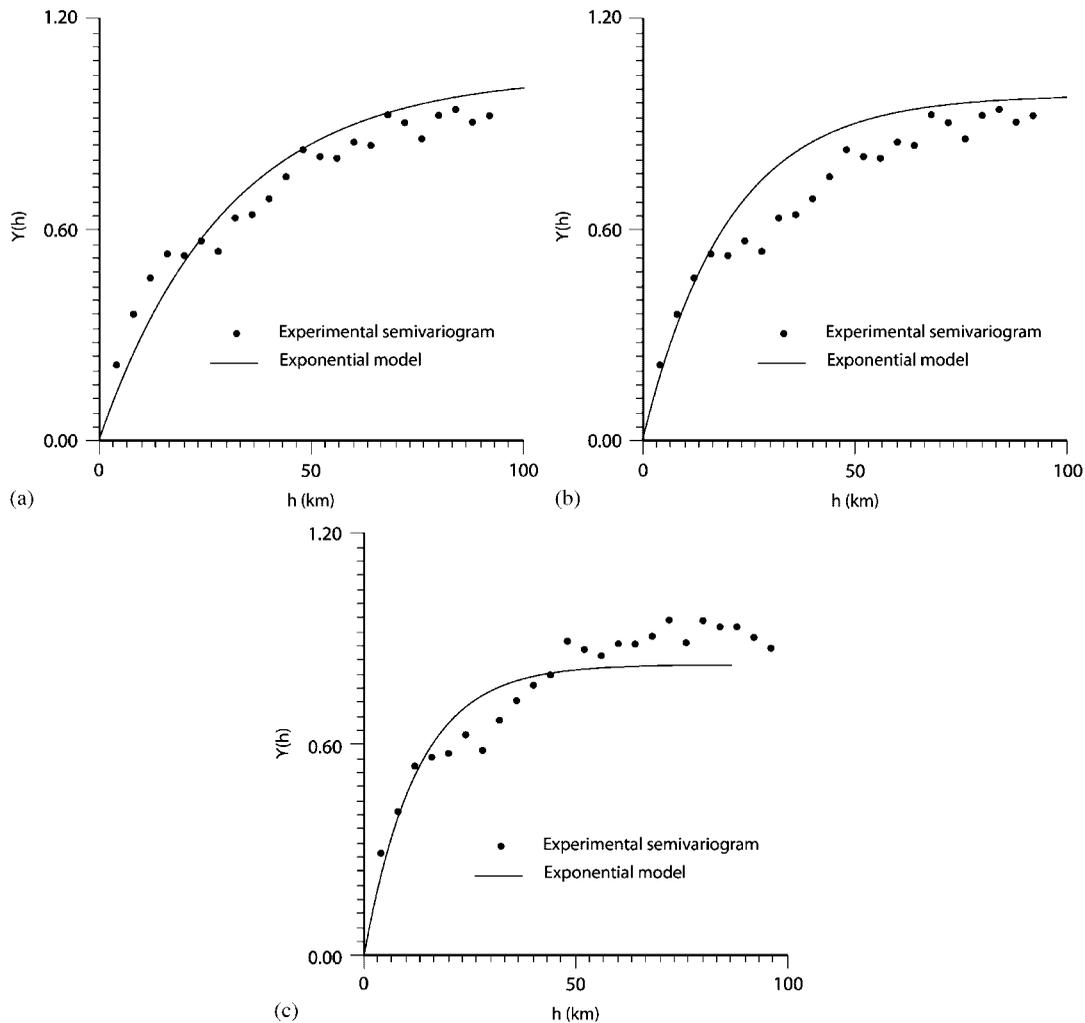


Figure 8. Semivariogram obtained using residuals computed based on Chi-Chi earthquake peak ground velocities: (a) residuals from [25] and semivariogram model from [10]; (b) residuals from [25] and semivariogram fitted to model the discrete values well at short separation distances and (c) residuals from [25], considering random amplification factors.

Figure 8(b) shows the experimental semivariogram (identical to the one shown in Figure 8(a)) along with an exponential function, which is manually fitted to model the experimental semivariogram values well at short separation distances. The range of this exponential model equals 55 km, which is much less than the range of 83.4 km mentioned earlier, and is closer to the results reported earlier for the Chi-Chi spectral accelerations.

The large range reported in Wang and Takada [10] may also be due to inaccuracies in modeling the local-site effects. As explained in Section 3.2, errors in capturing the local-site effects will cause systematic errors in the predicted ground motions that will result in an increase in the

range of the semivariogram. Using a constant amplification factor of 2.0 (without considering the actual local-site effects) will produce even larger systematic errors in the predicted ground motions than considered previously. Consider a complementary hypothetical example in which the ground-motion amplification factor for each site is considered to be an independent random variable, uniformly distributed between 1.0 and 2.0. Randomizing the ground-motion amplification will break up the correlation between the prediction errors in a cluster of closely spaced sites. The semivariogram of residuals obtained by considering such random amplification factors is shown in Figure 8(c). The range of this semivariogram equals 43 km, which is less than the 55 km from Figure 8(b). The true amplifications are neither constant at 2.0, nor are totally random between 1.0 and 2.0. Hence, the range of the semivariogram is expected to lie within 43 km and 55 km, which is close to the range observed using short period spectral accelerations in the current work.

Boore *et al.* [11] estimated correlations between residuals computed from the Northridge earthquake PGAs. They observed that the correlations dropped to zero when the inter-site separation distance was approximately 10 km. This matches with the range of 10 km estimated in the current work using the Northridge earthquake PGAs (Figure 2(a)). Those results appear to be consistent with the results shown here (and it is interesting to note that the two efforts used different estimation procedures and data sets).

The observations in the current work are also consistent with those reported in Goda and Hong [9] who reported a more rapid decrease in correlations with distance for the Northridge earthquake ground motions than for the Chi-Chi earthquake ground motions. They also reported that the decay of spatial correlation of the residuals computed from spectral accelerations is more gradual at longer periods, a feature observed and analyzed in the current research work. The current work adds plausible physical explanations for these empirically observed trends.

6. CONCLUSIONS

Geostatistical tools have been used to quantify the correlation between spatially distributed ground-motion intensities. The correlation is known to decrease with increase in the separation between the sites, and this correlation structure can be modeled using semivariograms. A semivariogram is a measure of the average dissimilarity between the data, whose functional form, sill and range uniquely identify the ground-motion correlation as a function of separation distance.

Ground motions observed during the Northridge, Chi-Chi, Big Bear City, Parkfield, Alum Rock, Anza and Chino Hills earthquakes were used to compute the correlations between spatially distributed spectral accelerations, at various spectral periods. The correlations were computed for normalized intra-event residuals, since the normalized intra-event residuals will be homoscedastic. The ground-motion model of Boore and Atkinson [1] was used for the computations, but the results did not change when the Chiou and Youngs [3] model was used instead.

It was seen that the rate of decay of the correlation with separation typically decreases with increasing spectral period. It was reasoned that this could be because long period ground motions at two different sites tend to be more coherent than short period ground motions, on account of lesser wave scattering during propagation. It was also observed that, at periods longer than 2 s, the estimated correlations were similar for all the earthquake ground motions considered. At shorter periods, however, the correlations were found to be related to the site V_s30 values. It was shown that the clustering of site V_s30 's is likely to result in larger correlations between residuals. Based on these findings, a predictive model was developed that can be used to select appropriate correlation

estimates for use in risk assessment of spatially distributed building portfolios or infrastructure systems.

The research work also investigates the effect of directivity on the correlations using pulse-like ground motions. The correlations obtained were similar to those estimated using all ground motions. The results, however, are not discussed in detail due to concerns about the reliability of the results on account of the small data set of pulse-like ground motions. The work also investigated the commonly used assumption of isotropy in the correlation between residuals using directional semivariograms. If directional semivariograms computed based on different azimuths are identical to the omni-directional semivariogram (which is obtained assuming isotropy), it can be concluded that the semivariograms (and therefore, the correlations) are isotropic. It was seen using empirical data that the correlation between Chi-Chi and Northridge earthquake intensities show isotropy at both short and long periods.

The results obtained were also compared with those reported in the literature [9–11]. Wang and Takada [10] report larger correlations using the PGVs computed using the Chi-Chi earthquake recordings than those reported in this work for spectral accelerations. It was shown that these larger correlations are a result of attempting to fit the experimental semivariogram reasonably well over the entire range of separation distances of interest (which is a typical result of using least-squares fits and eye-ball fits that produce results similar to least-squares fits), and of using a ground-motion model that does not account for the effect of local-site conditions. Typically, a semivariogram model should represent correlations accurately at small separations since ground motions at a site are more influenced by ground motions at nearby sites. The method of least squares assigns equal importance to all separation distances and is therefore, inefficient. In the current research work, semivariogram models are fitted manually with emphasis on accurately modeling correlations at small separations.

This study illustrates various factors that affect the spatial correlation between ground-motion intensities, and provides a basis to choose an appropriate model using empirical data. The proposed predictive model can be used for obtaining the joint distribution of spatially distributed ground-motion intensities, which is necessary for a variety of seismic hazard calculations.

ACKNOWLEDGEMENTS

The authors thank the two anonymous reviewers for their helpful reviews of the manuscript. The authors also thank Paolo Bazzurro and Jaesung Park from AIR Worldwide Co. for useful discussions on this research topic. This work was supported by the U.S. Geological Survey (USGS) via External Research Program award 07HQGR0031. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the USGS.

REFERENCES

1. Boore DM, Atkinson GM. Ground-motion prediction equations for the average horizontal component of PGA, PGV and 5% damped SA at spectral periods between 0.1 s and 10.0 s. *Earthquake Spectra* 2008; **24**(1):99–138.
2. Abrahamson NA, Silva WJ. Summary of the Abrahamson & Silva NGA ground-motion relations. *Earthquake Spectra* 2008; **24**(1):99–138.
3. Chiou BS-J, Youngs RR. An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra* 2008; **24**(1):173–215.
4. Campbell KW, Bozorgnia Y. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.1 to 10 s. *Earthquake Spectra* 2008; **24**(1):139–171.

5. Lee R, Kiremidjian AS. Uncertainty and correlation for loss assessment of spatially distributed systems. *Earthquake Spectra* 2007; **23**(4):743–770.
6. Bazzurro P, Luco N. Effects of different sources of uncertainty and correlation on earthquake-generated losses. *Presented at IFED: International Forum on Engineering Decision Making*, Stoos, Switzerland, 2004.
7. Park J, Bazzurro P, Baker JW. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *Tenth International Conference on Application of Statistic and Probability in Civil Engineering (ICASP10)*, Tokyo, Japan, 2007.
8. Jayaram N, Baker JW. Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of the Seismological Society of America* 2008; **98**(5):2231–2243.
9. Goda K, Hong HP. Spatial correlation of peak ground motions and response spectra. *Bulletin of the Seismological Society of America* 2008; **98**(1):354–365.
10. Wang M, Takada T. Macrospatial correlation model of seismic ground motions. *Earthquake Spectra* 2005; **21**(4):1137–1156.
11. Boore DM, Gibbs JF, Joyner WB, Tinsley JC, Ponti DJ. 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* 2003; **93**(6):2737–2751.
12. Deutsch CV, Journel AG. *Geostatistical Software Library and User's Guide*. Oxford University Press: Oxford, New York, 1998.
13. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press: Oxford, New York, 1997.
14. Chiou B, Darragh R, Gregor N, Silva W. NGA project strong-motion database. *Earthquake Spectra* 2008; **24**(1):23–44.
15. Baker JW, Jayaram N. Effects of spatial correlation of ground-motion parameters for multi-site risk assessment: collaborative research with Stanford University and AIR. *Technical Report*, Report for U.S. Geological Survey National Earthquake Hazards Reduction Program (NEHRP) External Research Program Awards 07HQGR0031, 2009. Available from: <http://earthquake.usgs.gov/research/external/reports/07HQGR0031.pdf>.
16. Bazzurro P, Park J, Tothong P, Jayaram N. Effects of spatial correlation of ground-motion parameters for multi-site risk assessment: collaborative research with Stanford University and AIR. *Technical Report*, Report for U.S. Geological Survey National Earthquake Hazards Reduction Program (NEHRP) External Research Program Awards 07HQGR0032, 2009. Available from: <http://earthquake.usgs.gov/research/external/reports/07HQGR0032.pdf>.
17. Der Kiureghian A. A coherency model for spatially varying ground motions. *Earthquake Engineering and Structural Dynamics* 1996; **25**:99–111.
18. Zerva A, Zervas V. Spatial variation of seismic ground motions. *Applied Mechanics Reviews* 2002; **55**(3):271–297.
19. Castellaro S, Mulargia F, Rossi PL. Vs30: proxy for seismic amplification. *Seismological Research Letters* 2008; **79**(4):540–543.
20. Borchardt RD. Estimates of site-dependent response spectra for design (methodology and justification). *Earthquake Spectra* 1994; **10**:617–653.
21. Baker JW. Quantitative classification of near-fault ground motions using wavelet analysis. *Bulletin of the Seismological Society of America* 2007; **97**(5):1486–1501.
22. CESMD database. Available from: <http://www.strongmotioncenter.org> (last accessed 6 June 2008), 2008.
23. Global Vs30 map server. Available from: <http://earthquake.usgs.gov/research/hazmaps/interactive/vs30/> (last accessed 21 September 2008), 2008.
24. KiK Net. Available from: <http://www.kik.bosai.go.jp/> (last accessed 12 December 2007), 2007.
25. Annaka T, Yamazaki F, Katahira F. Proposal of peak ground velocity and response spectra based on JMA 87 type accelerometer records. *Proceedings, 27th JSCE Earthquake Engineering Symposium*, Tokyo, Japan, vol. 1, 1997; 161–164.
26. Hayashi T, Fukushima S, Yashiro H. Effects of the spatial correlation in ground motion on the seismic risk of portfolio of buildings. *First European Conference on Earthquake Engineering and Seismology*, Geneva, Switzerland, 2006.
27. Kutner MH, Nachtsheim CJ, Neter J, Li W. *Applied Linear Statistical Models*. The McGraw-Hill Companies Inc.: New York, 2005.