FINAL TECHNICAL REPORT

AWARD NUMBER: 07HQGR0031 / 07HQGR0032

EFFECTS OF SPATIAL CORRELATION OF GROUND MOTION PARAMETERS FOR MULTI-SITE SEISMIC RISK ASSESSMENT: COLLABORATIVE RESEARCH WITH STANFORD UNIVERSITY AND AIR

PI: Jack Baker
Report co-author: Nirmal Jayaram
Dept. of Civil & Environmental Engineering
Yang & Yamasaki Environment & Energy Building
473 Via Ortega, Room 283
Stanford CA 94305-4020
650-725-2573 (phone)
650-723-7514 (fax)
bakerjw@stanford.edu

in collaboration with
Dr. Paolo Bazzurro, Dr. Jaesung Park and Dr. Polsak Tothong (AIR Worldwide)

December 2008

Research supported by the U. S. Geological Survey (USGS), Department of Interior, under USGS award number 07HQGR0031. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U. S. Government.

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, two 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); and b) the seismic waves travel over a similar path from source to site. Modern ground motion attenuation equations implicitly recognize the first cause of dependency via a specific inter-event error term. 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, have not yet been investigated.

In this report, 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 to previously published results, and apparent discrepancies among the models are explained. The analysis shows that the spatial correlation reduces with increase in separation between the sites of interest. The rate of decay of correlation typically decreases with increasing spectral acceleration period. At periods longer than 2 seconds, 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 Vs30 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 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. The research work investigates the effect of directivity on the correlations using pulse-like ground motions. It is seen that the correlations between intensities of pulse-like ground motions are slightly larger than correlations between ground motions without pulses. If ground-motion models that account for directivity effects accurately are developed, the correlations between near-fault ground motion intensities can be expected to similar to correlations between ground-motion intensities at other sites.

This report summarizes findings from project activities related to USGS award number 07HQGR0031. A companion final technical report for USGS award number 07HQGR0032 details related findings from the collaborative team members for this project.

Contents

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., Boore and Atkinson, 2008; Abrahamson and Silva, 2008; Chiou and Youngs, 2008; Campbell and Bozorgnia, 2008). 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 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., Lee and Kiremidjian, 2007; Bazzurro and Luco, 2004). 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. (2007) 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:

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

where Y_{ij} denotes the ground-motion parameter of interest (e.g., $S_a(T_1)$, the spectral acceleration at period T_1); $\bar{Y}_{ij}$ 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 part of the ground-motion model and are a function of the spectral period of interest, and in some models also a function of the earthquake magnitude and the distance of the site from the rupture. The intra-event residuals (ε_{ij}) can be normalized using σ_{ij} to obtain normalized intra-event residuals $(\grave{\varepsilon}_{ij})$. During an earthquake, the inter-event residual (η_j) computed at any particular period is a constant across all the sites.

Jayaram and Baker (2008) showed that a vector of spatially-distributed intra-event residuals $\boldsymbol{\varepsilon_j} =$

$(\varepsilon_{1j}, \varepsilon_{2j}, \cdots, \varepsilon_{dj})$ 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 $\varepsilon_{i_1 j}$ and $\varepsilon_{i_2 j}$ 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 (Goda and Hong, 2008; Wang and Takada, 2005; Boore et al., 2003). Boore et al. (2003) used observations of peak ground acceleration (PGA, which equals $S_a(0)$) from the 1994 Northridge earthquake to compute the spatial correlations. Wang and Takada (2005) computed the correlations using observations of peak ground velocities (PGV) from several earthquakes in Japan and the 1999 Chi-Chi earthquake. Goda and Hong (2008) 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 seconds and 3 seconds. 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. (2003) 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 (2005) extend past 100 km. Further, Goda and Hong (2008) observe 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) (Deutsch and Journel, 1998; Goovaerts, 1997). 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 $Z = \left(Z_{u_1}, Z_{u_2}, \cdots, Z_{u_d}\right)$ denote a spatially-distributed random function, where u_i denotes the location of 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 (Goovaerts, 1997). Let u and u' denote two sites separated by $\mathbf{h}$. The semivariogram $(\gamma(u,u'))$ is computed as half the expected squared difference between Z_u and $Z_{u'}$.

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

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:

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

Equation 2 can be replaced with equation 3 if the random function (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:

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

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 semi-

variogram values for any possible separation **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 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 **h** is replaced by a scalar separation length $\|\mathbf{h}\|$, also denoted as h), is expressed as follows:

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

where a and b are the sill and the range of the semivariogram function respectively (Figure 1). 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:

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

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

The Spherical model is as follows:

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

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:

$$\gamma(h) = a\left[I\left(h > 0\right)\right] \tag{8}$$

where $I\left(h > 0\right)$ 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 (Goovaerts, 1997):

$$\gamma(h) = a\left(1 - \rho\left(h\right)\right) \tag{9}$$

where $\rho\left(h\right)$ denotes the correlation coefficient between Z_u and Z_{u+h}. It can also be easily 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. Second-order stationarity is assumed in the development of most correlation models, and is, in fact, assumed even in the ground-motion models. In the current work, like many other works involving to spatial-correlation estimation, the semivariograms are also assumed to be isotropic. The assumption of isotropy is investigated below, and shown to be reasonable.

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,

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

Let $\grave{\varepsilon}_i$ 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.) $\grave{\varepsilon}_i$ is computed as follows:

$$\grave{\varepsilon}_i = \frac{\varepsilon_i}{\sigma_i} \tag{11}$$

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:

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

While assessing covariances, it is convenient to work with $\grave{\varepsilon}$'s rather than ε's, since $\grave{\varepsilon}$'s are homoscedastic (i.e., constant variance) with unit variance unlike the ε's. If required, the semivariogram function of $\grave{\varepsilon}$ can be easily extended to obtain the semivariogram function of ε.

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 $\grave{\varepsilon}$ can be obtained as follows (based on equation 4):

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

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 $\grave{\varepsilon}$ (i.e., the sill of $\hat{\gamma}(h)$) should equal 1 since the $\grave{\varepsilon}$'s have a unit variance. Hence, based on equation 9, it can be concluded that:

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

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 $\frac{\eta}{\sigma_{u_\alpha}} = \frac{\eta}{\sigma_{u_\alpha+h}}$. This approximation, vanishes, however, when the standard deviation of the intra-event residuals depends only on the period at which the residuals are computed. This is the case with the ground-motion model of Boore and Atkinson (2008), which is used in the current work. It is, however, to be noted that the Boore and Atkinson (2008) ground-motion model does not specify the variance of residuals (inter-event and intra-event residuals) computed in arbitrary directions, but rather, specifies the variance of the orientation-independent residuals (Boore and Atkinson, 2008; Boore et al., 2006). Hence, slight approximations are involved in computing the normalized intra-event residuals using the Boore and Atkinson (2008) model. The ground motion databases typically report recordings in two orthogonal horizontal directions. For instance, the PEER NGA Database (2005) 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 correla-

tions 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. Incidentally, though this work only uses the Boore and Atkinson (2008) ground-motion model, the results obtained were found to be similar when an alternate model, namely, the Chiou and Youngs (2008) model, was used.

3.1 Construction of experimental semivariograms using empirical data

Figure 1 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 $\in \left(h - \frac{\delta h}{2}, h + \frac{\delta h}{2}\right)$ 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 result in approximate experimental semivariograms. In the current work, experimental semivariograms are obtained using δh = 2 km (unless stated otherwise), since this was seen to be the smallest w that results in a reasonable number of site pairs in the bins.

The semivariogram shown in Figure 1 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):

$$\gamma(h) = 1 - exp\,(-3h/40) \tag{15}$$

The correlation function corresponding to this model equals $1 - \gamma(h) = exp\,(-3h/40)$ (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:

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

where $\gamma_0(\mathbf{h})$ is a pure nugget effect (which takes a value of 1 at $\|\mathbf{h}\| > 0$ and 0 otherwise); $\gamma_n(\mathbf{h})$ is either a spherical, an exponential or a Gaussian model (which can be isotropic or anisotropic, but is assumed to be isotropic in most cases in the current work); 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 form, the range and the contribution of the models can be systematically varied to obtain the best

fit to the experimental semivariogram values. In the literature, many research works determine a fit that minimizes the fitting error over the entire range of separation distances considered (i.e., all considered values of h) (e.g., Goda and Hong, 2008; Wang and Takada, 2005; Boore et al., 2003). This is typically done using the method of least squares or simply visually (Sometimes these visual/ eye-ball fits are obtained by visually minimizing the fitting error over a wide range of separations. These fits are similar to least-squares fits and are not addressed separately.) Fitting the experimental semivariogram values manually has a few advantages over the method of least squares and these are explained in the next paragraph and in detail in section 6.

One reason for preferring manual fitting to least-squares fitting is that it is more important to model the semivariogram structure well at short separation distances than at long separation distances. This is because the data at sites far away from i hardly influence the data values at site i on account of the low correlations between data at well-separated sites. In addition to this already low correlation, the impact of these widely separated sites is further reduced by an effective 'shielding' of their influence by more closely-located sites (Goovaerts, 1997). Hence, while obtaining the fit manually, emphasis is placed on modeling the experimental semivariogram values accurately at small separations (as compared to modeling the values accurately at large separations). Figure 2 shows sample semivariograms fitted to a data set using the 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 discussion on the advantages of using manual-fitting rather than least-squares fitting follows in section 6.

3.2 1994 Northridge earthquake recordings

This section discusses the ranges of semivariograms estimated using observed Northridge earthquake ground motions. As mentioned earlier, the sill of $\hat{\gamma}(h)$ should equal 1, though in practice, the sill does not exactly equal 1 due to limited size of the sample used (In other words, the variance of the observed residuals does not equal the variance predicted by the ground-motion model.) Hence, the data values ($\tilde{\varepsilon}$'s, which are used in the construction of $\hat{\gamma}(h)$) are scaled so that the sill of the semivariograms obtained will equal 1 (Scaling involves dividing the $\tilde{\varepsilon}$ values by the variance of the $\tilde{\varepsilon}$ sample.) It is to be noted that the computed correlations will not be affected by this scaling of the $\tilde{\varepsilon}$ values. In the rest of the report, $\tilde{\varepsilon}$ refers to this scaled sample that has a unit sample variance.

Figure 3 shows a semivariogram obtained using the $\tilde{\varepsilon}$ values computed at a period of 2 seconds. The semivariogram shown in this figure has an exponential form with a sill of 1 and a range of 38 km.

This model can be expressed as follows:

$$\gamma(h) = 1 - exp\,(-3h/38) \tag{17}$$

Of three models considered, namely, the exponential, the Gaussian and the spherical models, the exponential model is found to provide the 'best fit' (particularly at small separations). In fact, the exponential model seems to provide the best fit for experimental semivariograms obtained using $\tilde{\varepsilon}$'s computed at several different periods, based on recordings from different earthquakes.

The manual fitting approach described previously is used to compute ranges of the semivariograms of $\tilde{\varepsilon}$'s (obtained based on the Northridge earthquake time histories) computed at seven periods ranging between 0 seconds and 10 seconds. Though long period ground motions (i.e., periods close to 10 seconds) are not frequently used in practice, they are considered in this work in order to cover the entire range of periods for which the considered ground-motion models are defined. The semivariograms obtained are shown in Figures 3-9 and a plot of the estimated ranges against period is shown in Figure 10. The following can be observed from these figures:
(a) The estimated range of the semivariogram increases with period.
(b) The exponential model seems to result in a good fit in all the cases. The constancy of the semivariogram function across periods makes it simpler to specify a standard correlation model for the $\tilde{\varepsilon}$'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.

As mentioned earlier, a larger range indicates a smaller rate of decay of correlation with separation distance. Hence, it can be inferred from the above observations that the $\tilde{\varepsilon}$ values computed at long periods show larger correlations than those computed at short periods. This behavior can be explained using the concept of coherency, which has been widely researched in the past. Coherency can be thought of as a measure of 'similarity' in the ground motions. Der Kiureghian (1996) reports that the coherency between earthquake ground motions at two different sites is reduced by the scattering of waves during propagation, and that this reduction is more for high frequency waves. This is because the high-frequency waves with short wavelengths tend to be more affected by the heterogeneities of the propagation path, and as a result tend to be less coherent than long period ground motions (Zerva and Zervas (2002)). It is reasonable to expect highly coherent ground motions to exhibit similar peak amplitudes as well. The residuals, which are computed using these amplitudes (spectral accelerations), therefore, tend to show larger correlations at long periods, which is reflected by an increase in the range of semivariograms at long periods.

As mentioned earlier, a larger range indicates a smaller rate of decay of correlation with sepa-

ration distance. Hence, it can be inferred that the $\tilde{\varepsilon}$ values computed at long periods show larger correlations than those computed at short periods. This behavior can be explained using the concept of coherency, which has been widely researched in the past. Coherency can be though of as a measure of 'similarity' in the ground motions. Der Kiureghian (1996) reports that the coherency between earthquake ground motions at two different sites is reduced by the scattering of waves in the ground during propagation, and that this reduction is more for high frequency waves. This is because the high-frequency waves with short wavelengths tend to be more affected by the heterogeneities of the propagation path and as a result tend to be less coherent than long period ground motions (Zerva and Zervas (2002)). It is reasonable to expect highly coherent ground motions to exhibit similar peak amplitudes as well. The residuals, which are computed using these amplitudes (spectral accelerations), therefore, tend to show larger correlations at long periods, which is reflected by an increase in the range of semivariograms at long periods.

The Northridge earthquake data used for the above analysis are obtained from the PEER NGA Database (2005). 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 (2008) ground-motion model are considered. For the purposes of this report, these records are denoted 'usable records'. The semivariograms of residuals computed at periods of 5, 7.5 and 10 seconds, however, are obtained using all available records in the PEER NGA Database (2005). This is on account of the limited number of Northridge earthquake recordings at extremely long periods. At 5 seconds, 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, and shown in Figure 7) can be compared to the approximate semivariogram 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, and shown in Figure 11). The ranges of the two semivariograms, SV_1 and SV_2, are 40 km 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. Hence, the advantage of obtaining a stable semivariogram using all records is accompanied by a small approximation in the semivariogram range estimated.

As mentioned in section 1, correlation between intensities estimated using the fault-normal components are discussed in this report. 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 $\tilde{\varepsilon}$'s computed at 2 seconds, 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 (Figure 12). The corresponding range for the semivariogram based on the fault-normal ground motions equals 42 km (Figure 3). 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 ranges of semivariograms of $\tilde{\varepsilon}$'s computed at different periods based on the Chi-Chi earthquake recordings are discussed. Figure 13 shows the experimental semivariogram values at discrete separation distances, obtained using the $\tilde{\varepsilon}$ values computed at 2 seconds. The most accurate model for the semivariogram function is a combination of a nugget effect with a contribution of 0.3 and an exponential semivariogram with a contribution of 0.7 and a range of 85 km, which is also shown in Figure 13. This model can be expressed as follows:

$$\gamma(h) = 0.3I(h > 0) + 0.7\,(1 - exp\,(-3h/85)) \tag{18}$$

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

As mentioned earlier, however, the use of a single model for all semivariograms is highly desirable in order to facilitate development of a standard correlation model for use in future predictions. Moreover, it is seen that the exponential model proves to be accurate in most cases and hence, it seems reasonable to fit an approximate exponential model even in cases where alternate accurate models are available. Hence, the semivariogram function for the $\tilde{\varepsilon}$ values computed at 2 seconds is approximated by an exponential model with a range of 36 km and a sill of 1, as shown in Figure 13. This semivariogram function fits the data reasonably well at small separations.

Figures 14 - 20 show semivariograms of $\tilde{\varepsilon}$'s computed at seven different periods based on the Chi-Chi earthquake ground-motion recordings. A summary plot of the estimated range for each period is shown in Figure 21. The following can be observed from the figures:
(a) As was seen with the Northridge earthquake data, the range of the semivariogram typically increases with period (with an exception when the peak ground accelerations (PGA) are considered).
(b) The ranges are higher, in general, than those observed based on the Northridge earthquake data (Figure 10). This is also consistent with observations made by other researchers considering Northridge and Chi-Chi earthquake data (e.g., Goda and Hong, 2008).

The relatively large ranges obtained using the Chi-Chi earthquake residuals as compared to those obtained based on the Northridge earthquake residuals 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. Errors can arise in the predicted ground-motion intensities if inaccurate V_s30 values are used 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's are measured values (PEER NGA Database, 2005). The V_s30 values inferred from the Geomatrix site class are likely not exact. Further, since closely-spaced sites are likely to belong to the same site class and posess 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. 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. (2008) 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 (Borcherdy, 1994)) 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 seconds) than at longer periods. This implies that ground-motion intensity predictions based on V_s30 will have errors, particularly at periods below 0.5 seconds.

Figures 22 and 23 show the V_s30 values at the recording stations of the Northridge earthquake and the Chi-Chi earthquake respectively. A visual inspection of the figures show that the V_s30 values at the Chi-Chi earthquake recording stations show much larger spatial correlations than those at the Northridge earthquake recording stations (in other words, V_s30 values at the closely-spaced Chi-Chi earthquake recording stations are similar). This is confirmed by Figures 24 and 25, which show semivariograms of the normalized V_s30 values (the V_s30 semivariogram is not to be confused with the $\tilde{\varepsilon}$ 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 24 shows significant scatter at all separation distances indicating zero correlation at all separations. In contrast, Figure 25 indicates that the Taiwan V_s30 values have significant spatial correlation. In other words, the V_s30 values at closely-spaced Chi-Chi earthquake recording stations show much larger

correlations than the V_s30 values at closely-spaced Northridge earthquake recording stations. Hence, based on the argument in the previous paragraph, larger predictions errors can be expected at the Chi-Chi earthquake recording stations than at the Northridge earthquake recording stations. Such correlated errors explain the larger observed ranges when the residuals are computed using the Chi-Chi earthquake ground motions.

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 possible reasons for the larger ranges estimated in section 3.3 than in section 3.2. There could be other factors such as the rupture area that can affect the correlations. These factors could not, however, be investigated with the limited data set available.

Further, as mentioned previously, one notable aberration in the plot of range versus period (Figure 21) is the large range observed when the residuals are computed at 0 seconds as compared to some of the longer periods. This does not fall in line with the coherency argument of the previous section. This can, however, be explained using the relationship between the range and the V_s30's described in the above paragraphs. As mentioned earlier, in the presence of clustered (correlated) V_s30's (which is indicative of soil homogeneity), 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 seconds) (Castellaro et al., 2008), which explains the larger correlation between the residuals (which ultimately results in the larger range observed) computed using PGAs.

3.4 Simulated ground-motion data

The correlations between residuals computed based on broadband ground-motion simulations for scenario earthquakes on the Puente Hills thrust fault system (Graves, 2006) are discussed in this section. The simulated time histories are available for five different rupture scenarios that differ in the rupture velocity and the rise time. In this work, ground motions due to the rupture scenario defined by a rupture velocity equaling 80% of the shear wave velocity and a rise time of 1.4 seconds are used for the analysis. The ground-motion time histories have been simulated at 648 sites covering the Los Angeles, San Fernando and San Gabriel basin regions. The time histories at locations with very low V_s30 values, however, were reported to be possibly inaccurate because the simulation algorithm does not yet fully account for non-linear site effects (Graves, 2007). Hence, in the current work, only the time histories at sites with V_s30 values exceeding 300m/s are considered for analysis.

Experimental semivariograms are obtained for $\tilde{\varepsilon}$'s computed at several different periods ranging from 0 - 10 seconds. The exponential model is found to provide a good fit at periods below 2 seconds. At longer periods, however, a spherical model provides a better fit than an exponential model. For

example, Figure 26 shows the experimental semivariogram and a fitted spherical model (unit sill and range equaling 32 km) based on residuals computed at 5 seconds.

$$\begin{aligned} \gamma(h) &= \frac{3}{2}\frac{r}{32} - \frac{1}{2}\left(\frac{r}{32}\right)^3 \text{ if } h \leq 32 \\ &= 1 \text{ otherwise} \end{aligned} \tag{19}$$

As explained earlier, for consistency with other results, exponential models that provide a reasonably good approximation at short separation distances (that are useful in practice) are used to model the semivariograms. For example, the experimental semivariogram can also be fitted with an exponential model which has a unit sill and a range of 60 km as shown in Figure 26. It can be seen from the figure that this exponential function models the correlations at small separations reasonably accurately.

A plot of the range of semivariograms as a function of period is shown in Figure 27. The trend of increasing range with period is seen in this figure as well. The computed ranges are reasonably similar to those seen from the Northridge earthquake data. It is to be noted, however, that the ground-motion simulations at short periods (periods $\leq$ 2 seconds) may not be entirely accurate, and hence, the ranges obtained using the Northridge and Chi-Chi earthquake data are more reliable estimates.

3.5 Japanese earthquake data

A few research works use the ground-motion recordings from earthquakes in Japan, based on the data provided in the KiK Net (2007). 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 (2007) 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 (2007) were not considered further for studying the ranges of semivariograms.

3.6 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 (2008). The V_s30 data used for these computations came from the CESMD database (2008) (for the Parkfield earthquake) and the U.S. Geological Survey

V_s30 maps (for the other earthquakes).

Exponential models are fitted to experimental semivariograms of $\tilde{\varepsilon}$'s computed using the time histories from the above-mentioned earthquakes, at periods ranging from 0 - 10 seconds. Figures 28-32 show plots of range versus period for the Big Bear City, Parkfield, Alum Rock, Anza and Chino Hills earthquake residuals respectively. The ranges of the semivariograms are 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 figures 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 (Figures 33-37). The semivariogram ranges computed using the Anza earthquake V_s30's was found to be the largest and equaled 40 km, while the ranges computed from the Chino Hills, Big Bear City, Alum Rock and Parkfield earthquake data were smaller at 35 km, 30 km, 18 km 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 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 summarized by plotting the range of PGA-based residuals against the range of the V_s30's, for the earthquakes considered in this work (Figure 38). 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. (2008) that the scatter in the plot of F_a versus V_s30 is more 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. Moreover, at long periods, 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) could obfuscate errors related to V_s30's. 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.

3.7 Discussion

The above sections presented 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 39 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 seconds. At short periods (below 2 seconds), however, there are considerable differences in the estimated ranges depending on the ground-motion time histories used. The previous sections explained these differences using the V_s30 values at the recording stations for these earthquakes. Hence, from a practitioner's perspective, the V_s30 values play an important role in decision making. The following cases can be considered for decision making:

Case 1: If the practitioner is confident about the accuracy of the ground-motion model and the quality of the V_s30 values, the smaller ranges reported in Figure 39 should be chosen. It is to be noted, however, that the local-site effects are not entirely captured by the V_s30's (Castellaro et al., 2008) and hence, predictions based on V_s30's are likely inaccurate. Therefore, it is unlikely that a practitioner would encounter Case 1.

Case 2: If the practitioner lacks confidence in the ground-motion model and/ or the V_s30 values, but the V_s30 values do not show or are not expected to show clustering, the smaller ranges reported in Figure 39 will still be appropriate.

Case 3: If the practitioner lacks confidence in the ground-motion model and/ or the V_s30 values, and the V_s30 values show or are expected to show clustering, the larger ranges reported in Figure 39 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 ($T < 1$ second), for cases 1 and 2:

$$b = 8.5 + 17.2T \tag{20}$$

At short periods ($T < 1$ second), for case 3:

$$b = 40.7 - 15.0T \tag{21}$$

At long periods ($T \geq 1$ second) (for all cases):

$$b = 22.0 + 3.7T \tag{22}$$

where b denotes the range of the exponential semivariogram (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 $\rho(h) = exp(-3h/b)$ (follows from equations 5 and 14). Recall from equation 13 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 40. It is to be noted that the model has been developed based on only seven different earthquakes. Yet, since the trends exhibited by the ranges computed based on the various earthquake data were found to be similar, it can be expected that the model would predict a reasonable range for all practical purposes.

4 Correlation between near-fault ground-motion intensities

Most currently available ground-motion models do not directly predict ground motions containing strong velocity pulses, such as those caused by near-fault directivity. As a result, the ground-motion intensities predicted by the models at sites that experience pulse-like ground motions will be different from the observed values. Such systematic prediction errors can increase the apparent correlation between the residuals computed at these sites. Hence, in this section, empirical data are used to verify whether the correlation between residuals at sites experiencing pulse-like ground motion is significantly different from the correlation between residuals at other sites.

Baker (2007) used wavelet analysis to extract velocity pulses from ground motions and developed a quantitative criterion for classifying a ground motion as pulse-like. Ninety one large-velocity pulses were found in the fault-normal components of the approximately 3500 strong ground-motion recordings in the PEER NGA Database (2005). It should be noted that not all of these pulses may be due to directivity effects, but this provides a reasonable data set for studying the potential impact of directivity. Of these, 30 pulses were found in the fault-normal components of the Chi-Chi earthquake recordings, while the rest of the earthquakes have far fewer recordings with pulses. In the current work, the pulse-like ground motions from the Chi-Chi earthquake are used to compute $\tilde{\varepsilon}$ values at different periods. The semivariograms of the residuals are obtained and compared to those estimated using all usable records (section 3.3).

Figures 41-47 compare experimental semivariograms of residuals (at seven different periods) computed using pulse-like ground motions to experimental semivariograms of residuals computed using all usable ground motions. The figures show the experimental semivariogram values at short separation distances, which are of interest in practice. On account of the fewer available records, it is to be noted that the experimental semivariograms obtained using pulse-like ground motions are less clearly defined than those obtained using all usable ground motions. Hence, it is difficult to fit robust models for the experimental semivariograms obtained using the pulse-like ground motions. As a result, the experimental semivariograms are compared as such, rather than by their models and

ranges.

It can be seen from Figures 41-47 that the experimental semivariogram values obtained using the pulse-like ground motions are slightly less than those obtained using all usable ground motions, particularly at separation distances below 10 km and at long periods (7.5 and 10 seconds). This is consistent with expectations as the pulses from this earthquake typically have periods of approximately 7 seconds and so, it is expected that this is the period range that would be most strongly influenced by directivity. In other words, the $\tilde{\varepsilon}$'s obtained using pulse-like ground motions show slightly larger correlations than those obtained using all usable ground motions. The difference in the correlations is typically around 0.1, with a maximum value of approximately 0.2.

While the increased correlations between the residuals at sites experiencing pulse-like ground motions is expected, the difference in the correlation seems reasonably small. Moreover, it is to be noted that the source of this additional correlation is the systematic prediction errors caused by the ground-motion models at sites experiencing pulse-like ground motions. Hence, if ground-motion models that account for directivity effects accurately are developed, the correlations between near-fault ground-motion intensities can be expected to the similar to the correlation between ground-motion intensities at other sites. That is, the directivity effects are best addressed through refinements to ground-motion models, rather than refinements to correlation models.

5 Isotropy of semivariograms

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

5.1 Isotropy/ Anisotropy 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 (Goovaerts, 1997). Directional semivariograms are obtained as shown in equation 4 except that the estimate is obtained using only pairs of $(z_{u_\alpha}, z_{u_\alpha+\mathbf{h}})$ 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. 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 dif-

fering azimuths have the same model and sill, but different ranges. Zonal anisotropy is indicated by a variation in the sill with direction. In the current work, ground-motion recordings from earthquakes are used to investigate the potential presence of anisotropy in ground-motions.

5.2 Construction of a directional semivariogram

A directional semivariogram is specified by several parameters, as illustrated in Figure 48. 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 ($\delta h \leq h$). 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 $\left[h-\frac{\delta h}{2}, h+\frac{\delta h}{2}\right]$, and with azimuths ranging between $[\phi-\delta\phi, \phi+\delta\phi]$ are considered. For example, let α be a site located in a 2 dimensional region, as shown in Figure 48. 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 value 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 48). 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 (b) 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 report are obtained by considering a bin separation of 4 km, a bin width of 4 km, an azimuth tolerance of 10 $^\circ$ and a bandwidth of 10 km. Typically, directional semivariograms are plotted for azimuths of 0 $^\circ$, 45 $^\circ$ and 90 $^\circ$ in order to capture the effects of anisotropy, if any.

5.3 Test for anisotropy using Northridge and Chi-Chi ground motion data

Figure 49 shows the omni-directional and the three experimental directional semivariograms of the 2 second $\tilde{\varepsilon}$'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, which was reported in section 3.2.

The figure shows that 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 are obtained when the earthquake used and the period at which the residuals are computed are varied. For example, Figures 50 and 51 show the omni-directional and the directional experimental semivariograms of the Chi-Chi earthquake $\tilde{\varepsilon}$'s computed at 1 second and 7.5 seconds. The ranges of the exponential models shown in this figure are identical to those reported in section 3.3 (in which the semivariograms were obtained considering isotropy). The exponential models match well with the directional experimental semivariograms, which is good evidence that the semivariograms of $\tilde{\varepsilon}$ are isotropic.

The directional semivariograms are also obtained using the simulated time histories of Graves (2006). Figure 52 shows the experimental semivariograms of $\tilde{\varepsilon}$'s computed from the simulated time histories at a period of 2 seconds. The exponential model developed in section 3.4 (assuming isotropy) is found to fit the experimental directional semivariograms reasonably well. When the $\tilde{\varepsilon}$ values are computed at long periods (e.g., 7.5 seconds), however, the directional semivariograms are seen to be different from each other. Figure 53 shows the directional semivariograms of residuals computed at 7.5 seconds. It can be seen from the figure that the omni-directional semivariogram model developed in section 3.4 does not fit the experimental directional semivariograms well enough even at short separation distances. This is an indication of anisotropy and hence, a better fit can be obtained only by using an anisotropic semivariogram. Figure 54 shows the four experimental directional semivariograms along with a fitted anisotropic model that has the following form (based on a linear combination of two spherical models):

$$\hat{\gamma}(\mathbf{h}) = 0.5\ Sph\left(\sqrt{\frac{h_1^2}{15^2} + \frac{h_2^2}{120^2}}\right) + 0.5\ Sph\left(\frac{h}{25}\right) \tag{23}$$

where the separation vector $\mathbf{h}$ has components h_1 and h_2; h denotes $|\mathbf{h}|$; and $Sph(r)$ equals $\frac{3}{2}r - \frac{1}{2}r^3$ if $r \leq 1$ and equals 1 otherwise. This model depends not only on the distance between sites (h), but also on the components of $\mathbf{h}$, and thereby on the orientation of the separation vector. As a result, this semivariogram is anisotropic.

As shown previously in this section, however, semivariograms obtained from recorded ground motions are isotropic even at long periods. Hence, in practice, it seems reasonable to assume isotropy while developing semivariograms for the residuals.

6 Comparison with previous research

Researchers, in the past, have computed the correlation between ground-motion intensities using observed peak ground accelerations, peak ground velocities 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 (2005) used the ground-motion relationship of Annaka et al. (1997) to compute the normalized auto-covariance function of residuals computed using the Chi-Chi earthquake peak ground velocities (PGV). They used an exponential model to fit the discrete experimental covariance values and reported a result which is equivalent to the following semivariogram:

$$\gamma(h) = 1 - exp(-3h/83.4). \tag{24}$$

This semivariogram has a unit sill and a range of 83.4 km (from equation 5). It is to be noted that the current work does not consider the spatial correlation between PGV-based residuals. The PGVs, however, are comparable to spectral accelerations computed at short periods, and hence, the semivariogram ranges of residuals computed from PGVs can be compared to the short period ranges estimated in this work (Figure 39).It can be seen that the range reported by Wang and Takada (2005) is substantially higher than the ranges observed in the current work.

In order to explain this inconsistency, the correlations computed by Wang and Takada (2005) are recomputed in the current work using the Chi-Chi earthquake time histories available in the PEER NGA Database (2005) and the ground-motion model of Annaka et al. (1997). The Annaka et al. (1997) ground-motion model does not explicitly capture the effect of local-site conditions. In order to account for the local-site effects, Wang and Takada (2005) 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 55 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 (2005) possibly due to the differences in processing carried out on the raw data or the specific recordings used). It is clear from Figure 55, as well as the results presented in Wang and Takada (2005), 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 one of the drawbacks of using least-squares

fits to an experimental semivariogram.

As mentioned earlier, in the literature, several research works use the method of least squares (or visual methods that produce results similar to those of the least-squares method), to fit a model to an experimental semivariogram (Goda and Hong, 2008; Hayashi et al., 2006; Wang and Takada, 2005). 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 (Goovaerts, 1997). 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 the 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 (Kutner et al., 2005) 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 problem. For this reason, experimental semivariograms are fitted manually rather than using the method of least squares in the current work (Deutsch and Journel, 1998). This approach allows one to overlook outliers and also to focus on the semivariogram model at values of separation ($\mathbf{h}$) that are of practical interest. Though this method is more subjective than the method of least squares, the results obtained are reasonably robust since the correlations are not extremely sensitive to minor changes in the function parameters.

Figure 56 shows the experimental semivariogram (identical to the one shown in Figure 55) 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 (2005) 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. This hypothesis can be illustrated using the complementary case in which the ground-motion amplifications are random, rather than identical. Consider a 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 considering such random amplification factors is shown in Figure 57. The range of this semivariogram equals 43 km, which is less than the 55 km from Figure 56. In practice, the 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. (2003) 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 10). Various possible reasons for observing a smaller range with the Northridge earthquake PGAs as compared to the Chi-Chi earthquake PGAs were provided in section 3.3.

The observations in the current work are also consistent with those reported in Goda and Hong (2008) who reported a more rapid decrease in correlations 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 also provides physical explanations for these empirically-observed trends.

7 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 are not computed for the total intensity, rather, are computed for normalized intra-event residuals, since the normalized intra-event residuals will be homoscedastic. For a given earthquake, the correlation between the residuals can be easily extended to compute the correlation between intensities, if required. The ground-motion model of Boore and Atkinson (2008) was used for the computations, but the results did not change when the Chiou and Youngs (2008) 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 seconds, 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 correlation 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 identified by Baker (2007). It was seen that correlations between intensities of pulse-like ground motions were slightly larger than correlations estimated using all usable ground motions. The source of these additional correlations is the additional prediction error introduced by the ground-motion models at sites experiencing pulse-like ground motions. Hence, if ground-motion models that account for directivity effects accurately are developed, the correlations between near-fault ground-motion intensities can be expected to the similar to the correlation between ground-motion intensities at other sites.

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 considering 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 to those reported in the literature (Goda and Hong, 2008; Wang and Takada, 2005; Boore et al., 2003). Wang and Takada (2005) 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 chosen correlation 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.

8 Acknowledgments

The authors thank Paolo Bazzurro and Jaesung Park from AIR Worldwide Co. for useful discussions on this research topic. This work was partially 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. (Add Brad and Rob to this)

9 Data sources

The data for the ground motions studied here came from:

PEER NGA Database (2005) (http://peer.berkeley.edu/nga - last accessed 18 May 2007)

KiK Net (2007) (http://www.kik.bosai.go.jp - last accessed 12 December 2007)

CESMD database (2008) (http://www.strongmotioncenter.org - last accessed 6 June 2008) (BRAD and ROB)

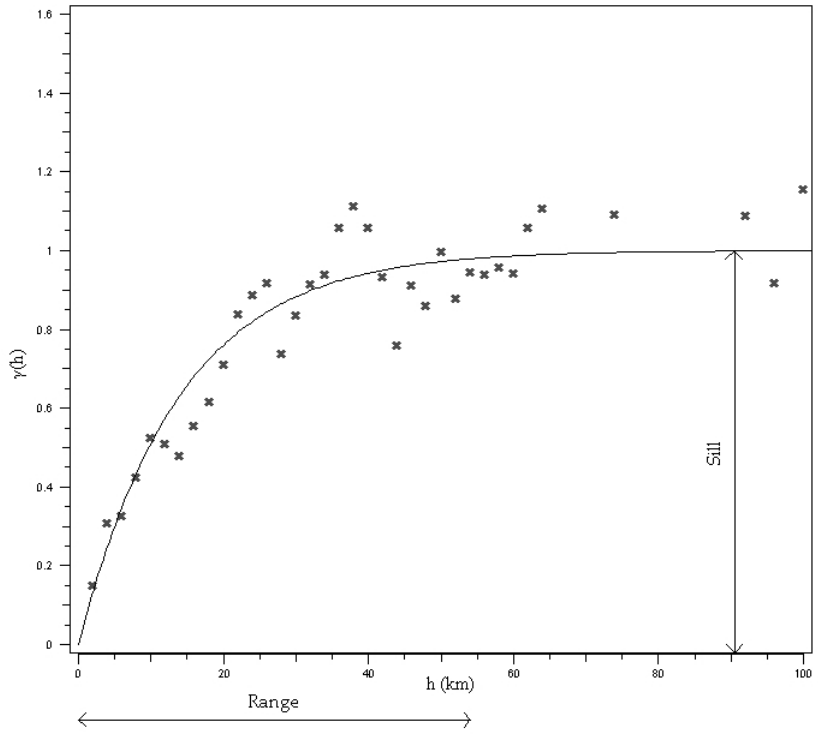

Figure 1: Parameters that describe a semivariogram

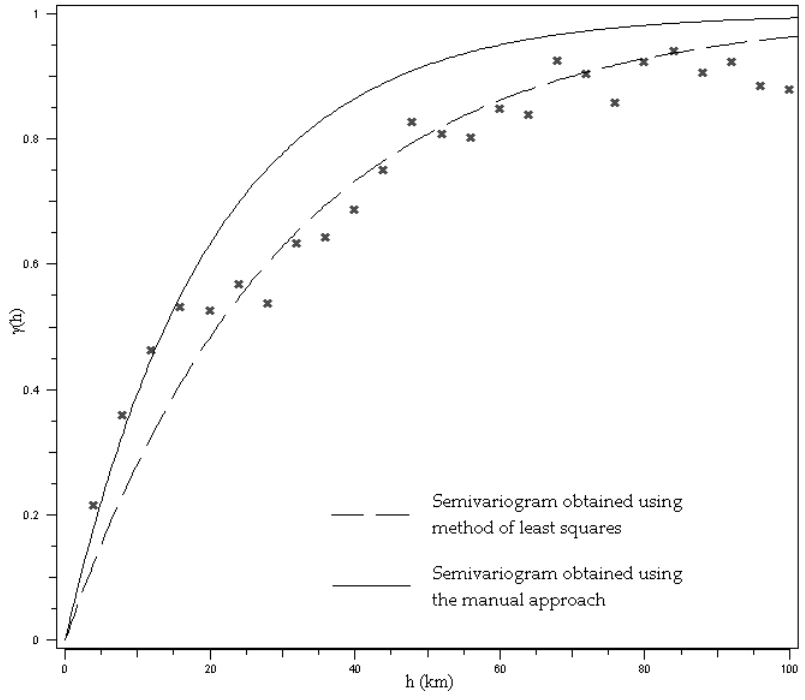

Figure 2: Semivariograms fitted to the same data set using the method of least squares and the manual approach

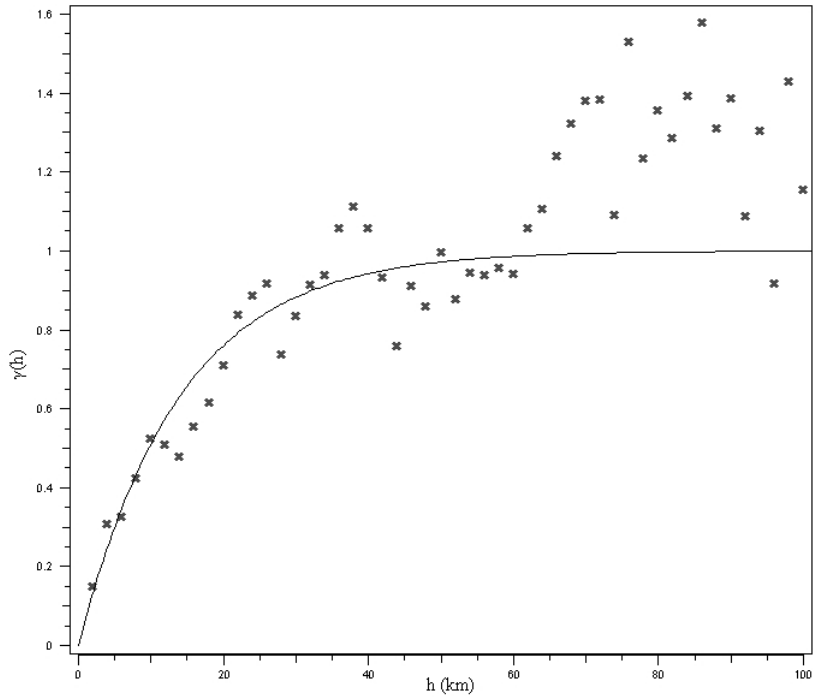

Figure 3: Semivariogram of $\tilde{\varepsilon}$ computed at 2 seconds based on the Northridge earthquake data

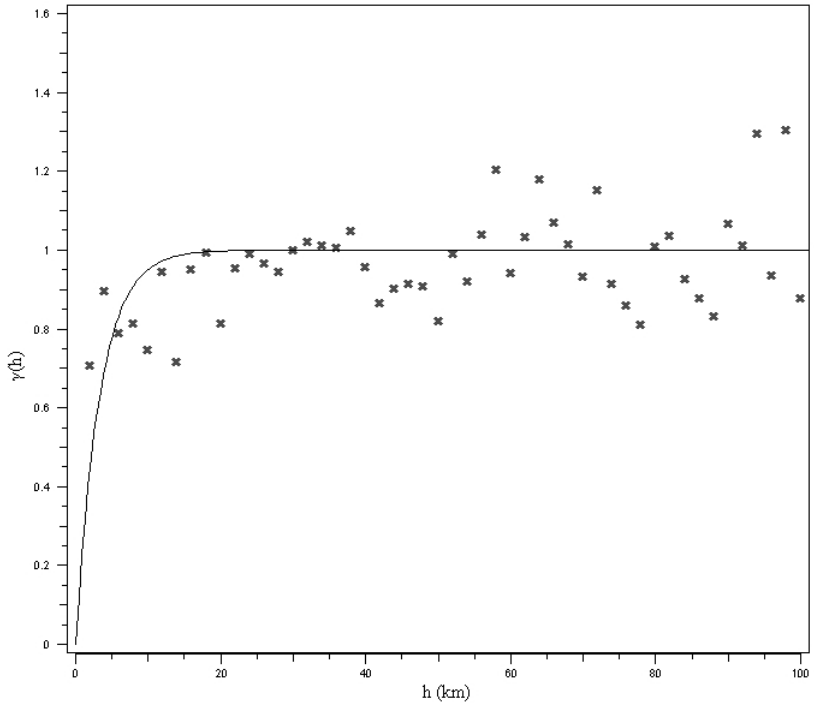

Figure 4: Semivariogram of $\tilde{\varepsilon}$ based on the peak ground accelerations observed during the Northridge earthquake data

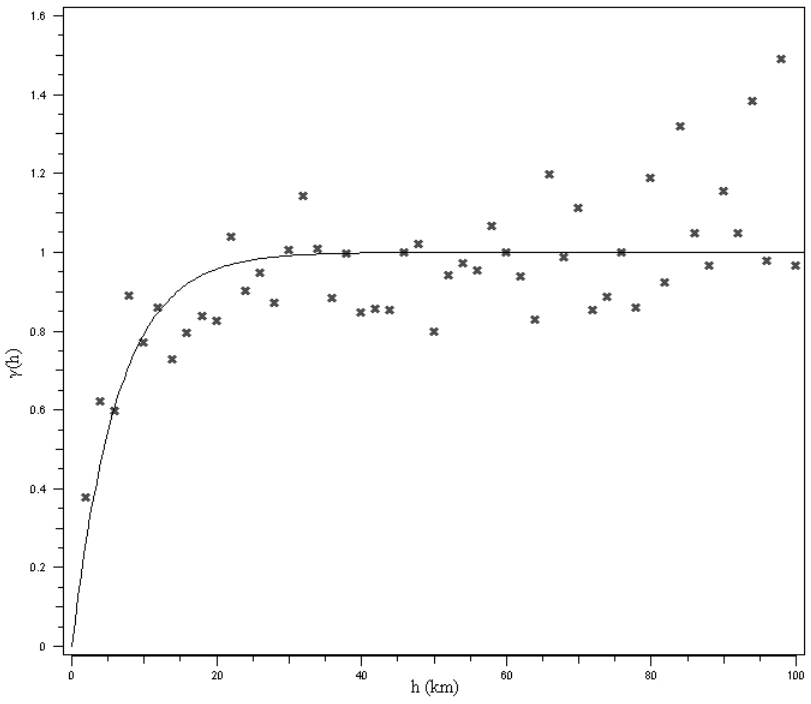

Figure 5: Semivariogram of $\tilde{\varepsilon}$ computed at 0.5 seconds based on the Northridge earthquake data

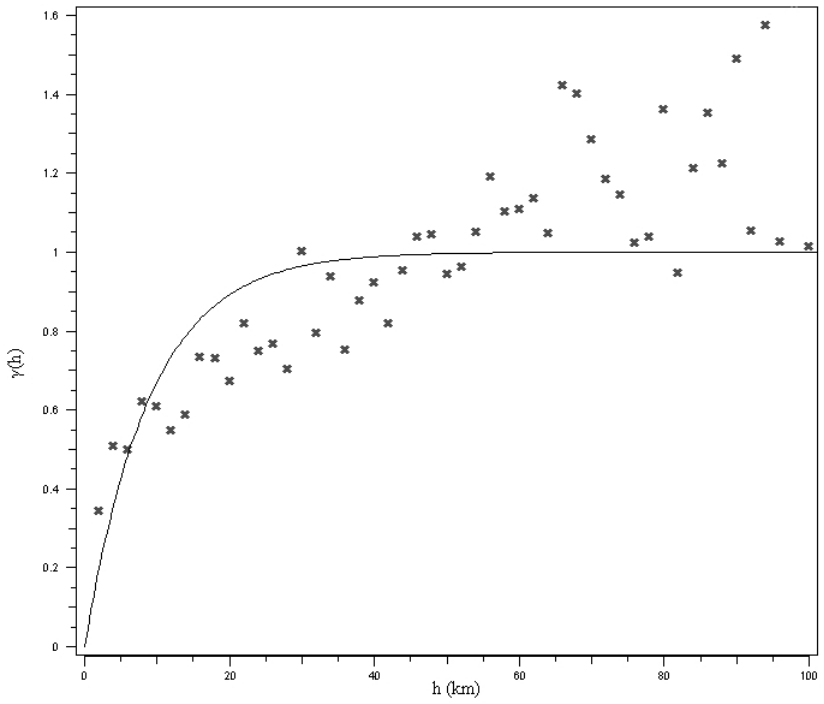

Figure 6: Semivariogram of $\tilde{\varepsilon}$ computed at 1 second based on the Northridge earthquake data

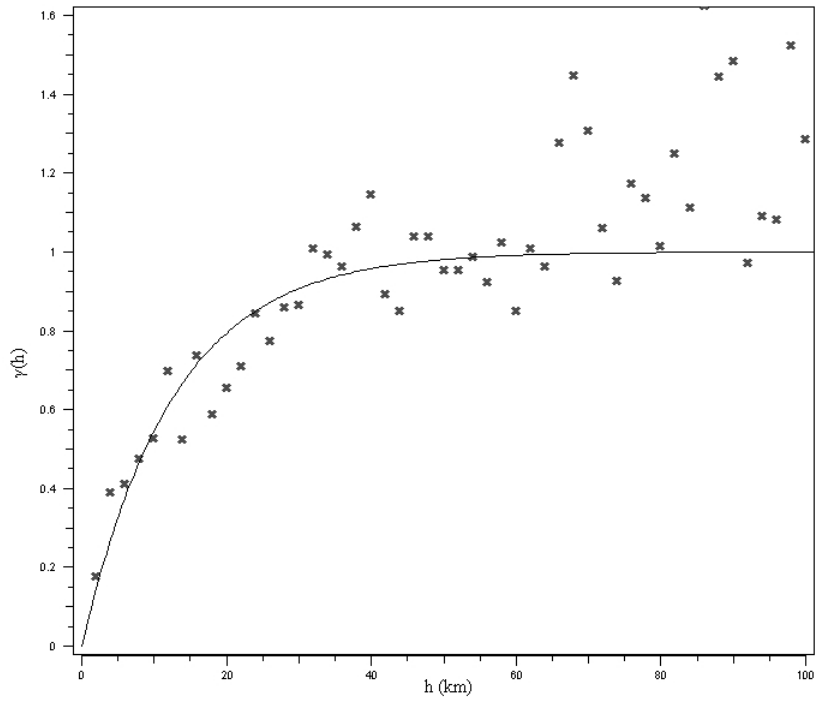

Figure 7: Semivariogram of $\tilde{\varepsilon}$ computed at 5 seconds based on the Northridge earthquake data

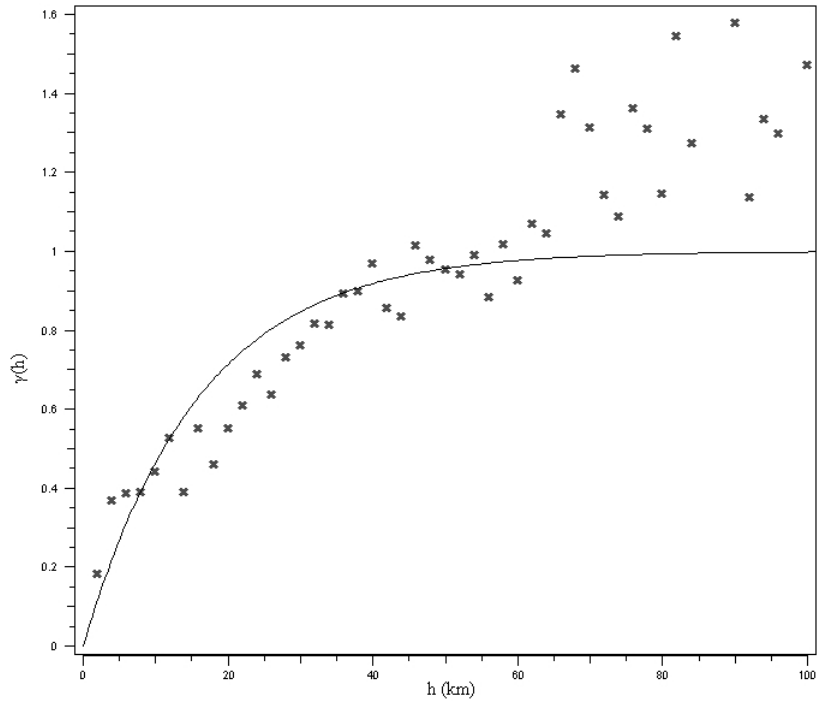

Figure 8: Semivariogram of $\tilde{\varepsilon}$ computed at 7.5 seconds based on the Northridge earthquake data

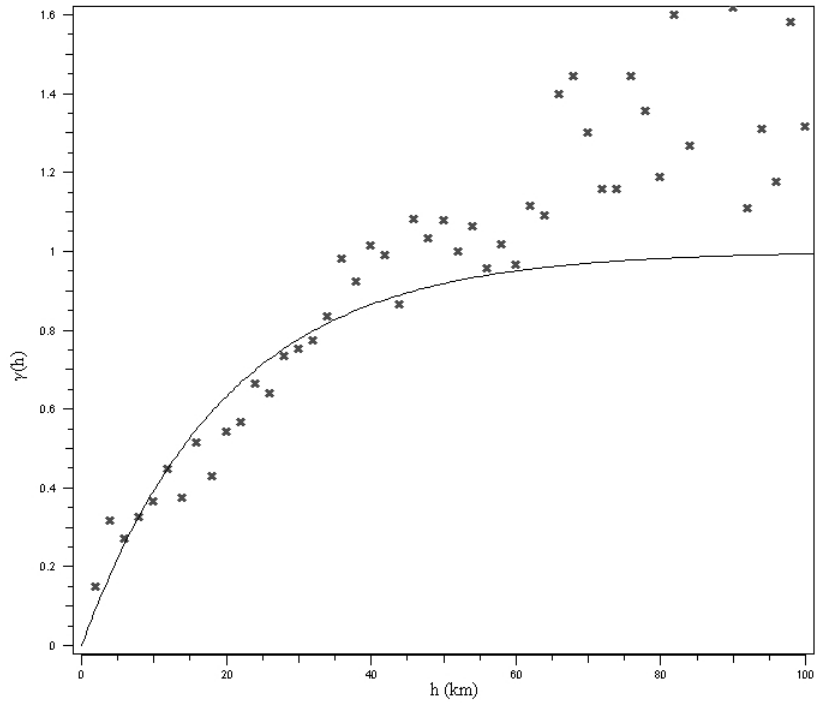

Figure 9: Semivariogram of $\tilde{\varepsilon}$ computed at 10 seconds based on the Northridge earthquake data

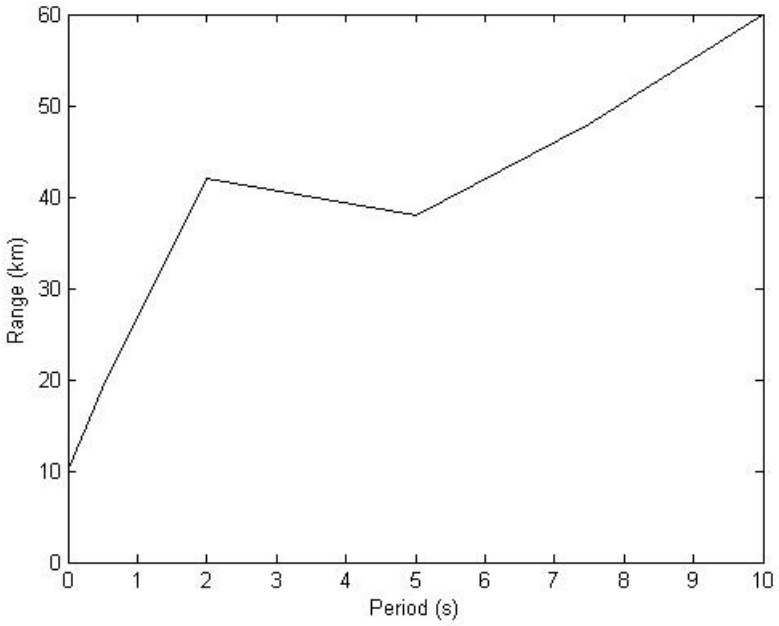

Figure 10: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Northridge earthquake data

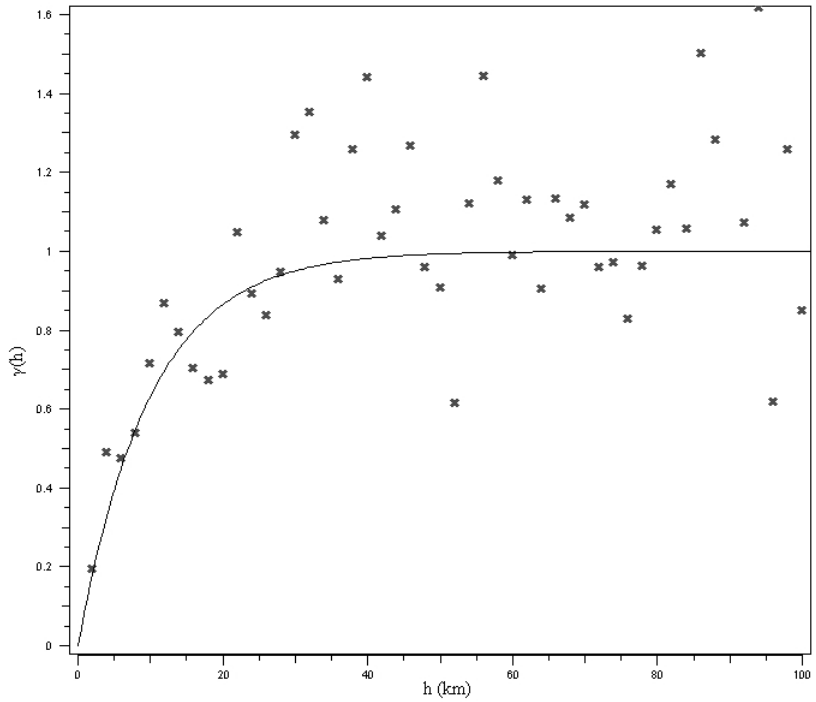

Figure 11: Semivariogram of $\tilde{\varepsilon}$ computed at 5 seconds based on the Northridge earthquake data. Only records used by the authors of the Boore and Atkinson (2007) ground-motion model are considered

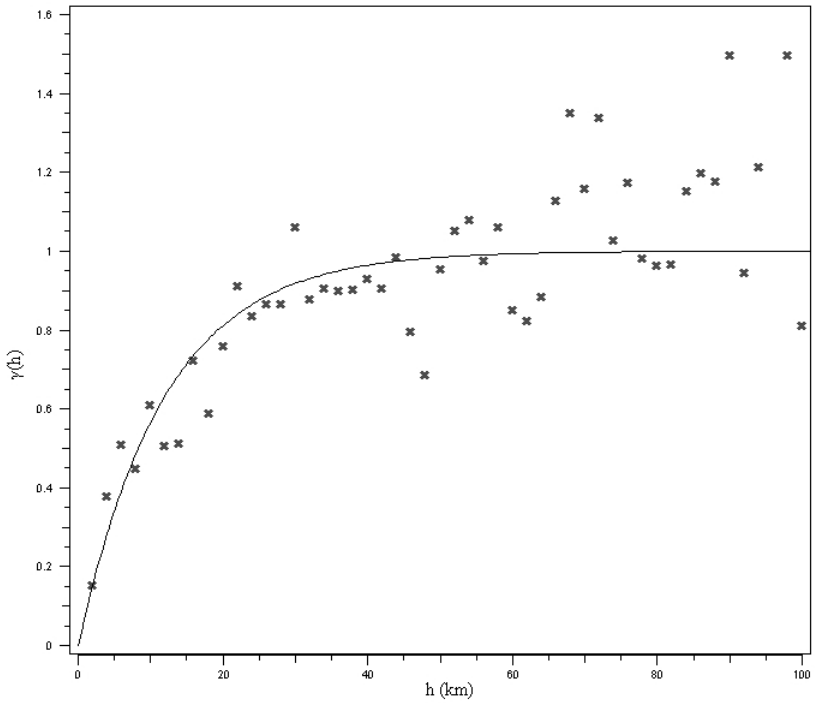

Figure 12: Semivariogram of $\tilde{\varepsilon}$ computed at 2 seconds based on the Northridge earthquake fault-parallel time histories

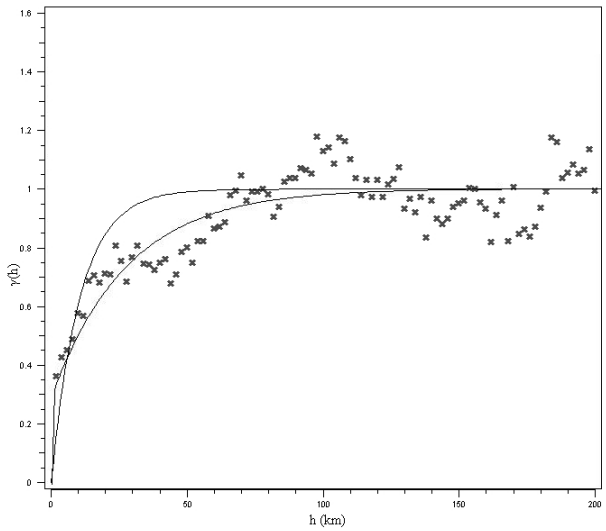

Figure 13: Experimental semivariogram of $\tilde{\varepsilon}$ computed at 2 seconds based on the Chi-Chi earthquake data. Also shown in the figure are two fitted semivariogram models: (i) An accurate exponential + nugget model and (ii) An approximate exponential model

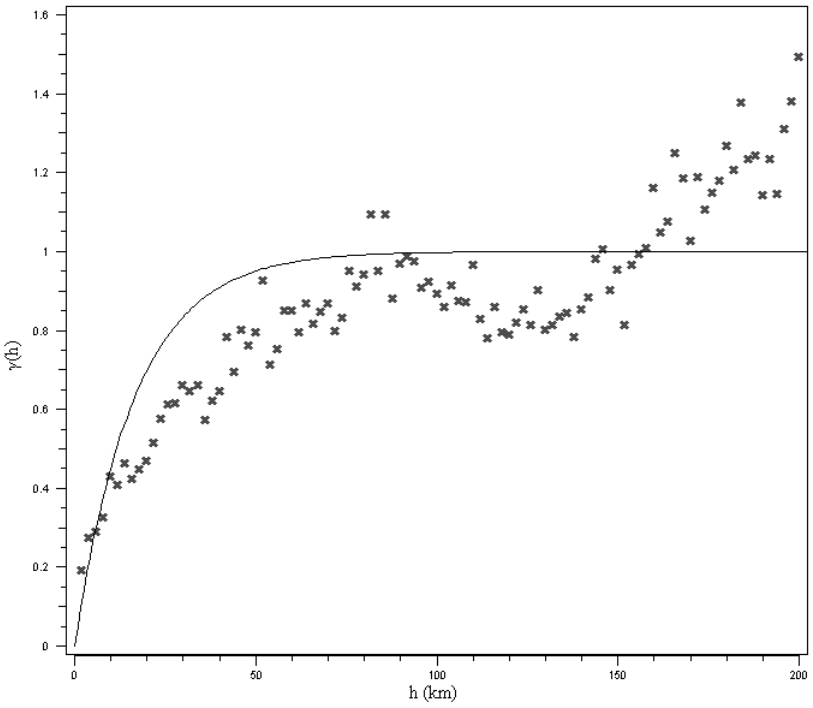

Figure 14: Semivariogram of $\tilde{\varepsilon}$ based on the peak ground accelerations observed during the Chi-Chi earthquake data

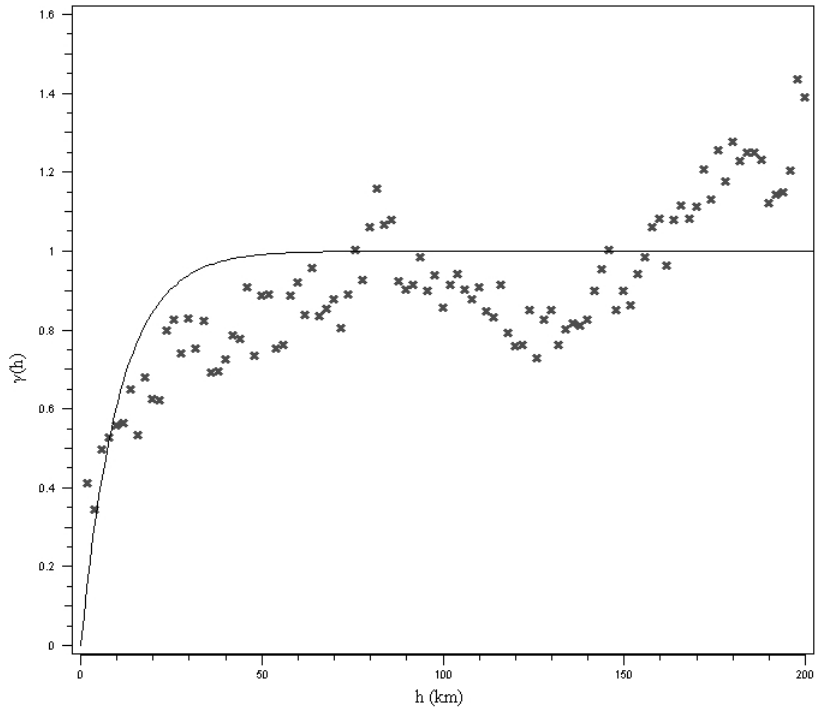

Figure 15: Semivariogram of $\tilde{\varepsilon}$ computed at 0.5 seconds based on the Chi-Chi earthquake data

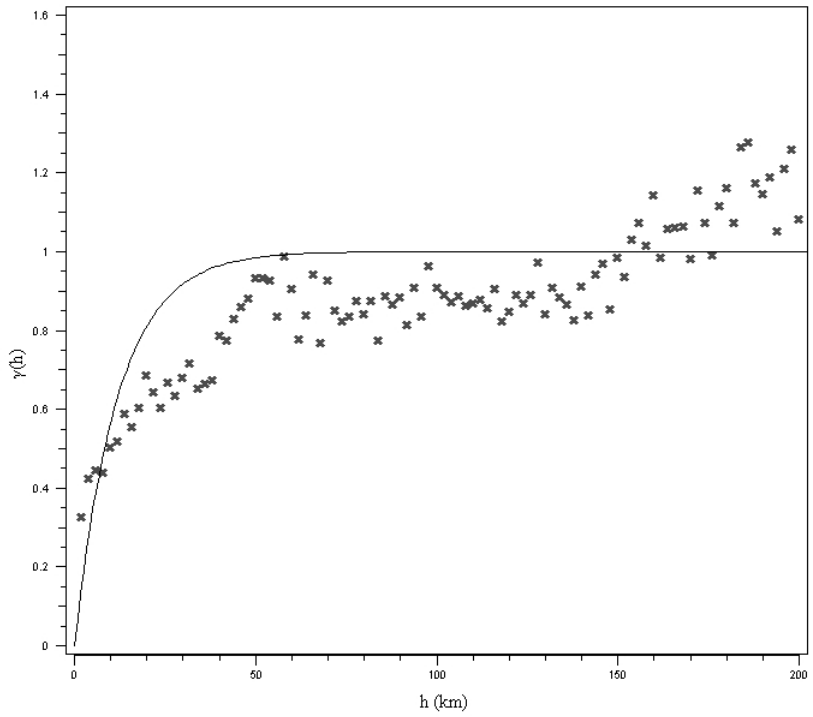

Figure 16: Semivariogram of $\tilde{\varepsilon}$ computed at 1 second based on the Chi-Chi earthquake data

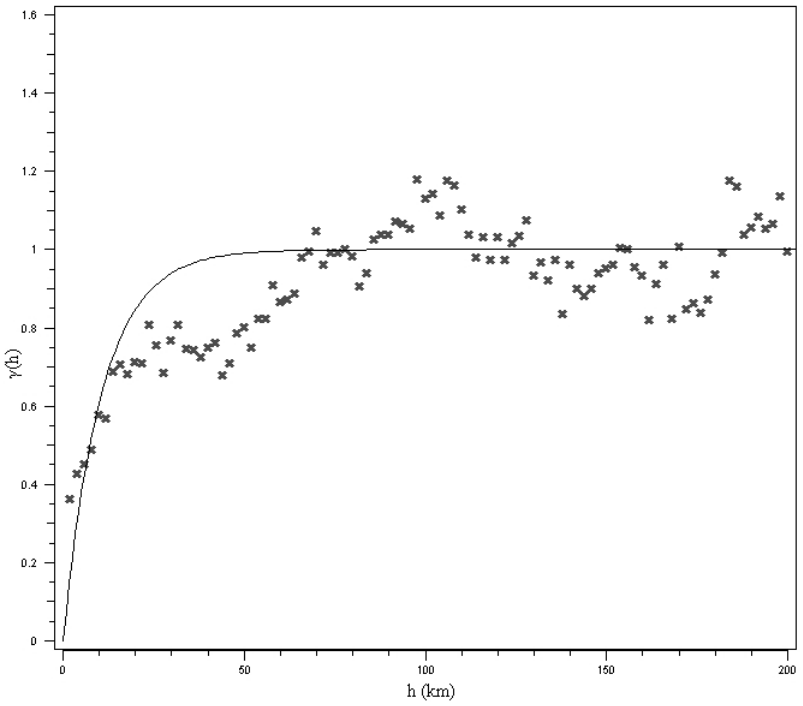

Figure 17: (Approximate) Semivariogram of $\tilde{\varepsilon}$ computed at 2 seconds based on the Chi-Chi earthquake data

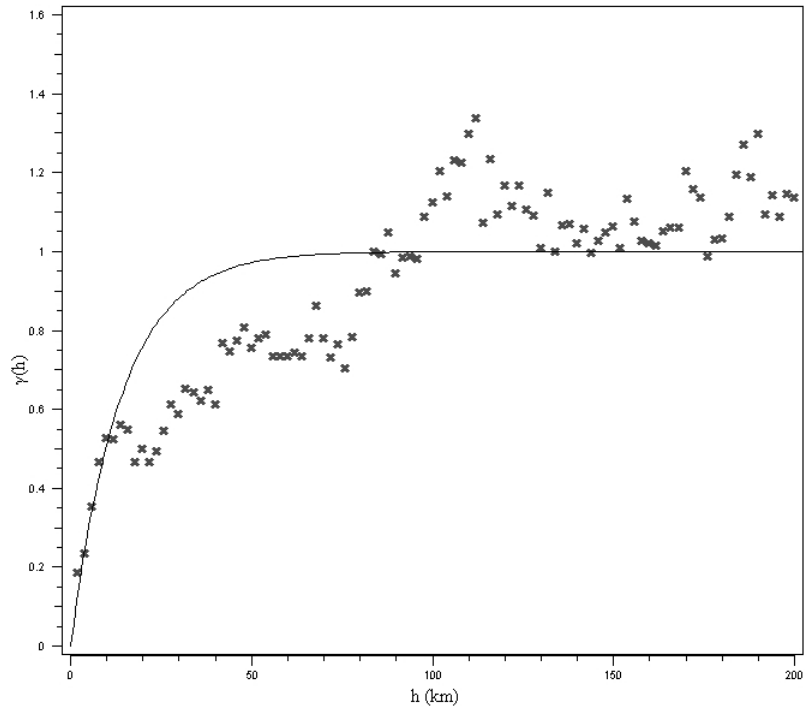

Figure 18: Semivariogram of $\tilde{\varepsilon}$ computed at 5 seconds based on the Chi-Chi earthquake data

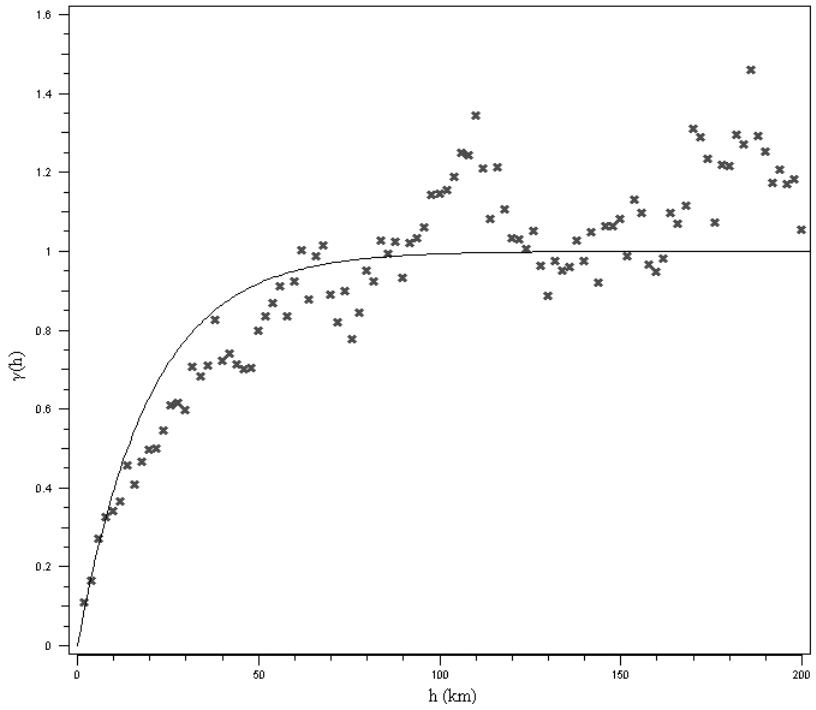

Figure 19: Semivariogram of $\tilde{\varepsilon}$ computed at 7.5 seconds based on the Chi-Chi earthquake data

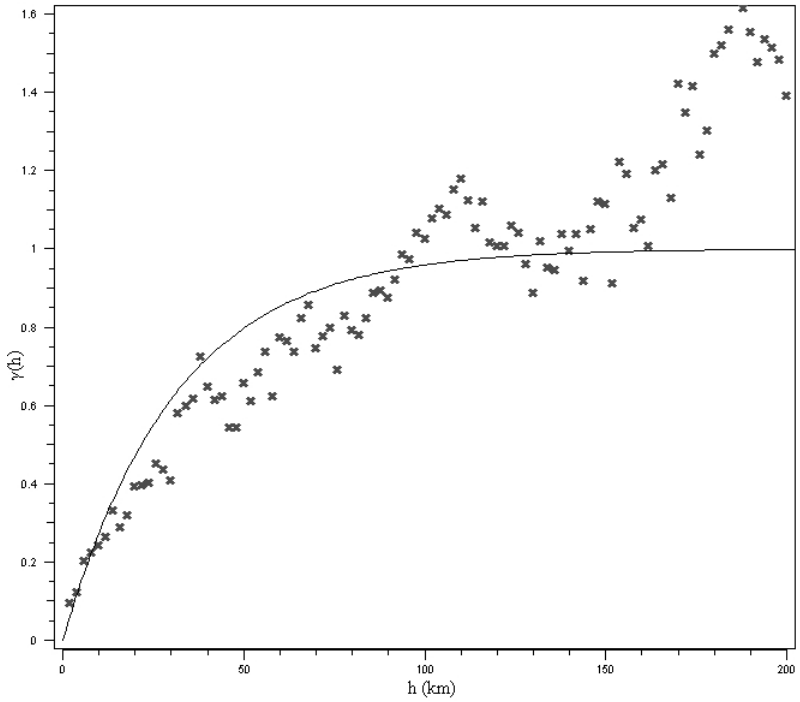

Figure 20: Semivariogram of $\tilde{\varepsilon}$ computed at 10 seconds based on the Chi-Chi earthquake data

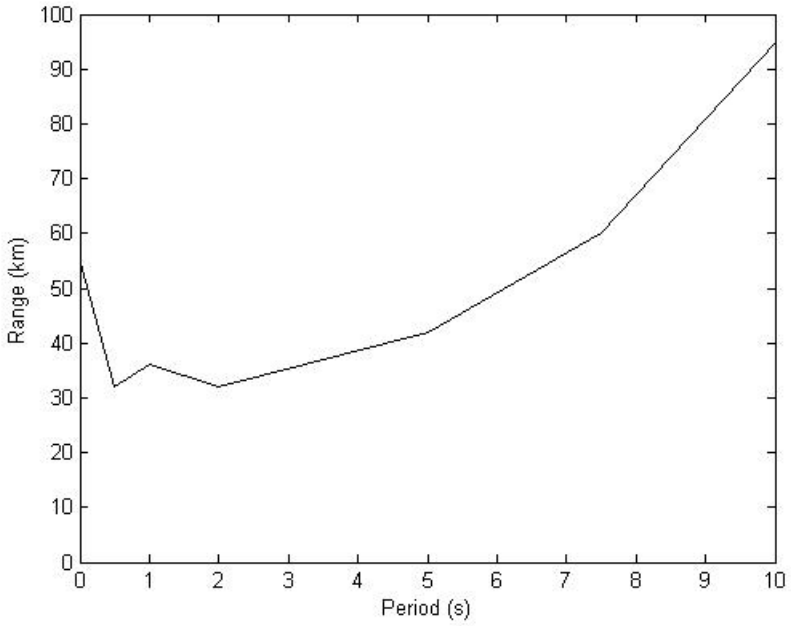

Figure 21: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Chi-Chi earthquake data

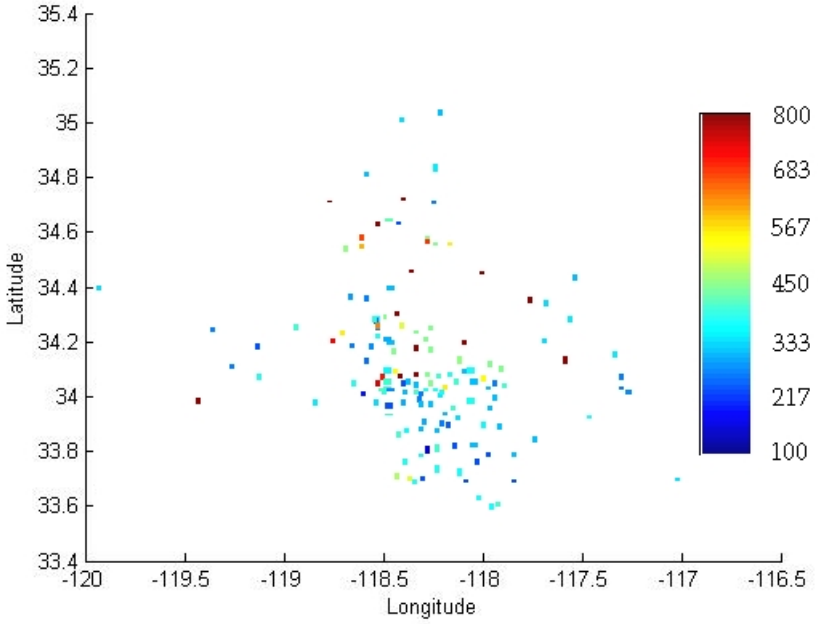

Figure 22: V_s30 at the recording stations of the Northridge earthquake.

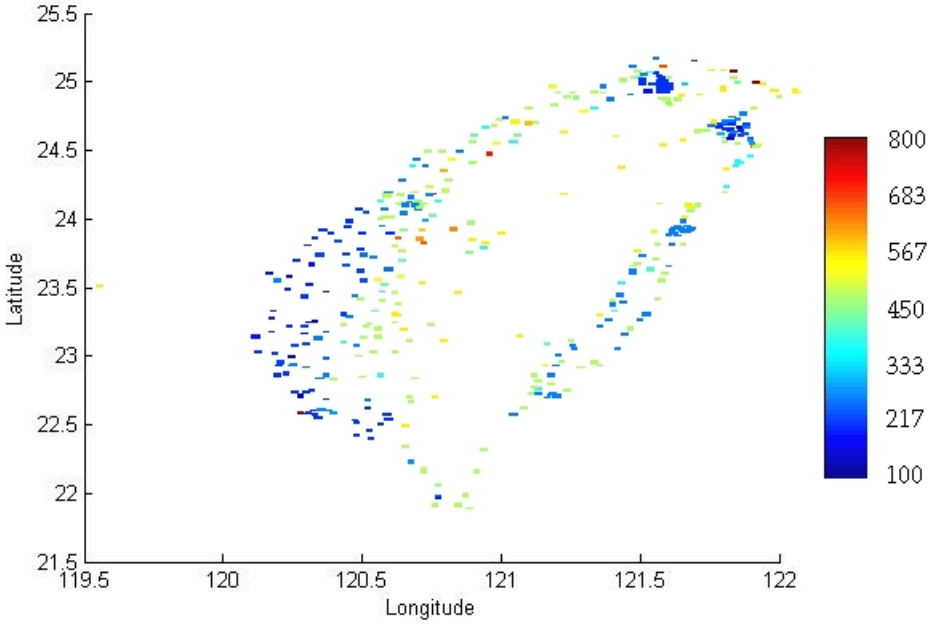

Figure 23: V_s30 at the recording stations of the Chi-Chi earthquake.

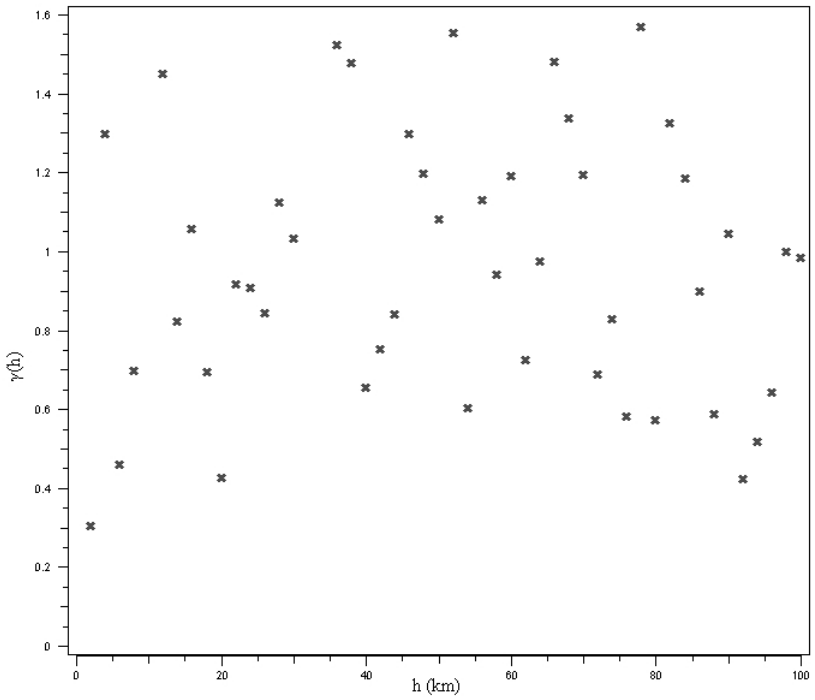

Figure 24: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Northridge earthquake. No semivariogram is fitted on account of the extreme scatter

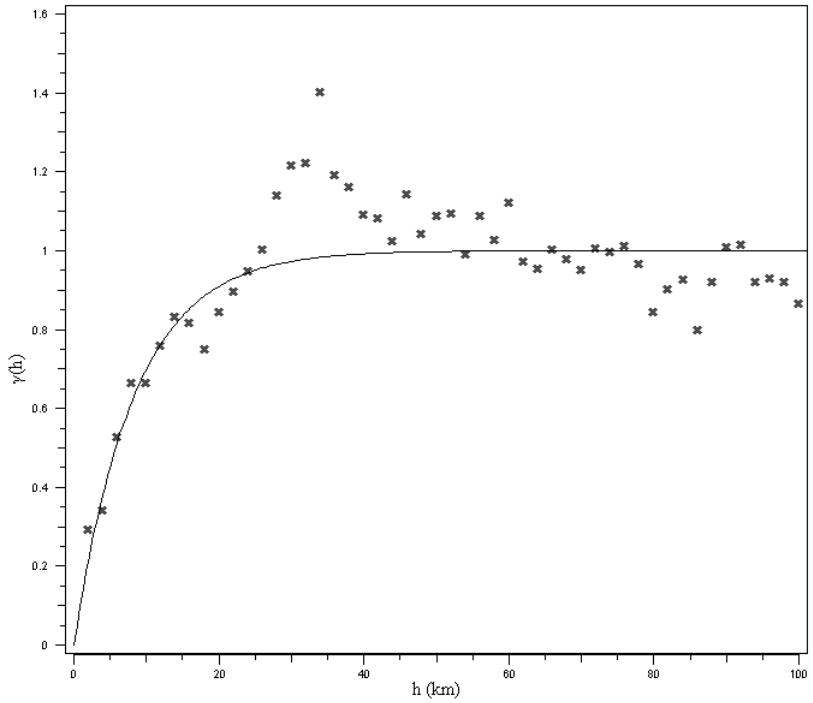

Figure 25: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Chi-Chi earthquake. The range of the fitted exponential semivariogram equals 25 km

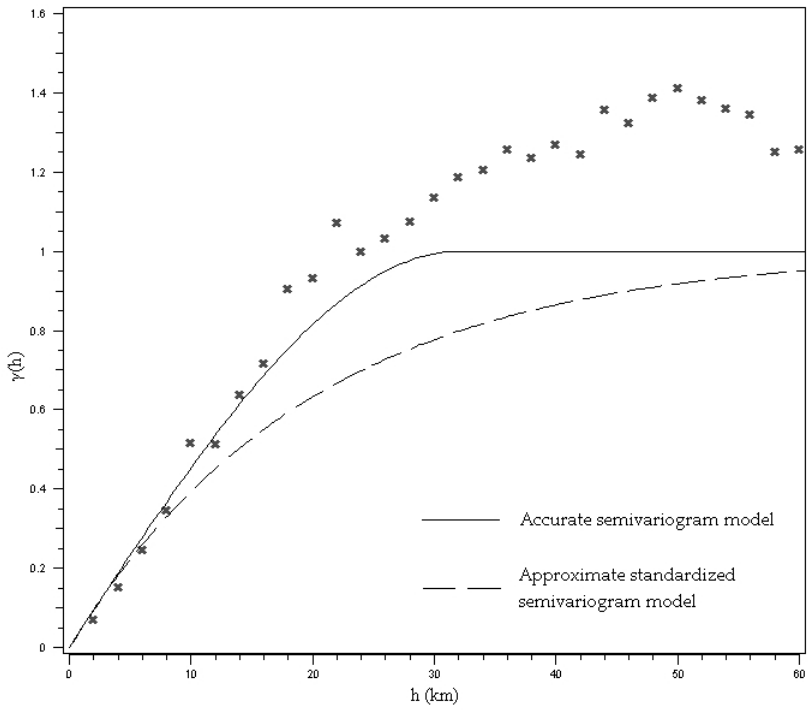

Figure 26: Experimental Semivariogram of $\tilde{\varepsilon}$ computed at 5 seconds based on the simulated ground-motion data. Also shown in the figure are two fitted semivariogram models: (i) An accurate spherical model and (ii) An approximate exponential model

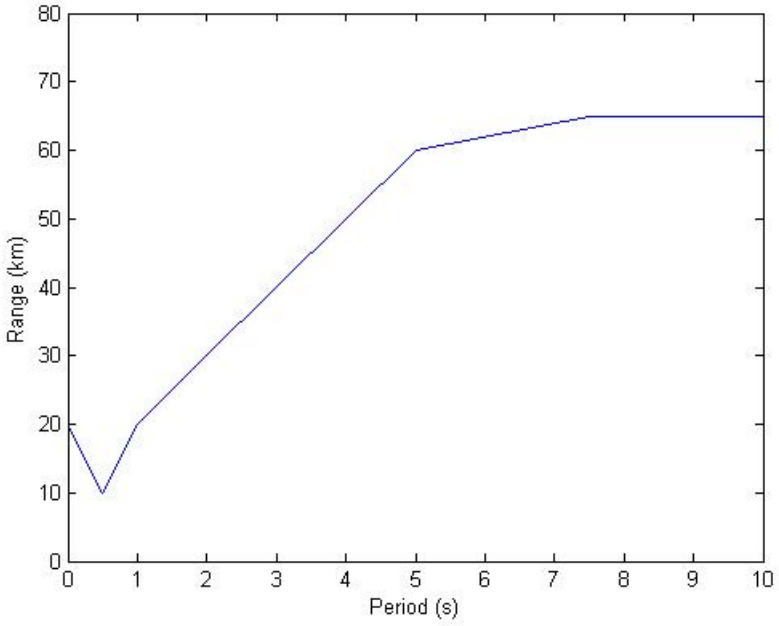

Figure 27: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the simulated ground-motion data

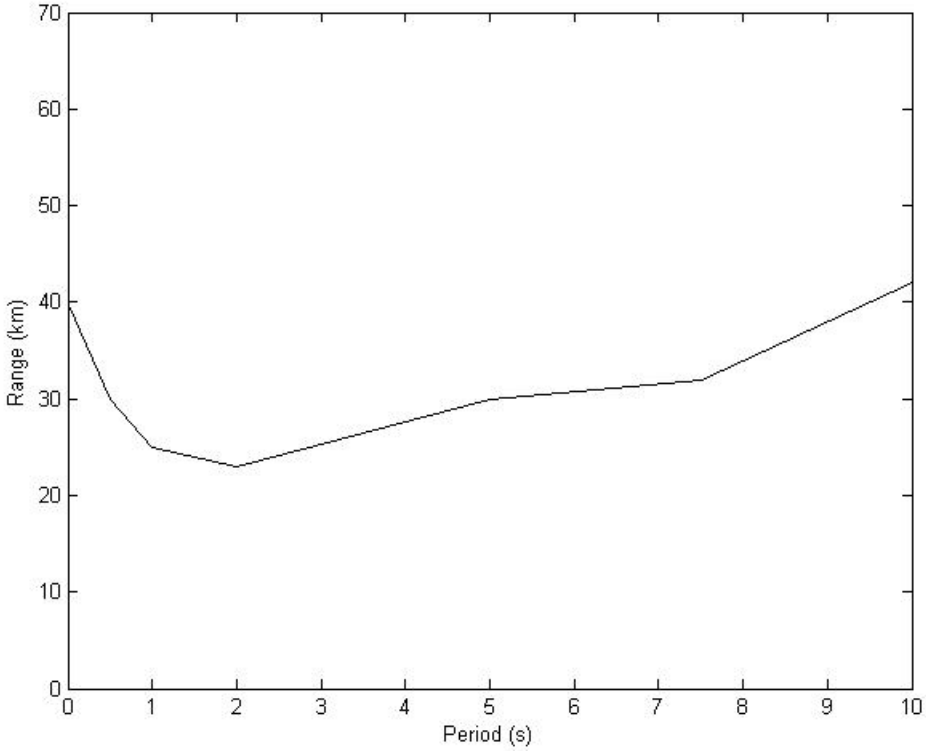

Figure 28: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Big Bear City earthquake data

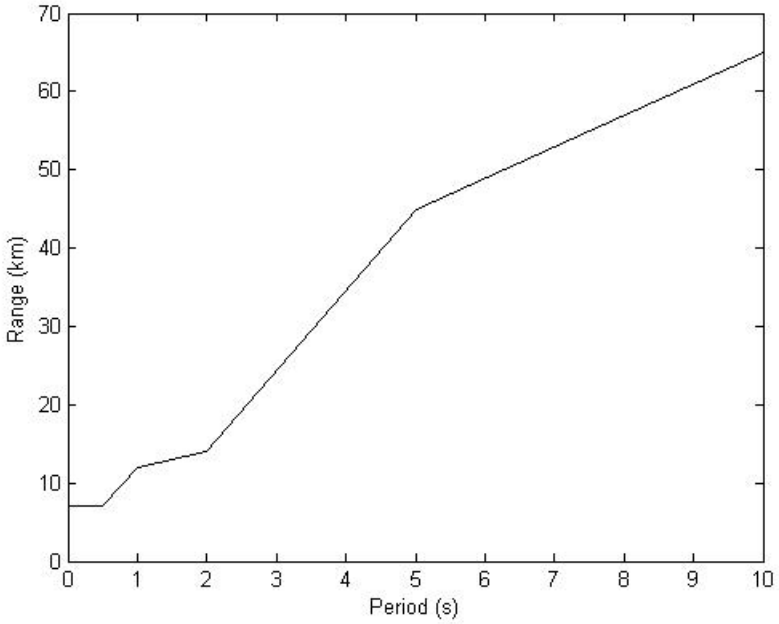

Figure 29: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Parkfield earthquake data

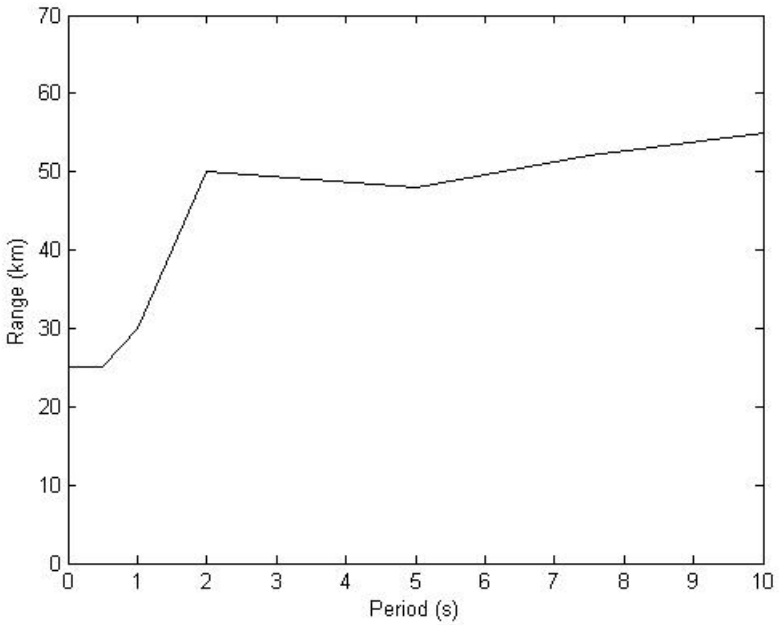

Figure 30: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Alum Rock earthquake data

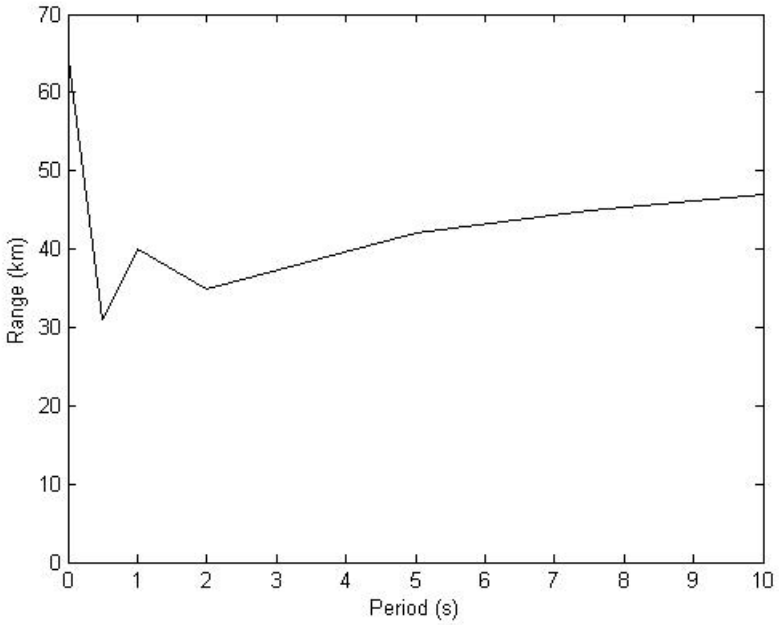

Figure 31: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Anza earthquake data

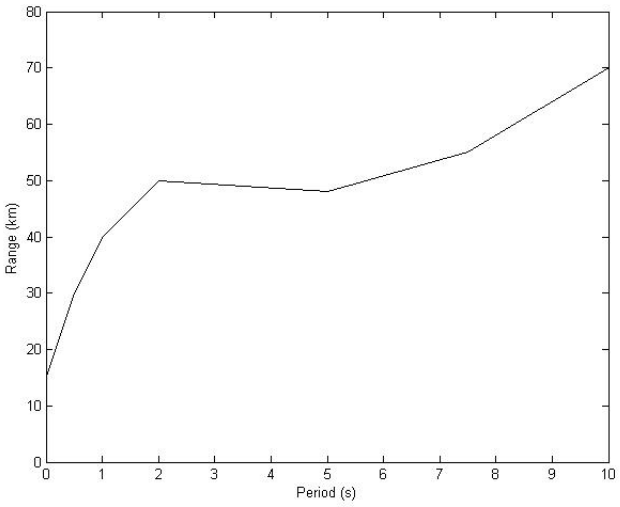

Figure 32: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained using the Chino Hills earthquake data

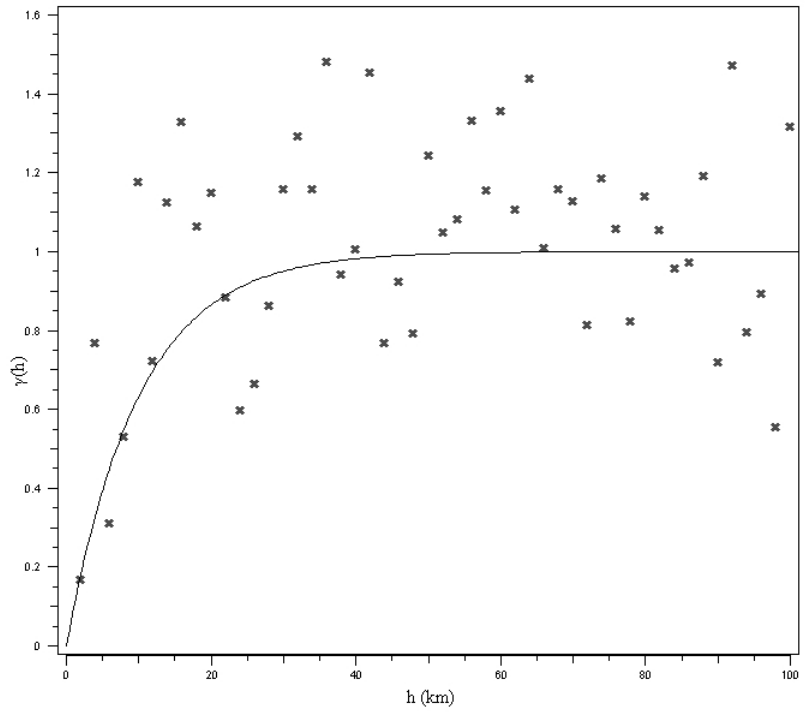

Figure 33: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Big Bear City earthquake. The range of the fitted exponential semivariogram equals 30 km

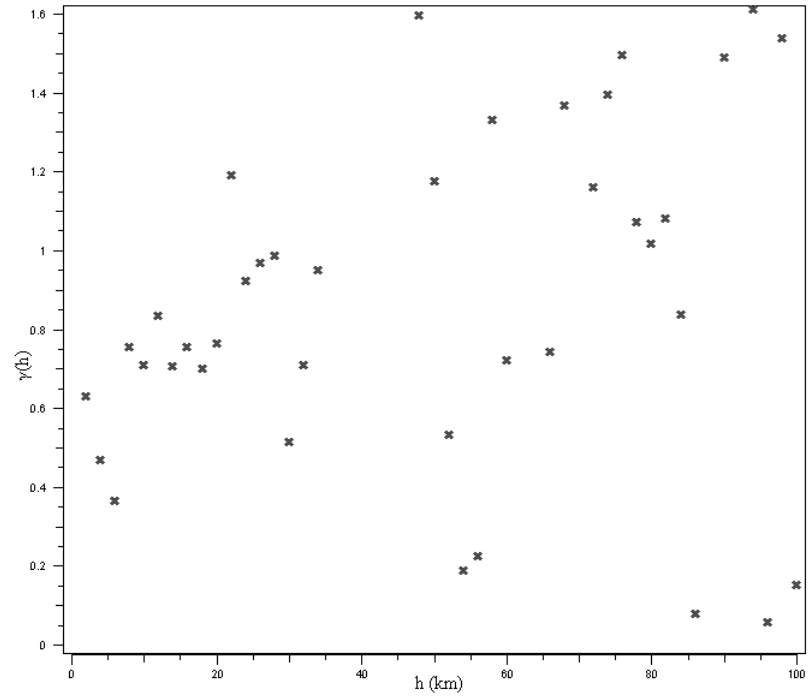

Figure 34: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Parkfield earthquake. No semivariogram is fitted on account of the extreme scatter

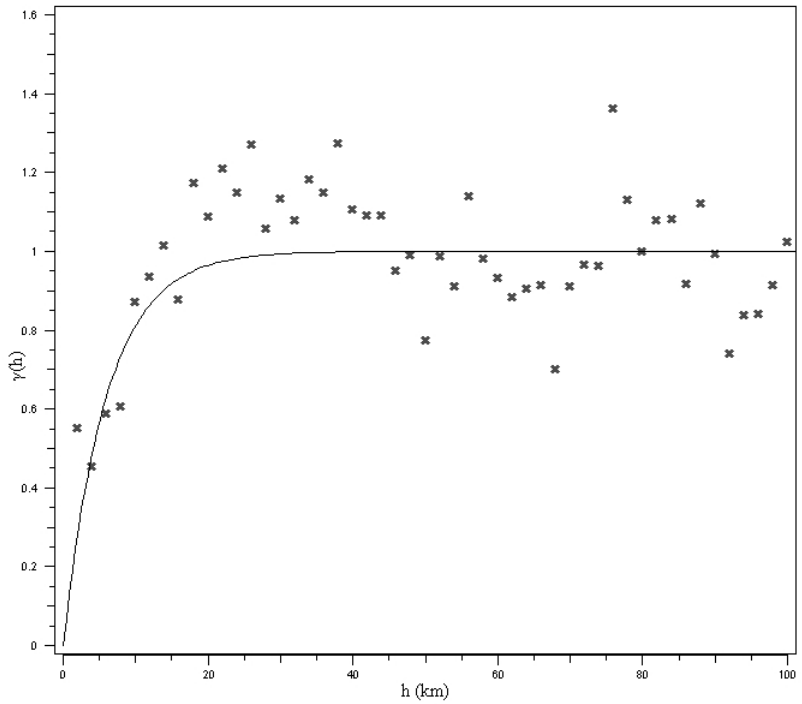

Figure 35: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Alum Rock earthquake. The range of the fitted exponential semivariogram equals 18 km

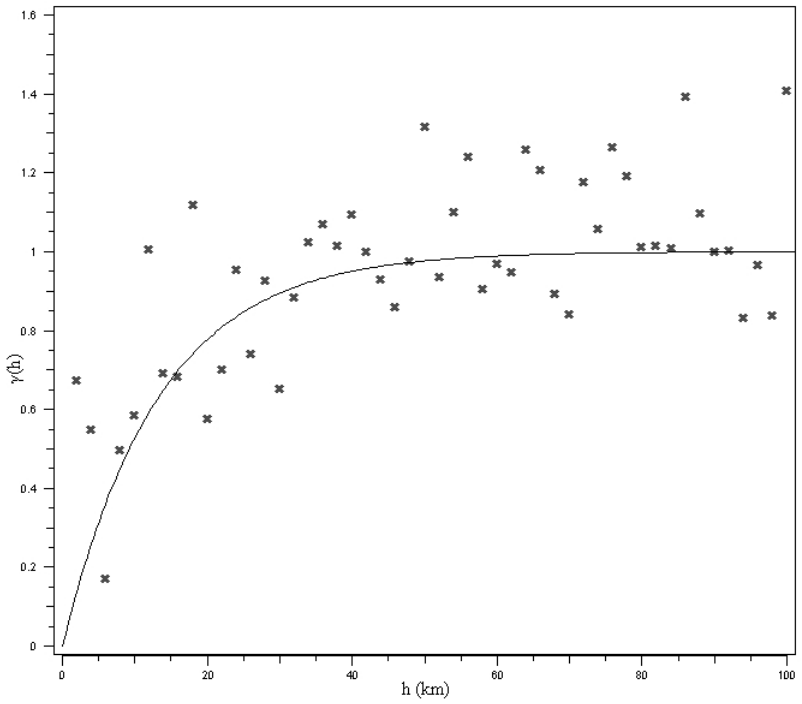

Figure 36: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Anza earthquake. The range of the fitted exponential semivariogram equals 40 km

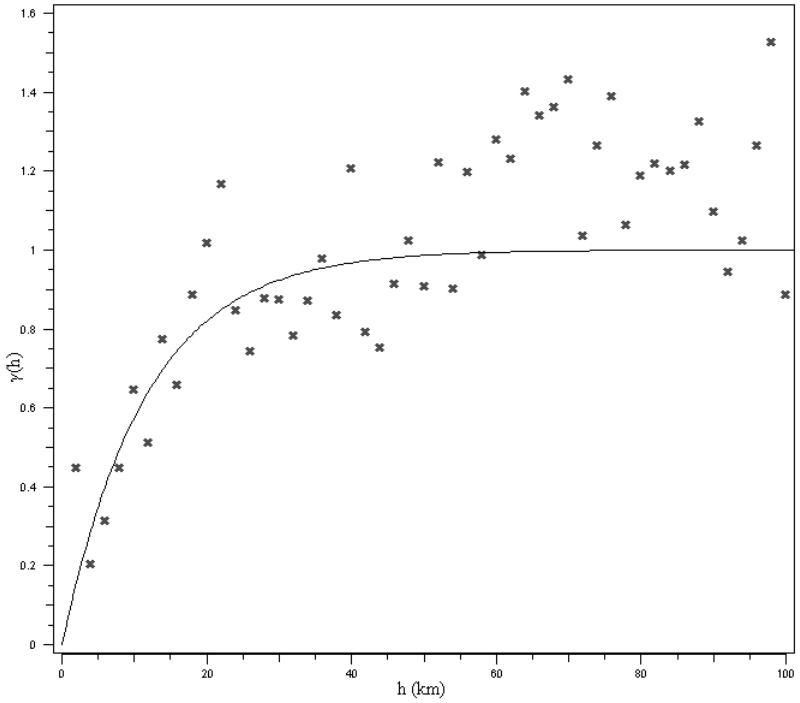

Figure 37: Experimental semivariogram obtained using normalized V_s30 at the recording stations of the Chino Hills earthquake. The range of the fitted exponential semivariogram equals 35 km

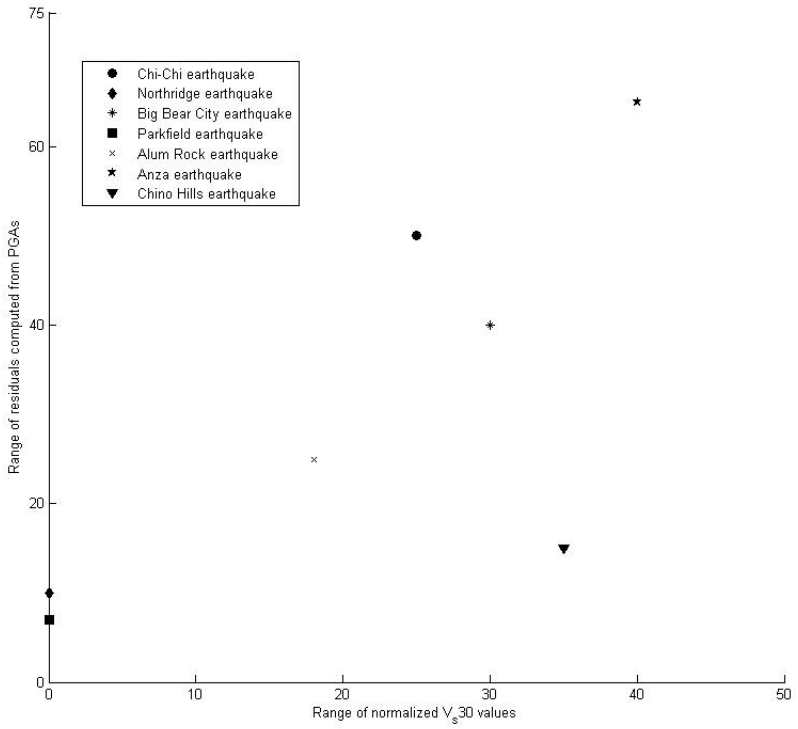

Figure 38: Ranges of residuals computed using PGAs versus ranges of normalized V_s30 values

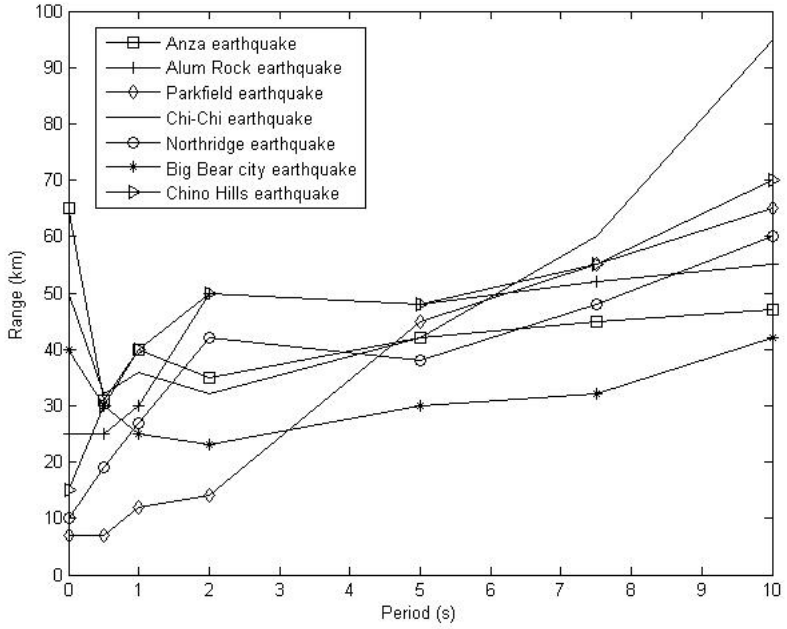

Figure 39: Range of semivariograms of $\tilde{\varepsilon}$, as a function of the period at which $\tilde{\varepsilon}$ values are computed. The residuals are obtained from six different sets of time histories as shown in the figure. The simulated ground motions are not reliable at short periods (below 2 seconds) and hence, the corresponding ranges are ignored

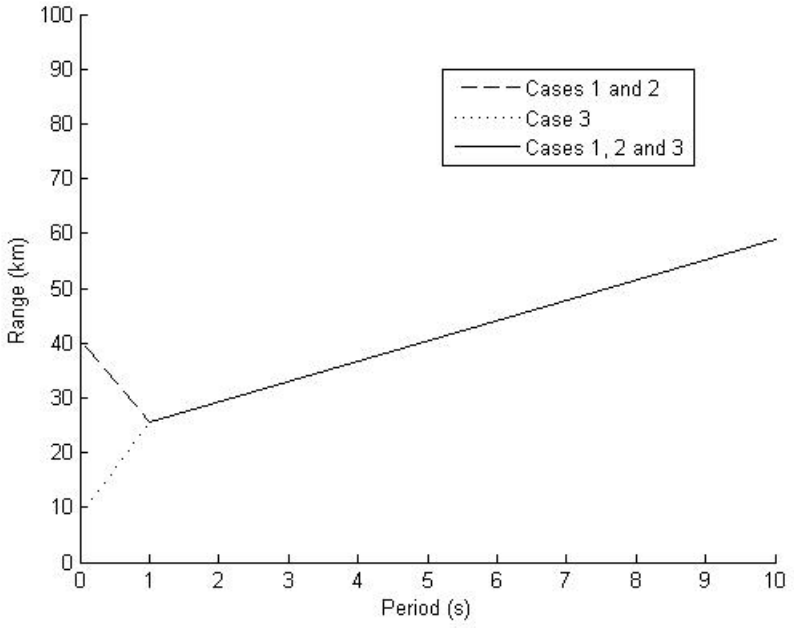

Figure 40: Range of semivariograms of $\tilde{\varepsilon}$ predicted by the proposed model as a function of the period

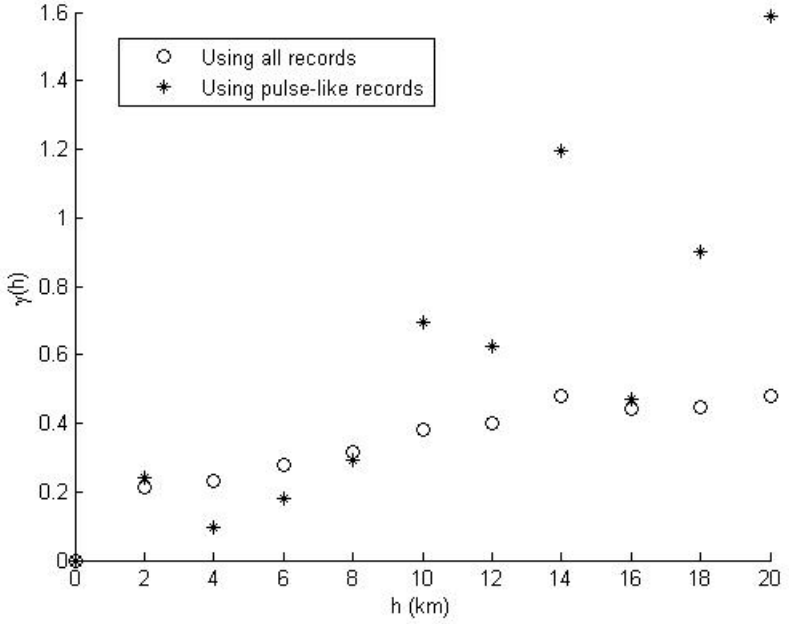

Figure 41: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are computed from peak ground accelerations

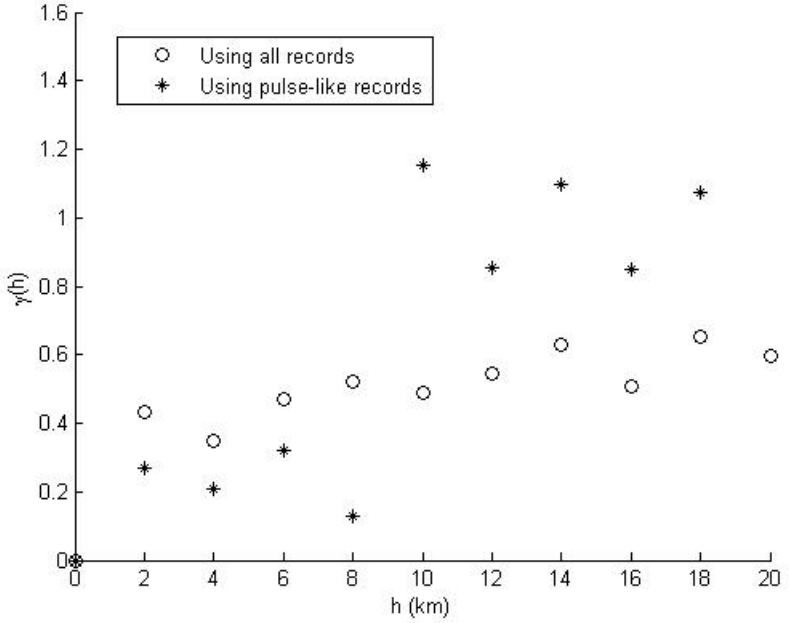

Figure 42: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are obtained from spectral accelerations computed at 0.5 seconds

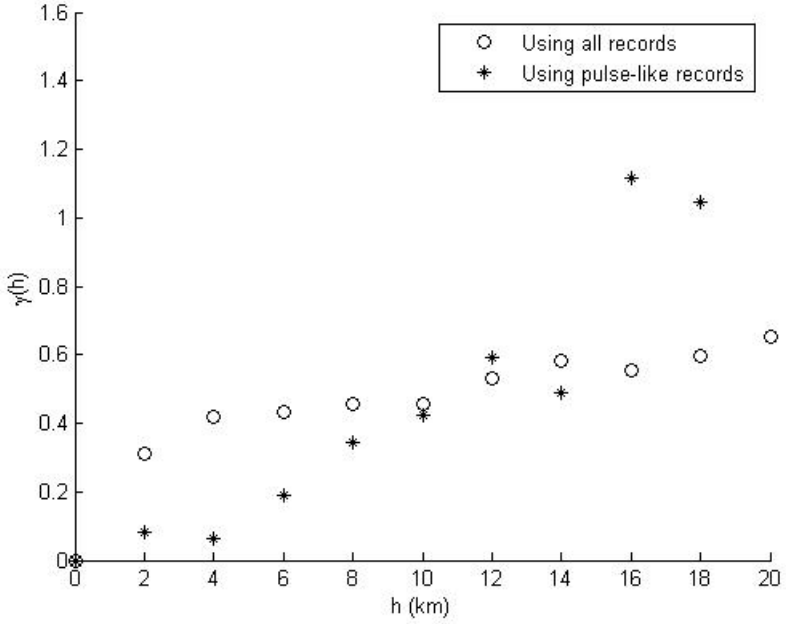

Figure 43: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are obtained from spectral accelerations computed at 1 second

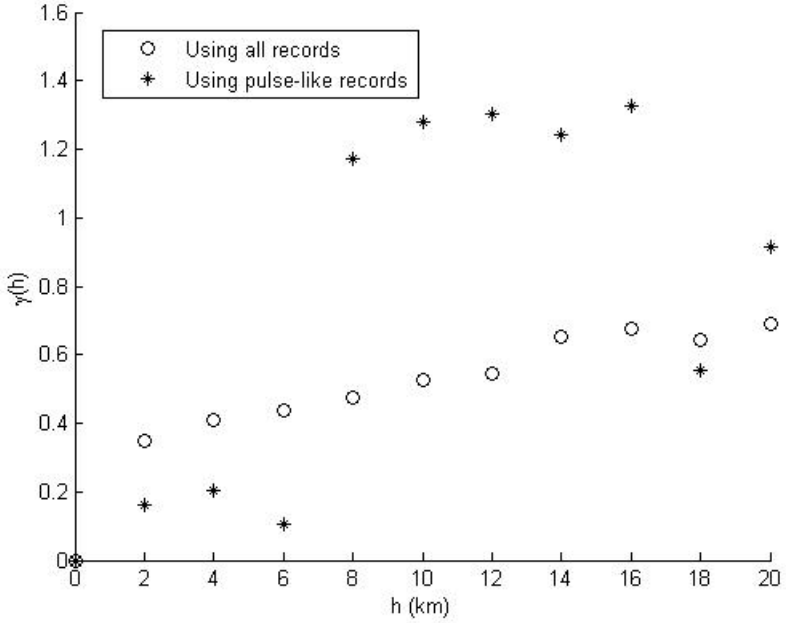

Figure 44: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are obtained from spectral accelerations computed at 2 seconds

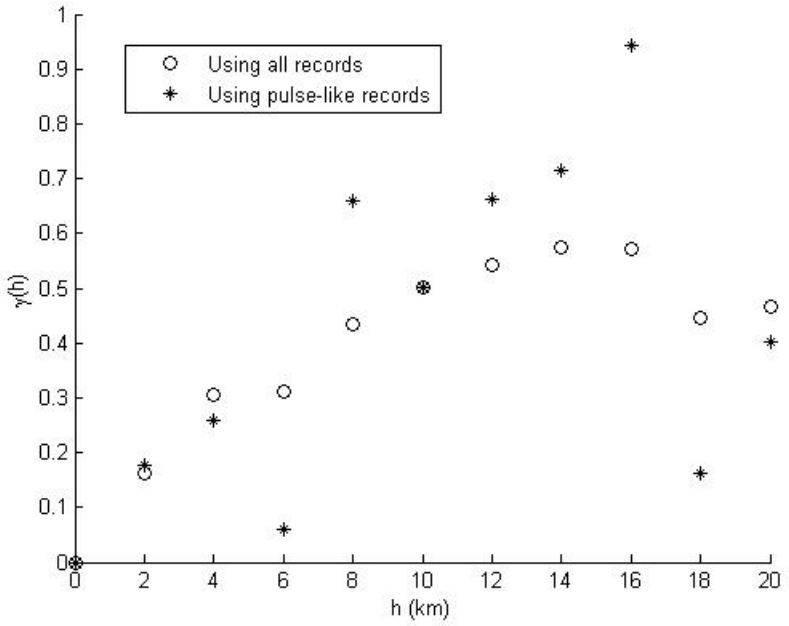

Figure 45: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are obtained from spectral accelerations computed at 5 seconds

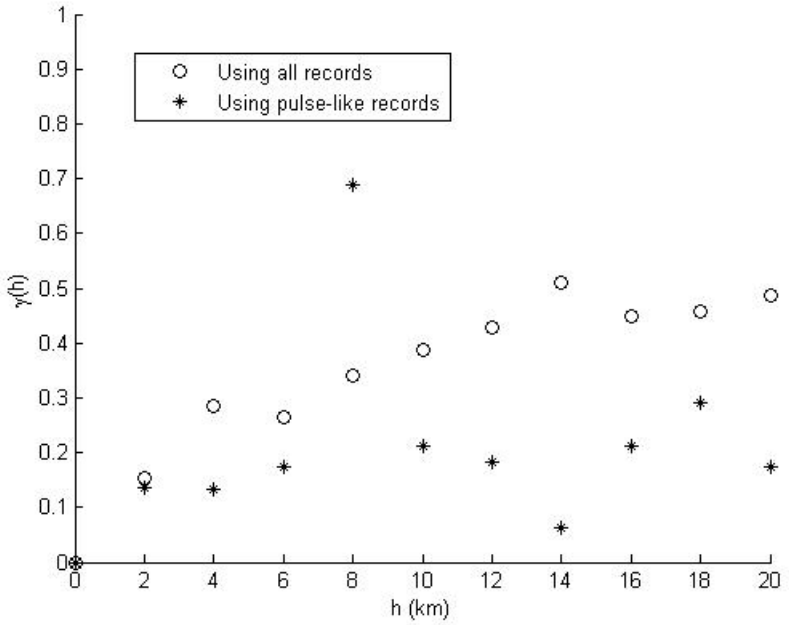

Figure 46: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are obtained from spectral accelerations computed at 7.5 seconds

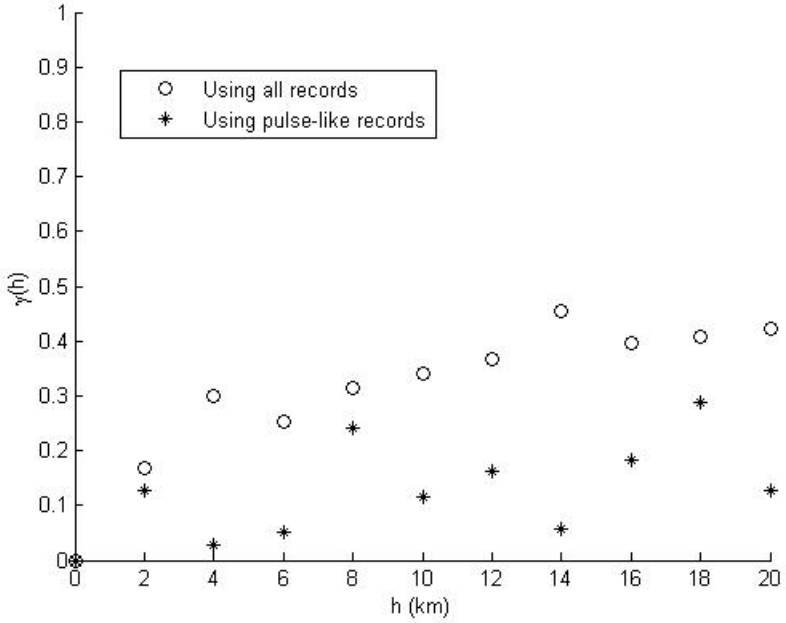

Figure 47: Comparison between the experimental semivariogram of $\tilde{\varepsilon}$'s computed using pulse-like ground motions and the experimental semivariogram of $\tilde{\varepsilon}$'s computed using all usable ground motions. The $\tilde{\varepsilon}$'s are obtained from spectral accelerations computed at 10 seconds

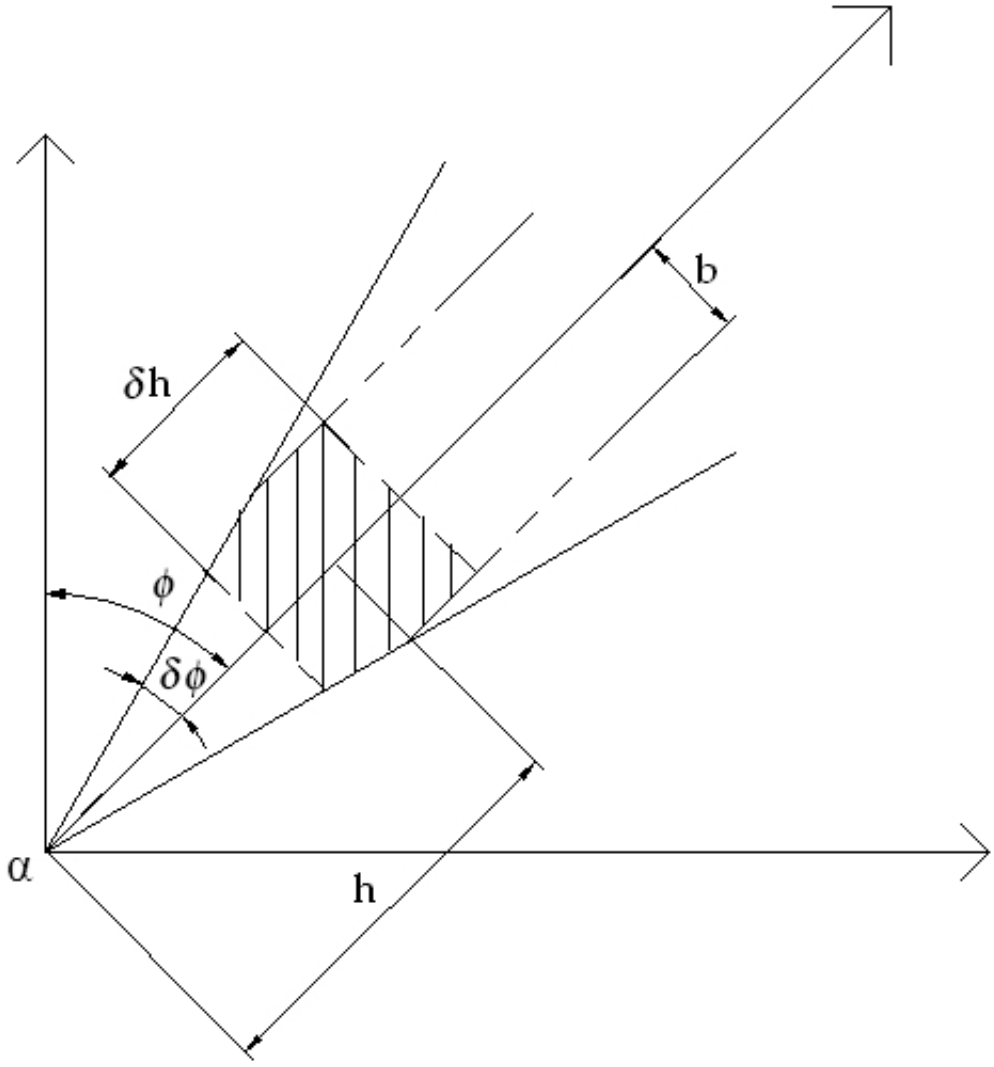

Figure 48: Parameters of a directional semivariogram

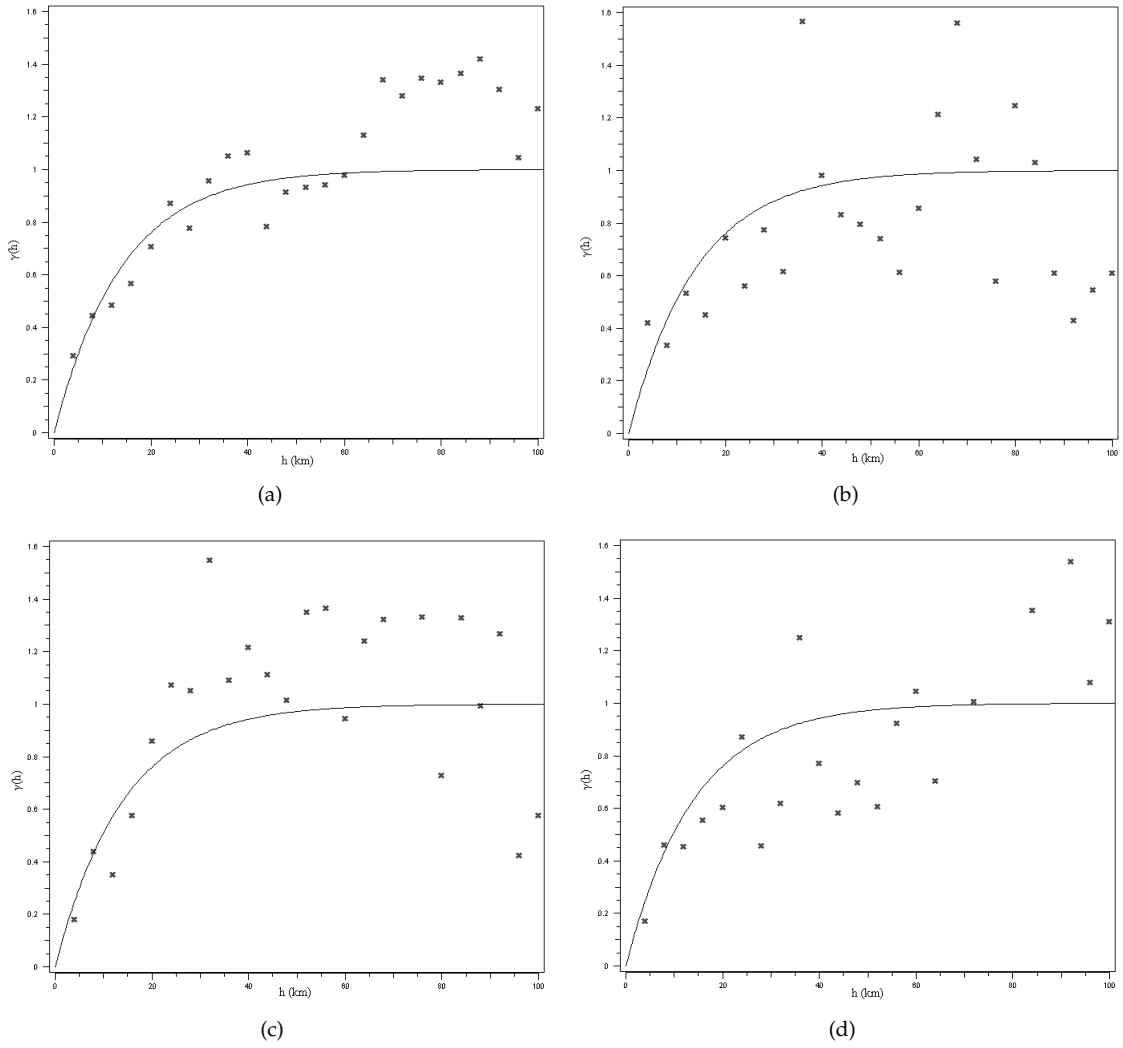

Figure 49: Experimental directional semivariograms at discrete separations obtained using the Northridge earthquake $\tilde{\varepsilon}$ values computed at 2 seconds. Also shown in the figures is the best fit to the omni-directional semivariogram: (a) Omni-directional; (b) Azimuth = 0; (c) Azimuth = 45 and (d) Azimuth = 90

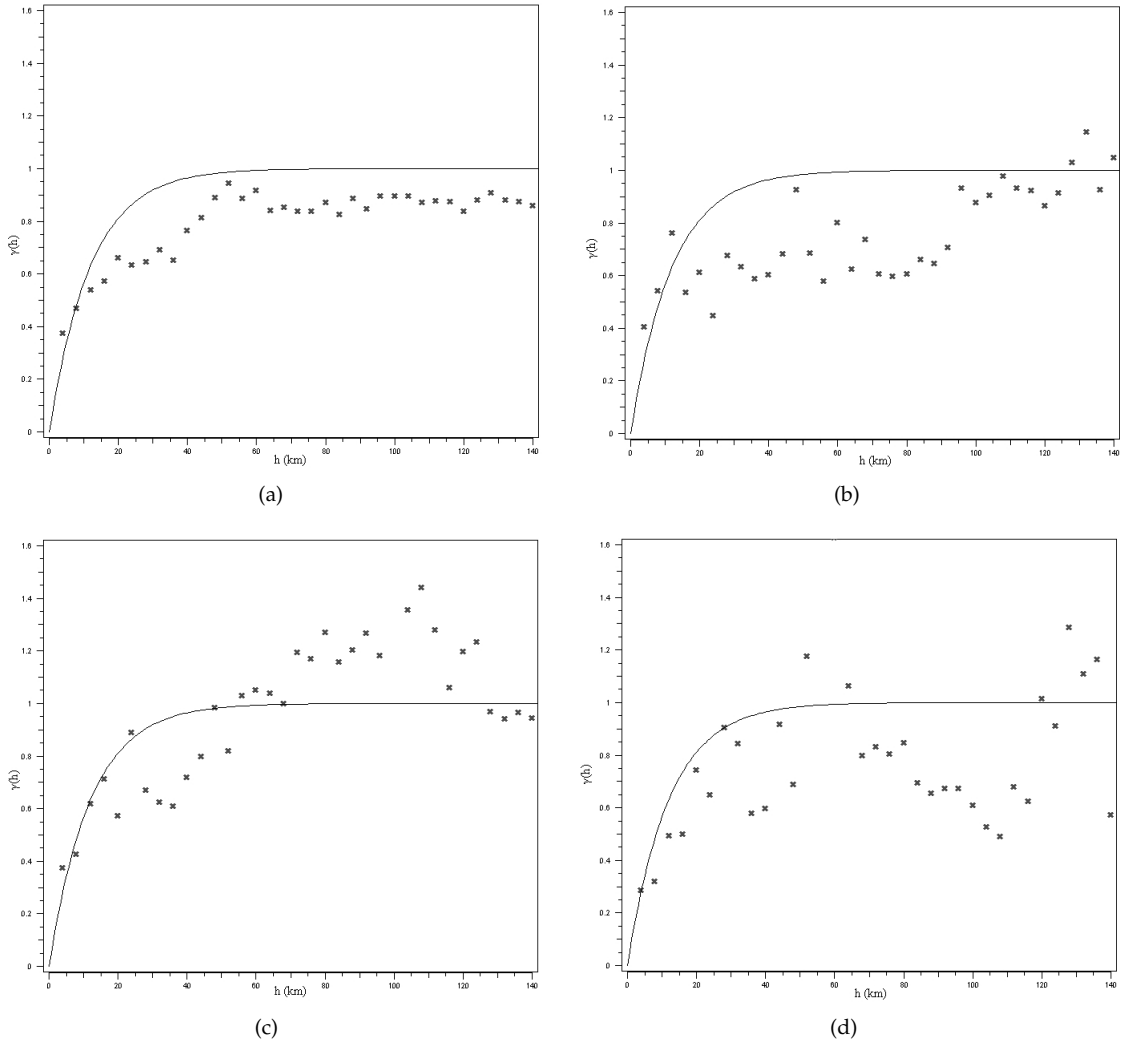

Figure 50: Experimental directional semivariograms at discrete separations obtained using the Chi-Chi earthquake $\tilde{\varepsilon}$ values computed at 1 second. Also shown in the figures is the best fit to the omni-directional semivariogram: (a) Omni-directional; (b) Azimuth = 0; (c) Azimuth = 45 and (d) Azimuth = 90

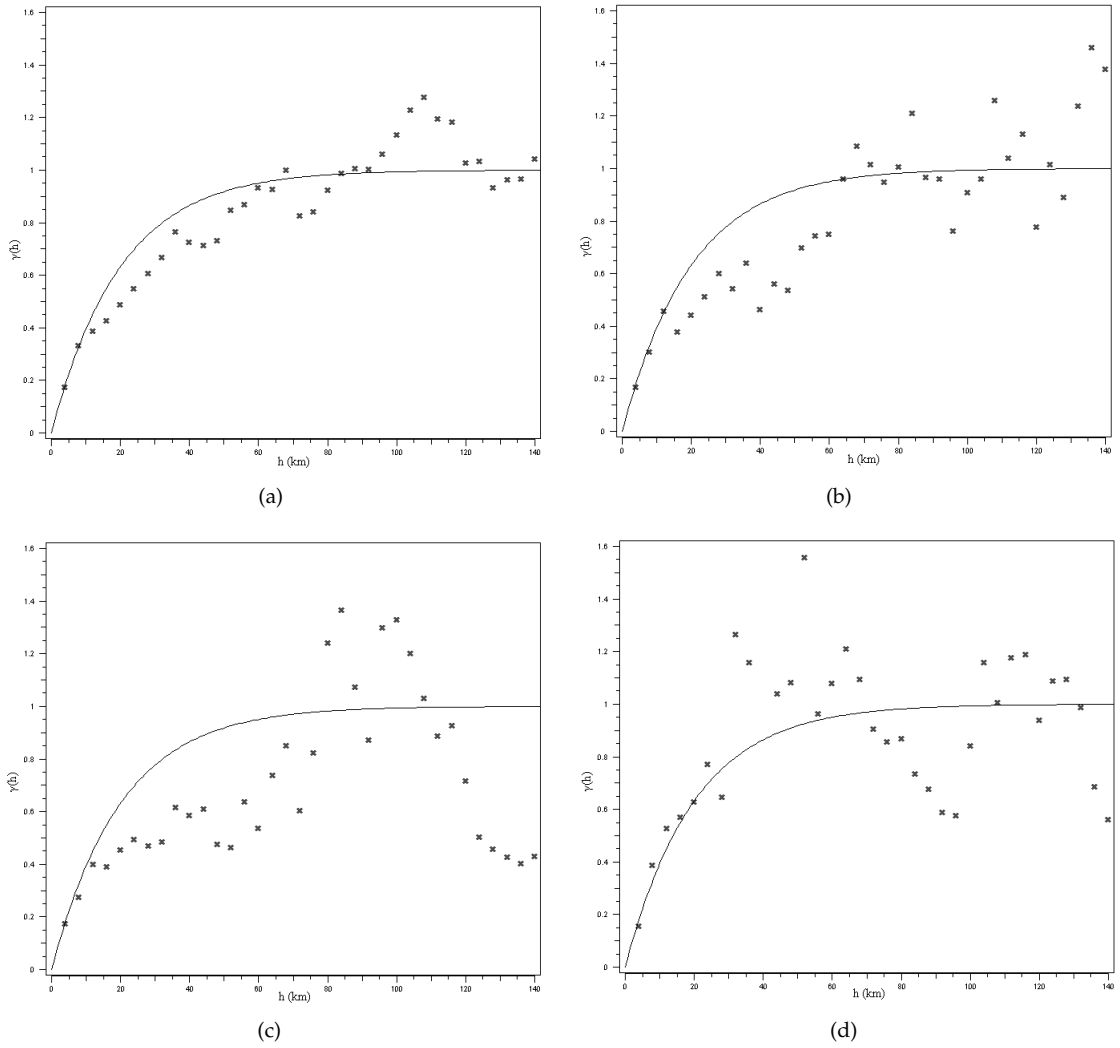

Figure 51: Experimental directional semivariograms at discrete separations obtained using the Chi-Chi earthquake $\tilde{\varepsilon}$ values computed at 7.5 seconds. Also shown in the figures is the best fit to the omni-directional semivariogram: (a) Omni-directional; (b) Azimuth = 0; (c) Azimuth = 45 and (d) Azimuth = 90

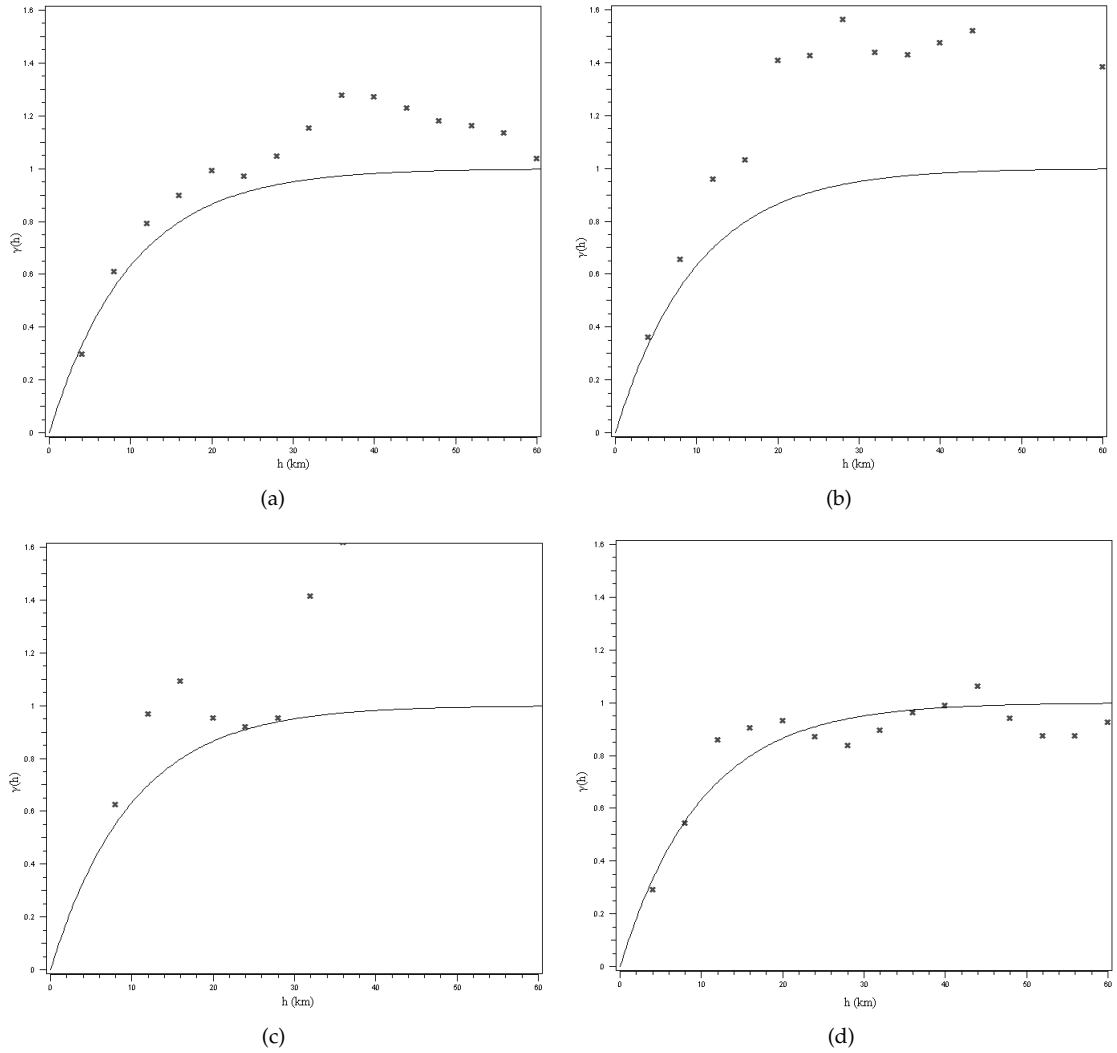

Figure 52: Experimental directional semivariograms at discrete separations obtained using the simulated time histories. The $\tilde{\varepsilon}$ values are computed at 2 seconds. Also shown in the figures is the best fit to the omni-directional semivariogram: (a) Omni-directional; (b) Azimuth = 0; (c) Azimuth = 45 and (d) Azimuth = 90

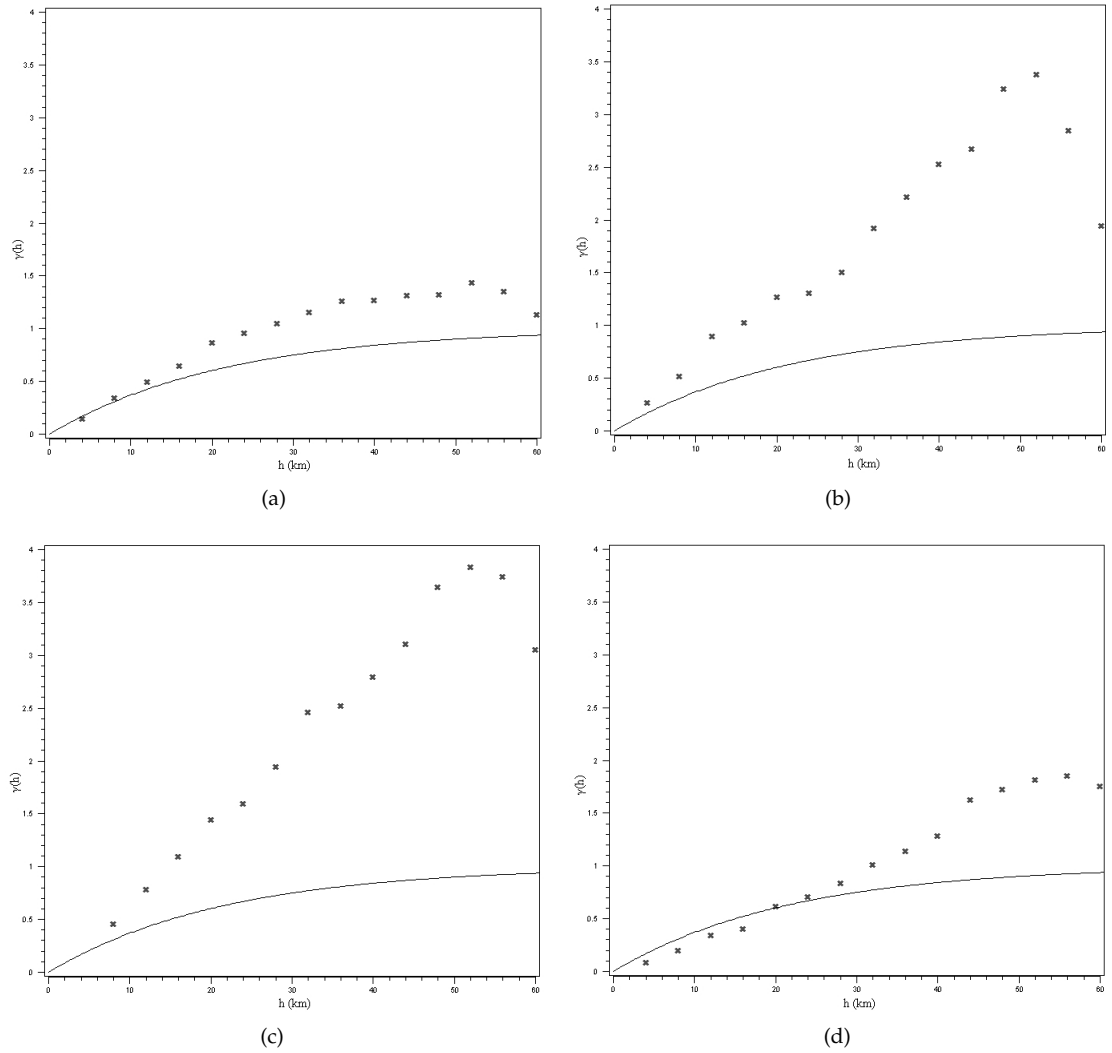

Figure 53: Experimental directional semivariograms at discrete separations obtained using the simulated time histories. The $\tilde{\varepsilon}$ values are computed at 7.5 seconds. Also shown in the figures is the best fit to the omni-directional semivariogram: (a) Omni-directional; (b) Azimuth = 0; (c) Azimuth = 45 and (d) Azimuth = 90

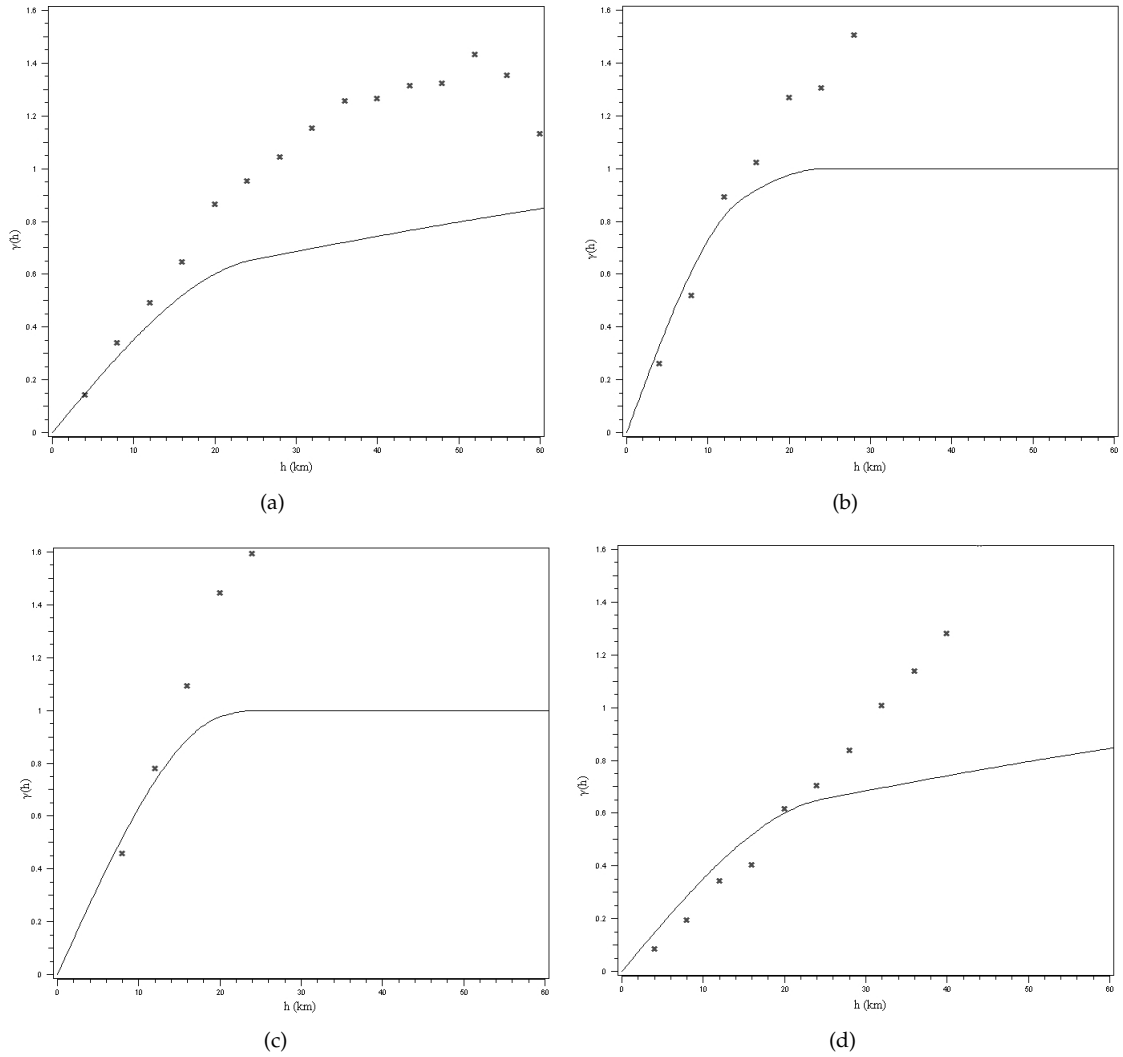

Figure 54: Experimental directional semivariograms at discrete separations obtained using the simulated time histories. The $\tilde{\varepsilon}$ values are computed at 7.5 seconds. Also shown in the figures is an anisotropic model that fits the four experimental semivariograms well (It is to be noted that an anisotropic semivariogram has different shapes in different directions.): (a) Omni-directional; (b) Azimuth = 0; (c) Azimuth = 45 and (d) Azimuth = 90

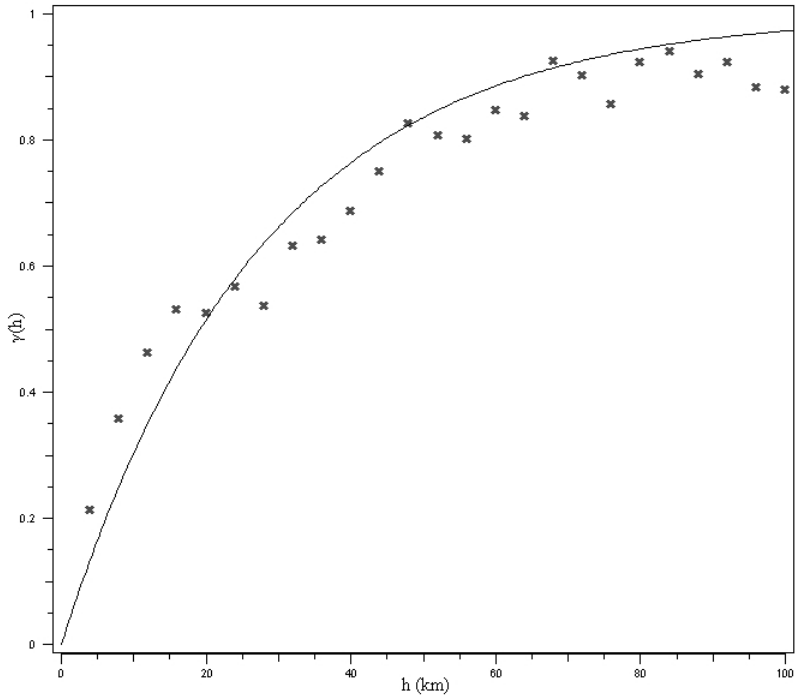

Figure 55: Semivariogram obtained using residuals computed based on Chi-Chi earthquake peak ground velocities. The exponential function has the range reported by Wang and Takada (2005)

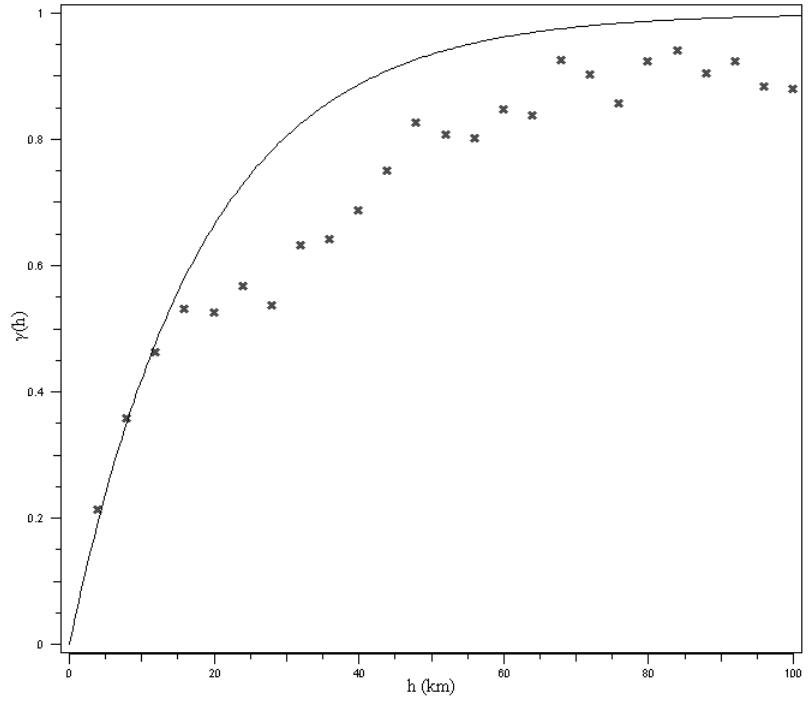

Figure 56: Semivariogram obtained using residuals computed based on Chi-Chi earthquake peak ground velocities. The exponential function is chosen to provide a good match at short separation distances

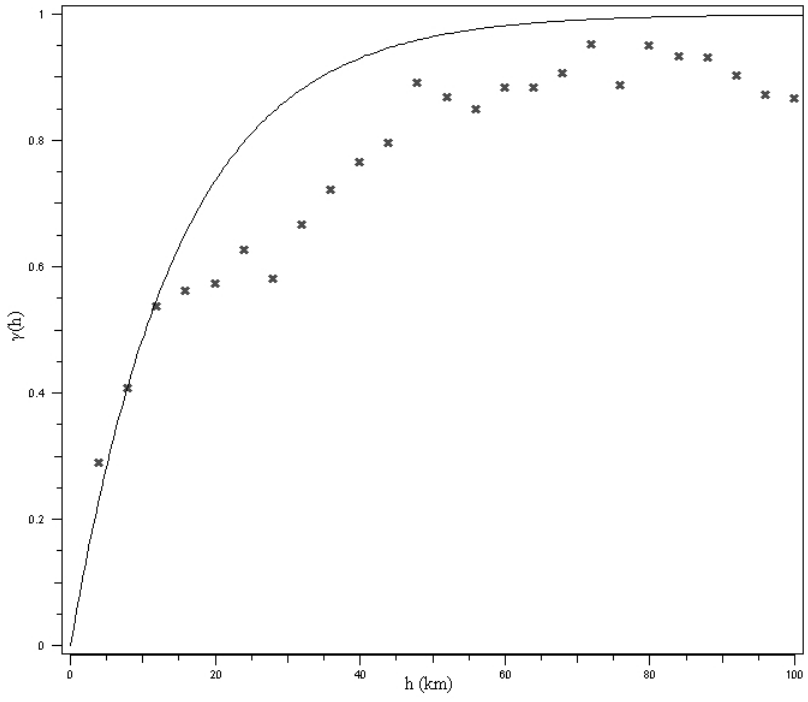

Figure 57: Semivariogram obtained using residuals computed based on Chi-Chi earthquake peak ground velocities. A random amplification factor is applied on the peak ground velocities predicted by the Annaka ground-motion model

References

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

T. Annaka, F. Yamazaki, and F. Katahira. Proposal of peak ground velocity and reponse spectra based on jma 87 type accelerometer records. *Proceedings, 27th JSCE Earthquake Engineering Symposium*, 1: 161–164, 1997.

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

P. Bazzurro and N. Luco. Effects of different sources of uncertainty and correlation on earthquake-generated losses. *Presented at IFED: International Forum on Engineering Decision Making, Stoos, Switzerland*, 2004.

D. M. Boore, J. F. Gibbs, W. B. Joyner, J. C. Tinsley, and D. J. Ponti. 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), 2003.

D. M. Boore, J. Watson-Lamprey, and N. A. Abrahamson. Orientation-Independent measures of ground motion. *Bulletin of the Seismological Society of America*, 96(4):1502–1511, 2006.

D.M. Boore and G.M. Atkinson. Ground-motion prediction equations for the average horizontal component of PGA, PGV and 5% damped SA at spectral periods between 0.01s and 10.0s. *http://peer.berkeley.edu/publications/peer_reports/reports_2007/reports_2007.html (Report PEER 2007/01; last accessed 29 November 2007)*, 2007.

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

R. D. Borcherdy. Estimates of site-dependent response spectra for design (methodology and justification). *Earthquake Spectra*, 10:617–653, 1994.

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

S. Castellaro, F. Mulargia, and P. L. Rossi. Vs30: Proxy for seismic amplification. *Seismological Research Letters*, 79(4):540–543, 2008.

CESMD database. http://www.strongmotioncenter.org (last accessed 6 June 2008), 2008.

B. S-J. Chiou and R. R. Youngs. An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra*, 24(1):173–215, 2008.

A. Der Kiureghian. A coherency model for spatially varying ground motions. *Earthquake engineering and structural dynamics*, 25:99–111, 1996.

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

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.

Pierre Goovaerts. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Oxford, New York, 1997.

R. Graves. Puente Hills scenario earthquake simulations for CEA. **TO BE FILLED**, 2006.

R. Graves. Puente Hills scenario earthquake simulations for CEA. Personal communication, 2007.

T. Hayashi, S. Fukushima, and H. Yashiro. 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.

N. Jayaram and J. W. Baker. Statistical tests of the joint distribution of spectral acceleration values. Bulletin of the Seismological Society of America, in press, 2008.

KiK Net. http://www.kik.bosai.go.jp/ (last accessed 12 December 2007), 2007.

M. H. Kutner, C. J. Nachtsheim, J. Neter, and W. Li. *Applied Linear Statistical Models*. The McGraw-Hill Companies Inc., New York, 2005.

R. Lee and A. S. Kiremidjian. Uncertainty and correlation for loss assessment of spatially distributed systems. *Earthquake spectra*, 23(4):743–770, 2007.

J. Park, P. Bazzurro, and J. W. Baker. Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. *10th International Conference on Application of Statistic and Probability in Civil Engineering (ICASP10), Tokyo, Japan*, 2007.

PEER NGA Database. http://peer.berkeley.edu/nga (last accessed 18 May 2007), 2005.

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

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