

A spatial cross-correlation model of spectral accelerations at multiple periods

Christophe Loth* & Jack W. Baker

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

SUMMARY

Many seismic loss problems (such as disruption of distributed infrastructure and losses to portfolios of structures) are dependent upon the regional distribution of ground-motion intensity, rather than intensity at only a single site. Quantifying ground-motion over a spatially-distributed region therefore requires information on the correlation between the ground-motion intensities at different sites during a single event. The focus of the present study is to assess the spatial correlation between ground-motion spectral accelerations at different periods. Ground motions from eight well-recorded earthquakes were used to study the spatial correlations. Based on obtained empirical correlation estimates, we propose a geostatistics-based method to formulate a predictive model that is suitable for simulation of spectral accelerations at multiple sites and multiple periods, in the case of crustal earthquakes in active seismic regions. While the calibration of this model and investigation of its implications were somewhat complex, the model itself is very simple to use for making correlation predictions. A user only needs to evaluate a simple equation relying on three sets of coefficients provided here, in order to compute a correlation coefficient for spectral values at two periods and at a specified separation distance. These results may then be used in evaluating the seismic risk of portfolios of structures with differing fundamental periods. Copyright © 2010 John Wiley & Sons, Ltd.

Received . . .

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

1. INTRODUCTION

Quantifying ground-motion over a spatially-distributed region rather than at just a single site is of interest for a variety of applications relating to risk of infrastructure or portfolios of properties. This requires information on the correlation between the ground-motion intensities at different sites during a single event. Researchers have previously estimated the correlations between residuals of spectral accelerations at the same spectral period at two different sites. But very few [1] have studied cross-correlations between residuals of spectral accelerations at different periods (or more generally between residuals of two different ground-motion intensity measures) at two different sites, which becomes important, for instance, when assessing the risk of a portfolio of buildings with different fundamental periods [2].

This research relies on the general framework of ground-motion models [3–6], that are defined as follows: for an earthquake j at a site i ,

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

*Correspondence to: Christophe Loth, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, U.S.A. E-mail: lothc@stanford.edu

Contract/grant sponsor: U.S. Geological Survey (USGS); contract/grant number: G10AP00046

where Y_{ij} refers to the ground-motion parameter of interest (e.g. $S_a(T)$ the spectral acceleration at period T); \bar{Y}_{ij} denotes the predicted (by the ground-motion model) median ground-motion intensity, a function of various parameters such as magnitude, distance, period and local-site conditions; ϵ_{ij} refers to the intra-event residual, a random variable of mean zero and standard deviation one; and η_j denotes the inter-event residual, also a random variable of mean zero and standard deviation one. The standard deviations σ_{ij} and τ_j are included in the ground-motion model prediction and depend on the spectral period of interest (in some models, they are also a function of the earthquake magnitude and the distance of the site from the rupture). For a given earthquake j , the inter-event residual η_j computed at any particular period is a constant across all the sites.

Previous studies have established that a vector of spatially distributed intra-event residuals $\epsilon_j = (\epsilon_{1j}, \epsilon_{2j}, \dots, \epsilon_{nj})$ follows a multivariate normal distribution [7]. Consequently, one can fully define the ϵ_j by specifying its mean vector and the covariance between all pairs of ϵ_{ij} 's. In our particular case, the mean vector of ϵ_j is 0 and hence we only need to know the variance-covariance matrix: for a given earthquake j ,

$$\Sigma(\text{event } j) = \begin{bmatrix} \text{cov}(\epsilon_{1j}, \epsilon_{1j}) & \cdots & \text{cov}(\epsilon_{1j}, \epsilon_{nj}) \\ \vdots & \ddots & \vdots \\ \text{cov}(\epsilon_{nj}, \epsilon_{1j}) & \cdots & \text{cov}(\epsilon_{nj}, \epsilon_{nj}) \end{bmatrix} \quad (2)$$

where $\text{cov}(\epsilon_{kj}, \epsilon_{lj})$ is the covariance between ϵ_{kj} , the residual at site k due to earthquake j , and ϵ_{lj} , the residual at site l due to earthquake j .

Spatial modeling of earthquake intensities has been investigated in the past by various researchers. For instance, the recent work of Foulser-Piggott and Stafford [8] aims at modeling the spatial correlation for $Y = I_\alpha$, the Arias intensity. Esposito and Iervolino [9] examined the spatial correlation of PGA and PGV based on European earthquake data. The modeling of the spatial correlation of the residuals of a single spectral acceleration period $Y = S_a(T)$ has also been addressed in previous contributions [1, 10–13], leading to predictive equations for the correlation coefficient as a function of the period of interest and the separation distance between two considered sites. The present work will generalize the modeling to a multivariate framework that accounts for several intensity measures $\{Y_i = S_a(T_i) | i = 1, \dots, n\}$, where less study has been done.

This study will begin with a presentation of geostatistical concepts relevant to the spatial modeling of correlations. Section 3 will describe our first attempt at using empirical data to estimate these correlations, to illustrate the encountered limitations and issues. An alternative technique will then be introduced in Section 4 as an improved solution, and a predictive model for covariance will be derived using it. Finally, as Goda and Hong [1] proposed to use the single period result combined with a Markov-type hypothesis to formulate a correlation model for the multi-period case, we will also evaluate this hypothesis in the last section, in comparison with the formulated model. In addition to the results presented here, supporting information is available in a related project report [14].

In this work, we used recorded ground-motion data from the Pacific Earthquake Engineering Research (PEER) Center's NGA database (http://peer.berkeley.edu/products/strong_ground_motion_db.html).

This database consists of recordings from many data centers around the world, as documented at the above referenced website. All conclusions are therefore applicable primarily for the crustal earthquakes in active seismic regions represented in that database. Figure 1 shows plots of the station locations for eight earthquakes which have been considered in this study: Northridge, Chi-Chi, Tottori, Niigata, Parkfield, Chuetsu, Iwate, El Mayor Cucapah. In this study, we used the Boore and Atkinson ground motion prediction model [3]. Furthermore, only recordings with a Joyner-Boore distance R_{jb} smaller than 200 km and within the lowest usable frequency limit were considered. Table I provides additional data regarding the selected earthquakes.

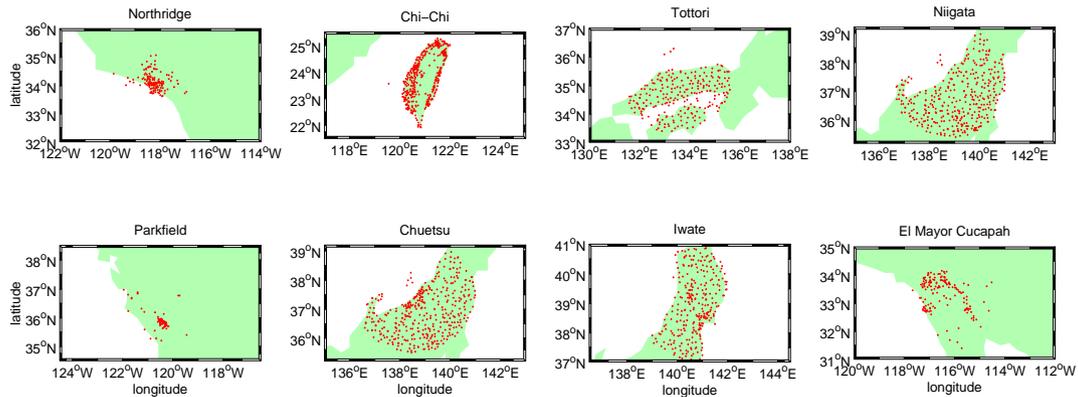


Figure 1. Locations of recordings from the eight considered earthquakes.

Table I. The eight considered earthquakes in the spatial correlation study

Name	Location	Year	Magnitude	Number of recordings
Northridge	California	1994	6.7	152
Chi-Chi	Taiwan	1999	7.6	401
Tottori	Japan	2000	6.6	235
Niigata	Japan	2004	6.6	365
Parkfield	California	2004	6.0	90
Chuetsu	Japan	2007	6.8	403
Iwate	Japan	2008	6.9	280
El Mayor Cucapah	California	2010	7.2	154

2. GEOSTATISTICAL MODELING OF CORRELATIONS

Spatially distributed random variables are often described by a semivariogram, which is a very popular tool in the geostatistics domain [15, 16]. The semivariogram is a so-called two-point statistic that characterizes the spatial decorrelation or dissimilarity. Its general formulation is given below for a pair of locations \mathbf{u} , \mathbf{u}' :

$$\gamma(\mathbf{u}, \mathbf{u}') = \frac{1}{2} \mathbb{E} \left[(Z(\mathbf{u}) - Z(\mathbf{u}'))^2 \right] \quad (3)$$

where $Z(\mathbf{u})$ is a random variable representing the value of interest at location \mathbf{u} , $\mathbb{E}[\cdot]$ denotes the expectation, and $\gamma(\mathbf{u}, \mathbf{u}')$ is the semivariogram value.

Since one often does not possess several observations of a random variable at a given pair of sites, the assumption of stationarity has to be made in order to evaluate Equation 3: one will typically retain that the semivariogram does not depend on the site locations $(\mathbf{u}, \mathbf{u}')$ but only on their separation vector $\mathbf{h} = \mathbf{u} - \mathbf{u}'$. Thus, for a stationary random variable Z , for instance $Z = \epsilon(T)$, the semivariogram is defined as follows:

$$\gamma(\mathbf{h}) = \frac{1}{2} \mathbb{E} \left[(Z(\mathbf{u} + \mathbf{h}) - Z(\mathbf{u}))^2 \right] \quad (4)$$

This semivariogram function can be empirically estimated with:

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha + \mathbf{h}) - z(\mathbf{u}_\alpha)]^2 \quad (5)$$

where z denotes an observation from the random variable Z , \mathbf{u}_α a recording location from the data, and $N(\mathbf{h})$ the number of pairs at separation vector \mathbf{h} available in the data.

Previous studies have indicated that the correlation structure of residuals from ground motion models was not dependent on the considered direction, and was therefore isotropic [1, 11, 13, 17]. This translates in Equations 4 and 5 by simply “removing” all vector notations, so that $h = \|\mathbf{h}\|$. It should also be noted that in the practical computation of the semivariogram with Equation 5, it is unlikely that two data points will be separated by the exact distance h . Therefore, a tolerance parameter Δ will have to be considered such that for a given lag distance h , all the pairs of points separated by a distance included in the interval $[h - \Delta, h + \Delta]$ will contribute to the evaluation of the empirical semivariogram $\gamma(h)$.

Furthermore, the covariance function can be defined as:

$$C(h) = \text{cov}(Z(u), Z(u+h)) = \mathbb{E}[(Z(u) - m)(Z(u+h) - m)] \quad (6)$$

where m is the mean of $Z(u)$ (and is also equal to the mean of $Z(u+h)$ under the stationarity hypothesis). This spatial covariance is directly related to the semivariogram function with:

$$C(h) = C(0) - \gamma(h) \quad (7)$$

Similarly, it can be noted that the correlation coefficient is defined as:

$$\rho(h) = \frac{C(h)}{C(0)} \quad (8)$$

Thus, semivariogram and covariance have “opposite” behaviors: the covariance is a measure of spatial similarity between $Z(u)$ and $Z(u+h)$. While one could conduct a covariance study on either one of those functions, the semivariogram is often preferred in geostatistical practice, as it does not require a prior estimation of the mean of the random field m .

In this research, we consider the cross-covariance structure of residuals of spectral acceleration at multiple periods. This means that one needs to extend the previous definitions to the multivariate case in order to estimate all spatial cross-correlation terms between $\epsilon(T_i)$ and $\epsilon(T_j)$, $T_i \neq T_j$. First, the definition of the semivariogram can easily be generalized to the multivariate case. Denoting two stationary random variables $Z_1 = \epsilon(T_1)$ and $Z_2 = \epsilon(T_2)$, one defines their cross-semivariogram:

$$\gamma_{12}(\mathbf{h}) = \mathbb{E}[(Z_1(\mathbf{u} + \mathbf{h}) - Z_1(\mathbf{u}))(Z_2(\mathbf{u} + \mathbf{h}) - Z_2(\mathbf{u}))] \quad (9)$$

which may again be empirically evaluated with:

$$\gamma_{12}(\mathbf{h}) = \frac{1}{2N_{12}(\mathbf{h})} \sum_{\alpha=1}^{N_{12}(\mathbf{h})} [(z_1(\mathbf{u}_\alpha + \mathbf{h}) - z_1(\mathbf{u}_\alpha))(z_2(\mathbf{u}_\alpha + \mathbf{h}) - z_2(\mathbf{u}_\alpha))] \quad (10)$$

with $N_{12}(\mathbf{h})$ is the minimum number of available data pairs in the data between Z_1 and Z_2 (the difference in the numbers of data for a period pair is due to the existence of a lowest usable frequency for a given earthquake recording). Again, it should be noted that the previously explained definitions 9 and 10 are true only for second order stationary random variables, meaning:

$$\begin{cases} \mathbb{E}[Z_i(\mathbf{u})] = m_i & \forall i \in 1, \dots, n, \text{ and for all locations } \mathbf{u} \\ \mathbb{E}[(Z_i(\mathbf{u}) - m_i)(Z_j(\mathbf{u} + \mathbf{h}) - m_j)] = C_{ij}(\mathbf{h}) & \forall i, j \in 1, \dots, n, \text{ and for all locations } \mathbf{u}, \mathbf{u} + \mathbf{h} \end{cases} \quad (11)$$

Equation 7 will extend in the multivariate case, by defining the isotropic semivariogram matrix function $\Gamma(h)$:

$$\mathbf{\Gamma}(h) = [\gamma_{ij}(h)] = \begin{bmatrix} \gamma_{11}(h) & \cdots & \gamma_{1n}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(h) & \cdots & \gamma_{nn}(h) \end{bmatrix} \quad (12)$$

Similarly, one denotes the isotropic covariance matrix function $\mathbf{C}(h)$ as follows:

$$\mathbf{C}(h) = [C_{ij}(h)] = \begin{bmatrix} C_{11}(h) & \cdots & C_{1n}(h) \\ \vdots & \ddots & \vdots \\ C_{n1}(h) & \cdots & C_{nn}(h) \end{bmatrix} \quad (13)$$

where $C_{ij}(h)$ is the covariance function between $Z_i = \epsilon(T_i)$ and $Z_j = \epsilon(T_j)$. With these notations, one obtains the following relationship:

$$\mathbf{C}(h) = \mathbf{C}(0) - \mathbf{\Gamma}(h) \quad (14)$$

These concepts will be the basis of the quantification of the spatial correlation between pairs of various spectral accelerations, presented in the following sections.

3. DIRECT SEMIVARIOGRAM FIT OF EMPIRICAL DATA

3.1. Computation of residuals

For each of the 8 selected earthquakes in this study, residuals were derived from Equation 1:

$$\epsilon(T) = \frac{\ln Sa(T) - \ln \overline{Sa}(T)}{\sigma(T)} \quad (15)$$

with $\ln Sa(T)$ the recorded logarithmic spectral acceleration at period T , $\overline{Sa}(T)$ and $\sigma(T)$ respectively the median spectral acceleration and logarithmic standard deviation from the ground motion model. Note that this computation does not take into account the inter-event term, since it is constant for a given earthquake and therefore does not affect the actual correlation structure of the intra-event residuals (a proof of this statement can be found in [13]).

3.2. Fitting technique

Geostatistics literature recommends a manual fit of the semivariogram, warning against regression methods that might misrepresent the actual information provided by the semivariogram [15]. Each estimated point of the semivariogram $\gamma(h)$ is subject to an error inherent to that point. This error will vary with the considered separation distance h , the extent of the region used in the semivariogram calculation, etc. For these reasons, semivariogram fitting cannot be reduced to a simple regression problem. However, given the quantity of data to be analyzed in the multivariate case (we consider 9 periods and 8 earthquakes, resulting in 360 different semivariograms), it was deemed reasonable to develop an automated fitting algorithm to speed up the process, as long as the result of the fit was consistent with independently obtained manual fits.

Not any function can be chosen to fit an empirical semivariogram. The covariance function, directly related to the semivariogram through Equation 7, must be positive definite. This is due to the fact that the variance of any linear combination of Z values must be non-negative; in other words, for any set of locations \mathbf{u}_α and any set of weights ω_α , the covariance function must satisfy:

$$\text{var} \left(\sum_{\alpha=1}^p \omega_\alpha Z(\mathbf{u}_\alpha) \right) = \sum_{\alpha=1}^p \sum_{\beta=1}^p \omega_\alpha \omega_\beta C(\|\mathbf{u}_\alpha - \mathbf{u}_\beta\|) \geq 0 \quad (16)$$

where $\text{var}()$ denotes the variance.

In practice, one ensures positive definiteness by modeling a semivariogram with a positive linear combination of admissible semivariogram models. These admissible models include, but are not limited to, the four following models. The exponential model is defined as:

$$\tilde{\gamma}(h) = S \left[1 - \exp \left(\frac{-3h}{R} \right) \right] \quad (17)$$

where $\tilde{\gamma}$ refers to the value from a model, S is the sill and R is the range of the semivariogram. The sill of a bounded semivariogram is equal to the variance of Z ; for the exponential semivariogram, it represents the value to which $\tilde{\gamma}(h)$ asymptotically converges as h tends to infinity. The range is then defined as the separation distance h at which $\tilde{\gamma}(h)$ is equal to 95% of the sill of the exponential semivariogram. This means that the range represents the distance at which 95% of the correlation is

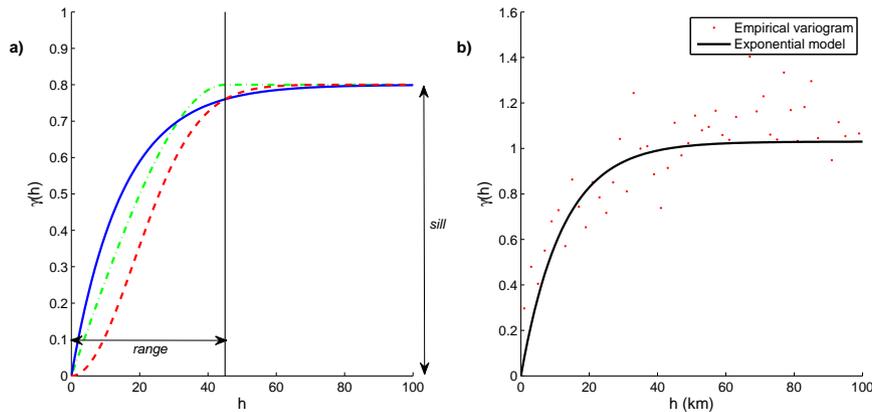


Figure 2. a) Spherical, exponential and Gaussian semivariograms with $S = 0.8$ and $R = 45$. For the spherical model, the range represents the distance at which 100% of the correlation is lost, whereas for the Gaussian and exponential semivariograms, it represents the distance at which 95% of the correlation is lost; b) Empirical semivariogram and fitted exponential model of the normalized residuals (ϵ) from the Northridge earthquake data, at $T_1 = T_2 = 1$ s.

lost. The spherical model is defined as:

$$\tilde{\gamma}(h) = \begin{cases} S \left[\frac{3h}{2R} - \frac{1}{2} \left(\frac{h}{R} \right)^3 \right] & \text{if } h \leq R \\ S & \text{otherwise} \end{cases} \quad (18)$$

With this model, the sill S is attained at $h = R$. The third common semivariogram model is the Gaussian model:

$$\tilde{\gamma}(h) = S \left[1 - \exp \left(\frac{-3h^2}{R^2} \right) \right] \quad (19)$$

The sill and the range of the Gaussian semivariogram are defined as for the exponential variogram. Finally, the nugget effect model is defined as:

$$\tilde{\gamma}(h) = \begin{cases} 0 & \text{if } h = 0 \\ S & \text{if } h > 0 \end{cases} \quad (20)$$

This semivariogram induces a complete lack of correlation at non-zero separation distance, therefore no range can be defined for the nugget effect. The first three semivariogram models are shown in Figure 2a. The entire correlation structure of the variables of study will be completely defined by the semivariogram model, which itself depends only on the corresponding sills and ranges. In this work, we first assumed that each cross-semivariogram γ_{ij} associated with $\epsilon(T_i)$ and $\epsilon(T_j)$ can be modeled with an isotropic exponential function, such that:

$$\tilde{\gamma}_{ij}(h) = S_{ij} \left[1 - \exp \left(\frac{-3h}{R_{ij}} \right) \right] \quad (21)$$

where S_{ij} is the sill and R_{ij} the range. Figure 2b shows the isotropic semivariogram function computed for data from the Northridge earthquake, and a fitted exponential model. This choice is motivated by results obtained by researchers in the past [11, 13], who observed an exponential decay of the correlation coefficient in the univariate case. Indeed, we know from Equation 7 that the semivariogram and the covariance are related as follows:

$$\gamma_{ij}(h) = C_{ij}(0) - C_{ij}(h) \quad (22)$$

and by extending Equation 8 to the multivariate case, we obtain the following model for the correlation coefficient:

$$\tilde{\rho}_{ij}(h) = \rho_{ij}(0) \exp\left(\frac{-3h}{R_{ij}}\right) \quad (23)$$

where $\rho_{ij}(0) = C_{ij}(0) = S_{ij}$ if the considered variables have unit standard deviations. It can be noted that other functional forms for the correlation coefficient have been used, such as the more general $\tilde{\rho}(h) = \exp(-\alpha h^\beta)$ by Goda and Hong [1], where α and β are constants (for $\beta = 1$, this model is equivalent to the exponential functional form). Boore et al. [10] used a particular case of this model with $\beta = 0.5$.

Previous studies have proposed empirical equations to predict the sill [7, 12, 18, 19], as it is equal to the covariance between $\epsilon(T_i)$ and $\epsilon(T_j)$ at the same site ($h = 0$):

$$S_{ij} = C_{ij}(0) \quad (24)$$

Three methods were investigated in order to achieve a robust estimation of the sill:

- (i) a direct computation of $C_{ij}(0)$ of the empirical data;
- (ii) calculating the mode of the histogram of the semivariogram values themselves;
- (iii) a refinement of (ii) using a Gaussian kernel function.

The last approach proved to be the most robust one, and it has been retained in lieu of the predictive model. Indeed, when fitting a least squares regression, it is critical to assess as correctly as possible the value of the sill, in order to achieve a correct estimate of the range.

The empirical estimation of S with a kernel function relies on a discretization of the observed semivariogram values, followed by a computation of a kernel weighted function:

$$\begin{cases} y_0 = 0, y_1 = 0.01, \dots, y_i = 0.01 i, \dots, y_{100} = 1 \\ \text{kernel}(i) = \sum_{k=1}^{k_{max}} \exp\left(-\frac{(\gamma(h_k) - y_i)^2}{\sigma}\right), h_{k_{max}} = 100 \text{ km}, \sigma = \text{constant} \\ S = y_{i_0} \text{ s.t. } \max_i(\text{kernel}(i)) = \text{kernel}(i_0) \end{cases} \quad (25)$$

Variations of the constant σ did not have a significant impact on the final result of the sill value. In this study, a value of $\sigma = 0.1$ has been used.

Once S is accurately determined, the range can be derived using weighted least squares regression. With the exponential semivariogram model, the problem can be linearized using a log transformation:

$$\tilde{\gamma}(h) = S \left[1 - \exp\left(\frac{-3h}{R}\right) \right] \Rightarrow \ln(S - \tilde{\gamma}(h)) = ah + b \text{ with } \begin{cases} a = \frac{-3}{R} \\ b = \ln S \end{cases} \quad (26)$$

The regression algorithm will evaluate the weighted sum of squares, as a function of the range R :

$$\begin{aligned} WSS(R) &= \sum_k \omega(h_k) [\ln(S - \min\{\gamma(h_k), S - 0.01\}) - \ln(S - \tilde{\gamma}(h_k))]^2 \\ &= \sum_k \frac{1}{h_k} \left[\ln(S - \min\{\gamma(h_k), S - 0.01\}) - \left(\left(\frac{-3}{R} \right) h_k + \ln S \right) \right]^2 \end{aligned} \quad (27)$$

where $\omega(h_k)$ is a weighting function giving more importance to the smallest separation distances (an inverse distance weighting has been used here, such that the weight on the error at lag h_k is equal to $1/h_k$). The semivariogram values are constrained to be less than $S - 0.01$ in order for the logarithm function to be well-defined. The value of R yielding the minimum of this WSS is retained as the range of the experimental semivariogram. It should be noted that this regression bears no intrinsic fundamental or theoretical meaning; it merely is a convenient way to obtain satisfying fits for a large amount of semivariograms.

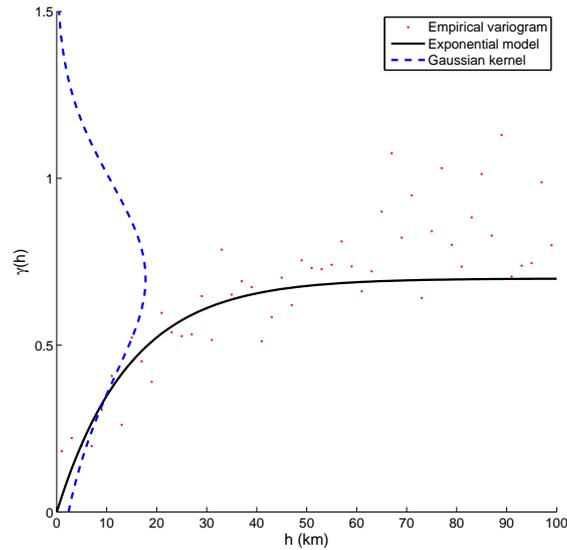


Figure 3. Direct cross-semivariogram fitting using the Northridge earthquake data for $T_1 = 1s$ and $T_2 = 2.5s$. The sill is determined as the y-axis value at which the Gaussian kernel attains its maximum.

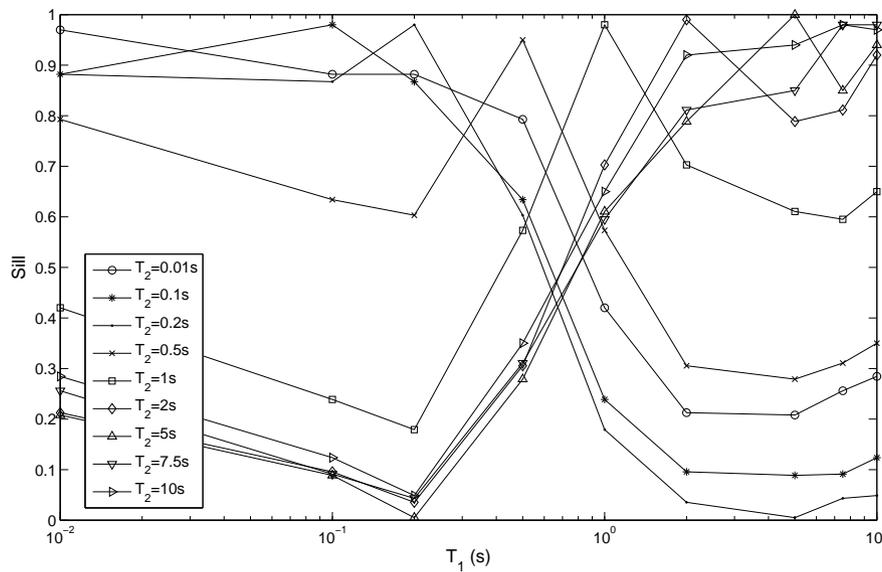


Figure 4. Sills of the cross-semivariograms obtained with the direct semivariogram fitting technique for the Northridge earthquake. Each line attains its maximum when $T_1 = T_2$.

3.3. Observed results

Figure 3 shows the result of the kernel fitting for the cross-semivariogram between $T_1 = 1s$ and $T_2 = 2.5s$ from the Northridge earthquake. The fitted semivariogram proves to be a good match with the data while also representing a likely outcome of a manual fitting. The kernel fitting provides accurate estimates of the sills for all period pairs of interest, as shown in Figure 4. It should be noted that the sills at $T_1 = T_2$ are theoretically 1, but approximations in both the semivariogram values and the fitted semivariogram lead to results that are not precisely unity. Results for cross-

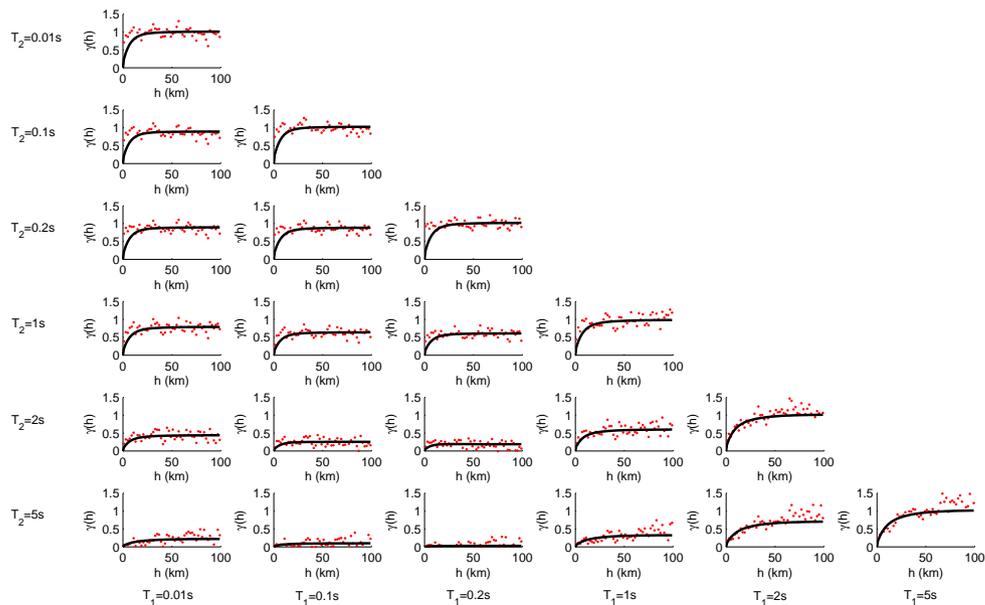


Figure 5. Direct cross-semivariogram fits for six pairs of periods from the Northridge earthquake data.

semivariograms for several period pairs obtained with this approach are shown in Figure 5. However, numerical instabilities may be encountered with residuals having low correlation, e.g. between ϵ 's with very short and very long periods. These cross-semivariograms are often just "noise" with an almost zero sill and lead to irrelevant estimates of the range due to a poor convergence of the least squares regression. Thus, raw results of the direct semivariogram fit have to be filtered, by dismissing the data leading to low sills. After filtering the results, we could observe clusters of data in the range versus sill plane, where the observations could be distinguished with respect to the value of the corresponding period pairs. It appeared that cross-semivariograms for two long periods (approximately, $T_1 > 1s$, $T_2 > 1s$) have a higher range of around 40 km, while they show a shorter range of approximately 25 km for two short periods ($T_1 < 1s$, $T_2 < 1s$). The presence of two structures of different range is indicated in Figure 6, which show both short and long range components. Furthermore, the very low ranges obtained at short periods (Figure 5) tend to indicate the presence of a so-called "nugget effect", which means that a more appropriate semivariogram function in this case might be the one from Equation 20. Indeed, the nugget semivariogram is nothing but a limiting case of an exponential semivariogram whose range tends to zero.

The 1999 Chi-Chi earthquake provides many recordings and thus is one of the most useful events for this study. The direct semivariogram fitting technique gives adequate representation of the data. The results did not show two different spatial structures as was the case for the Northridge earthquake, but longer semivariogram ranges were noticed in average, meaning that the correlation between spectral accelerations generally holds for longer distances.

Similar work was performed for the aforementioned six other earthquakes. The same quality of fit could be observed, although the empirical semivariograms did not demonstrate such a clear exponential trend in some cases, possibly due to the relative lack of data (e.g., for Parkfield and El Mayor). While a simple average of the sills and ranges over all earthquakes may be proposed for the development of a predictive model, some limitations prevent the use of these results for the formulation of the covariance matrix.

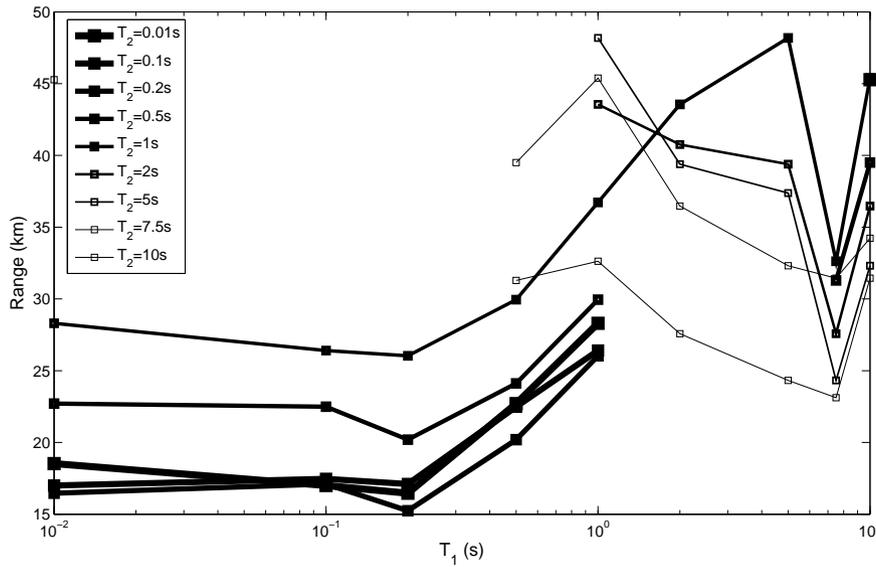


Figure 6. Filtered ranges of the cross-semivariograms from the Northridge earthquake data (9 periods were considered).

3.4. Limitations

The direct semivariogram fit developed in this study proved to be a useful tool to evaluate the spatial correlation of the empirical data. One may very well use these results to estimate any correlation coefficient between spectral acceleration at two different periods at two different sites. However, a more general objective of this study was to formulate a predictive model for the covariance matrix of a given set of ϵ 's, based on these estimations and Equation 14.

However, for \mathbf{C} to be an acceptable covariance matrix, the same condition of positive definiteness as in the univariate case (see Equation 16) must be satisfied: the variance of any weighted linear combination of n variables at p sites must be non-negative. This results in the following requirement for the multivariate case [20]:

$$\text{var} \left(\sum_{i=1}^n \sum_{\alpha=1}^p \omega_{\alpha}^i Z_i(\mathbf{u}_{\alpha}) \right) = \sum_{i=1}^n \sum_{j=1}^n \sum_{\alpha=1}^p \sum_{\beta=1}^p \omega_{\alpha}^i \omega_{\beta}^j C_{ij} (\|\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}\|) \geq 0 \quad (28)$$

where ω_{α}^i is the weight associated with the value of Z_i at location \mathbf{u}_{α} . Unfortunately, the direct semivariogram fitting approach described above takes no such constraint into account when evaluating empirical sills and ranges, and thus will not lead to a positive definite covariance matrix in most cases. It is possible to “fix” this matrix by merely changing its eigenvalues to make it positive definite ([21]: section 6.2). This is achieved by performing an eigenvalue decomposition of \mathbf{C} , such that:

$$\mathbf{C} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \quad (29)$$

where each column of \mathbf{Q} is the eigenvector \mathbf{q}_i of \mathbf{C} and $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues, i.e., $\Lambda_{ii} = \lambda_i$. The eigenvalue matrix $\mathbf{\Lambda}$ is then transformed into $\mathbf{\Lambda}^+$ by changing the negative coefficients to 0 :

$$\Lambda_{ii}^+ = \begin{cases} \lambda_i & \text{if } \lambda_i \geq 0 \\ 0 & \text{if } \lambda_i < 0 \end{cases} \quad (30)$$

Finally, $\mathbf{\Lambda}^+$ is recombined with the eigenvector matrix \mathbf{Q} to obtain the positive definite matrix \mathbf{C}^+ :

$$\mathbf{C}^+ = \mathbf{Q}\mathbf{\Lambda}^+\mathbf{Q}^T \quad (31)$$

We have observed that only minor changes need to be made to the fitted covariance matrix in order to transform it into a positive definite one (i.e., $\mathbf{C}^+ \approx \mathbf{C}$).

While modifying the eigenvalues is relatively easy to do, it does not allow much control on how much the covariance matrix will be changed. It also makes it difficult to access the “new” actual values of the ranges of the different cross-semivariograms. Another approach involves the computation of cross-covariance terms from convolution integrals of the direct covariances [22], such that $C_{ij}(\mathbf{h}) = \int_{\mathbb{R}^2} \rho_i(\mathbf{u})\rho_j(\mathbf{u} + \mathbf{h})d\mathbf{u}$. Although this approach will lead to valid models, it makes it quite difficult to fit the cross-covariance coefficients to empirical data.

An alternative solution to ensure positive definiteness is to impose a single range for all direct and cross-semivariograms. The covariance matrix function will become:

$$\mathbf{C}(h) = \rho(h) \cdot \mathbf{C}(0) \quad (32)$$

where $\rho(h)$ is a scalar function (for instance in the present case, $\rho(h) = \exp(-3h/R)$). This formulation of the covariance matrix function is called the separable model [23]. For the full covariance matrix to be positive definite in this case, one has only to ensure that the covariance matrix at a single site $\mathbf{C}(0)$ is positive definite. This is a much simpler task than ensuring the full covariance for the number of periods n times the number of sites p to be positive definite, since $\mathbf{C}(0)$ is only defined for the number of periods n . Unfortunately, fitting a single range to the data is not possible, as it does not reflect the underlying structures discovered in this section. The next section will introduce an extension of this separable model that can incorporate more than one range.

4. THE LINEAR MODEL OF COREGIONALIZATION

While independently fitting each empirical semivariogram may not provide an admissible correlation model, it does give some insight into the spatial characteristics of the considered variables. From the direct semivariogram fits developed for the Northridge earthquake especially, we noticed clear contributions of three different structures: a short range component acting on small periods and a large range component acting on longer periods, as well as a nugget effect for very short periods. To take the effect of multiple spatial scales into account, a model was considered that assumes all variables to be linear combinations of the same basic structural components. Analytically, for a given set of n mean-zero random variables (Z_1, Z_2, \dots, Z_n) [15]:

$$Z_i(\mathbf{u}) = \sum_{l=0}^L \sum_{k=1}^{n_l} a_{ik}^l Y_k^l(\mathbf{u}) \quad \forall i = 1, \dots, n \quad (33)$$

with

- $\mathbb{E}[Y_k^l(\mathbf{u})] = 0$
- $\text{cov}(Y_k^l(\mathbf{u}), Y_{k'}^{l'}(\mathbf{u} + \mathbf{h})) = \begin{cases} c_l(h) & \text{if } k = k' \text{ and } l = l' \\ 0 & \text{otherwise} \end{cases}$

This is the so-called linear model of coregionalization. This model has become a widely used tool in multivariate geostatistics. The decomposition of Z into independent components Y^l yields the following formulation of the semivariogram matrix (in the isotropic case):

$$\mathbf{\Gamma}(h) = \sum_{l=0}^L \mathbf{B}^l g^l(h) \quad (34)$$

where \mathbf{B}^l are the coregionalization matrices and $g^l(h)$ are admissible semivariogram functions chosen a priori by the user. One can note that the case of $L = 0$ corresponds to the separable model of Equation 32. The coregionalization matrices can be interpreted as specific contributions to the sill or variance of each structure $g^l(h)$. It can be shown that in order to ensure the positive definiteness of the covariance matrix, one only needs to provide positive definite \mathbf{B}^l matrices.

4.1. Fitting technique

Goulard and Voltz [24] proposed an automated algorithm to fit a Linear Model of Coregionalization (LMC) in a positive definite manner. Its objective is to minimize a weighted sum of squares comparable to the one presented in Equation 27:

$$WSS = \sum_{k=1}^K \sum_{i=1}^{N_v} \sum_{j=1}^{N_v} \omega(h_k) \cdot \frac{[\tilde{\gamma}_{ij}(h_k) - \gamma_{ij}(h_k)]^2}{\sigma_i \sigma_j} \quad (35)$$

where $\tilde{\gamma}_{ij}(h_k)$ denotes the value of the semivariogram model, $\gamma_{ij}(h_k)$ is the actual semivariogram empirical value, $\omega(h_k)$ the weight at lag h_k , σ_i the observed standard deviation of $Z_i = \epsilon(T_i)$ (here, $\sigma_i = 1$). The WSS is simply a weighted sum of the standardized squared errors between the empirical semivariogram and the model, over all periods and all discrete separation distances. The Goulard algorithm has become popular in multivariate geostatistics involving coregionalization studies, as it provides a fast and elegant way to fit all crosssemivariograms while ensuring the positive definiteness of the resulting covariance matrix.

The algorithm is executed as follows:

1. Initialize the $L + 1$ coregionalization matrices \mathbf{B}^l with any values.
2. Remove one of the $L + 1$ semivariogram models ($g^{l_0}(h)$) and compute the difference between the initial empirical model and the LMC deprived of the l_0^{th} structure:

$$\Delta_{l_0} \Gamma(h_k) = \hat{\Gamma}(h_k) - \sum_{\substack{l=0 \\ l \neq l_0}}^L \mathbf{B}^l g^l(h_k) \quad (36)$$

3. Compute the symmetric matrix:

$$\mathbf{G}_{l_0} = \sum_{k=1}^K \omega(h_k) \cdot \Delta_{l_0} \Gamma(h_k) \cdot g^{l_0}(h_k) \quad (37)$$

4. Obtain the spectral decomposition of $\mathbf{G}_{l_0} = \mathbf{Q}_{l_0} \mathbf{\Lambda}_{l_0} \mathbf{Q}_{l_0}^T$. Set all negative eigenvalues to 0 by forming: $\mathbf{G}_{l_0} = \mathbf{Q}_{l_0} \mathbf{\Lambda}_{l_0}^+ \mathbf{Q}_{l_0}^T$ where $\mathbf{\Lambda}_{l_0}^+$ is $\mathbf{\Lambda}_{l_0}$ with all the negative diagonal terms changed to 0 (this is similar to what was discussed in Equations 29 to 31).
5. Compute the new coregionalization matrix corresponding to the l_0^{th} structure:

$$\hat{\mathbf{B}}^{l_0} = \frac{\mathbf{G}_{l_0}^+}{\sum_{k=1}^K \omega(h_k) \cdot [g^{l_0}(h_k)]^2} \quad (38)$$

6. Increment $l_0 \leftarrow l_0 + 1$ ($l_0 \leftarrow 0$ if $l_0 > L$) and loop over steps 2 to 5 until WSS is smaller than a user-specified threshold.

This algorithm is equivalent to fitting one structure at a time to the empirical data, while ensuring positive definiteness of each coregionalization matrix at step 4. The procedure is not guaranteed to converge in theory, but the experience has shown that the algorithm almost always converges whatever the initial choice of the coregionalization matrices at step 1 [16]. We were able to confirm this experience with the ground motion data considered here.

The first step in fitting a coregionalization model is to choose a set of basic structures $g^l(h)$. At this point, insights from the earlier direct semivariogram fits are useful to identify which structures should be included in the model. While one could consider a model that includes all the different exponential functions previously fitted for each period pair, it is better to minimize the number of structures to simplify both calculation and later interpretation. Thus, we propose to keep one short range exponential function (of 20 kilometers) and one long range exponential function (of 70 kilometers), as well as a nugget (constant) component, so that the semivariogram matrix function

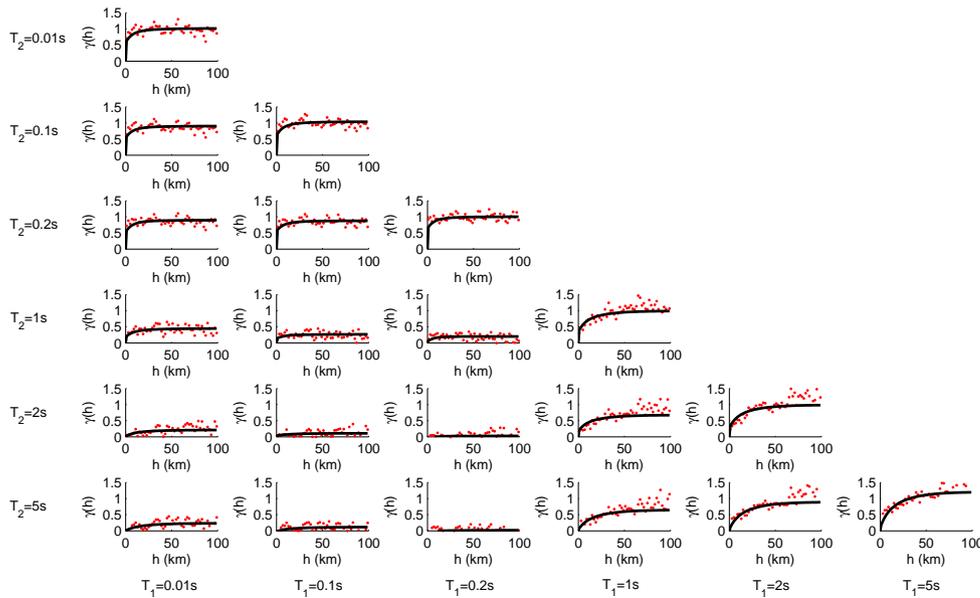


Figure 7. Northridge earthquake cross-semivariograms obtained using the Linear Model of Coregionalization.

can be expressed as:

$$\Gamma(h) = \mathbf{B}^1 \left(1 - \exp\left(\frac{-3h}{20}\right) \right) + \mathbf{B}^2 \left(1 - \exp\left(\frac{-3h}{70}\right) \right) + \mathbf{B}^3 \quad (39)$$

This choice is motivated by the observations obtained from the direct semivariogram fits. The values of 20 and 70 km were picked based on the analysis of the 8 studied earthquakes, in order for all semivariogram ranges to fall approximately within these boundaries. The addition of a nugget semivariogram is to further ensure an adequate fit at small separation distances for shorter periods. The retained weighting was the same as for the direct semivariogram fitting, such that $\omega(h_k) = 1/h_k$.

4.2. Observed results

All results have to be checked visually in order to ensure an acceptable model. We present here the fitting of a linear model of coregionalization to the empirical semivariograms previously examined. Figure 7 shows plots of the cross-semivariograms from the Northridge earthquake, where we can observe that the coregionalization model matches quite well with the observed data. However, we noted a relatively high value of the *WSS* compared to the other earthquakes. This is mainly explained by the noise in the empirical semivariograms, which can be seen for instance on the bottom-right plot of the semivariogram for $T_1 = T_2 = 5s$. We also noticed a good fit for distances smaller than 50 km, while the larger distances show much more noise in the empirical semivariogram and a poorer fit. Even so, the weighted fitting is not sensitive to these large distance values and so provides a robust estimation of the data at short distances. The values of the coefficients of the coregionalization matrices over the different period pairs show that the short range matrix \mathbf{B}^1 makes a larger contribution to the semivariogram at small periods, while the long range matrix \mathbf{B}^2 has a more significant impact on large periods. This result is in agreement with the observations made in the direct semivariogram fit.

Direct semivariogram fits of the Chi-Chi earthquake residuals showed a somewhat different spatial behavior, in the sense that we could not identify two structures as clearly as with the

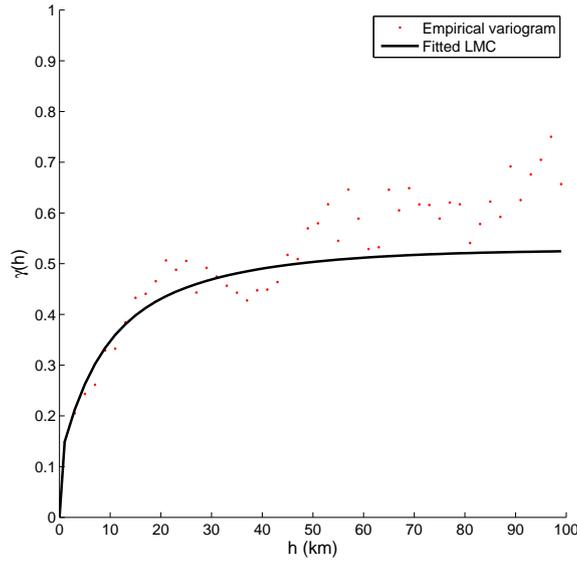


Figure 8. Chi-Chi earthquake cross-semivariogram obtained using the Linear Model of Coregionalization for $T_1 = 1s$ and $T_2 = 2s$.

Northridge earthquake. However, it has been observed that the same linear coregionalization model used with the Northridge data still provided very accurate fits of every semivariogram, mainly because the observed ranges in the direct semivariogram fit were also within 20 to 70 kilometers. Figure 8 shows the fitted cross-semivariogram between 1s and 2s data, and the coregionalization model provides a good fit of the empirical data, particularly at distances of less than 50 km.

Other earthquakes were investigated, and the linear model of coregionalization lead once again to high quality semivariogram fits. It was observed that the values of the coregionalization matrices look quite similar from earthquake to earthquake, except for the high periods of the Northridge earthquake ($T > 3s$), because those motions were recorded on analog instruments, causing much of the long-period data to be unusable. While the presence of some interevent variability in the spatial correlation has been previously established by Goda [25], the lack of systematically differing patterns between these cases suggests that this data set does not provide evidence to build a model that varies by region, or by earthquake magnitude. This lack of observed variation is not proof that no such trends exist, but rather that if they exist they are subtle enough that they cannot be detected using currently available earthquake strong motion data. The individual results for each event will be incorporated in the proposed predictive equation.

4.3. Consistency

To verify the robustness of this fitting procedure, the empirical covariance matrix functions developed for the Chi-Chi earthquake were used to generate ϵ data via Monte-Carlo simulations at the locations of the recordings from the same Chi-Chi earthquake. From the simulated ϵ values, another model of coregionalization was fit, and compared to the original model of coregionalization which the simulations were based on. A measure of the difference between the initial coregionalization matrices and the ones fitted to the generated data was defined as follows:

$$\begin{cases} \Delta_{ij}^{SR} = \mathbf{B}_{initial,ij}^1 - \mathbf{B}_{fitted,ij}^1 \\ \Delta_{ij}^{LR} = \mathbf{B}_{initial,ij}^2 - \mathbf{B}_{fitted,ij}^2 \\ \Delta_{ij}^{nug} = \mathbf{B}_{initial,ij}^3 - \mathbf{B}_{fitted,ij}^3 \end{cases} \quad (40)$$

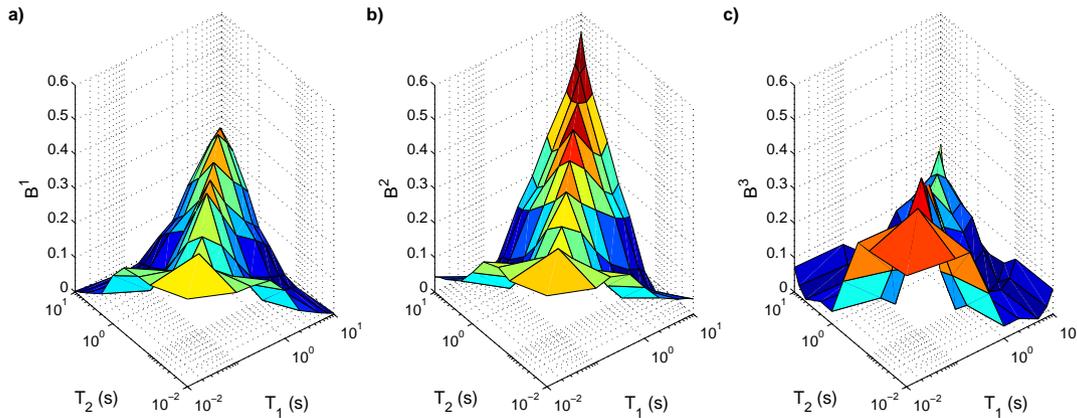


Figure 9. Average over the considered earthquakes of the short range \mathbf{B}^1 (a), long range \mathbf{B}^2 (b) and nugget effect \mathbf{B}^3 (c) coregionalization matrices.

Very little differences were observed (i.e. $|\Delta_{ij}^{SR}|$, $|\Delta_{ij}^{LR}|$ and $|\Delta_{ij}^{nug}|$ were all less than 0.1 for all i, j), which indicates the robustness and unbiasedness of the fitting method. Similar numerical results are provided in [14] to further quantify this conclusion.

4.4. Observations

Extending the simple framework of the separable model, the linear model of coregionalization proved to be a reliable technique to fit many cross-covariances at once. The Goulard algorithm is both fast and easy to use, as it requires only the empirical semivariograms and the set of basic structures $g^l(h)$. The goodness of fit obtained with this new method is comparable to the results from the direct semivariogram fitting, but the linear model of coregionalization also provides an admissible model for simulation purposes because the positive definiteness of the full covariance matrix for ϵ 's having arbitrary periods and locations is ensured.

We next use the fitted coregionalization models to build a predictive equation for the covariance matrix function $\mathbf{C}(h)$ at lag h .

5. FORMULATION OF A PREDICTIVE MODEL

5.1. Predictive model for the covariance

From all investigated earthquake data, we propose a model to predict the covariance matrix function $\mathbf{C}(h)$ from a sampling of nine periods ranging from 0.01 to 10 seconds, by averaging all the fitted coregionalization matrices over the various earthquakes. The resulting matrices are plotted in Figure 9.

One can extract any subsample of periods and use the corresponding coregionalization submatrices for simulation purposes. If one wants to consider a period that is not included here, linear interpolation between periods can be used as long as the positive definiteness of the resulting coregionalization matrices is verified. If the resulting coregionalization matrix is not positive definite, then setting the eigenvalues of the non-positive definite matrix to 0 will lead to an admissible model (see the procedure described in Equations 29 to 31).

The semivariogram matrix function is first modeled using Equation 39 with the coregionalization matrices \mathbf{B}^1 and \mathbf{B}^2 . The covariance matrix function $\mathbf{C}(h)$ can be obtained from the semivariogram matrix with Equation 14 by noting that:

$$\mathbf{C}(0) = \lim_{h \rightarrow +\infty} \mathbf{\Gamma}(h) = \mathbf{B}^1 + \mathbf{B}^2 + \mathbf{B}^3 \quad (41)$$

Table II. Short range coregionalization matrix, \mathbf{B}^1

Period (s)	0.01	0.1	0.2	0.5	1	2	5	7.5	10
0.01	0.30								
0.1	0.24	0.27							
0.2	0.23	0.19	0.26				sym.		
0.5	0.22	0.13	0.19	0.32					
1	0.16	0.08	0.12	0.23	0.32				
2	0.07	0	0.04	0.14	0.22	0.33			
5	0.03	0	0	0.09	0.13	0.23	0.34		
7.5	0	0	0	0.06	0.09	0.19	0.29	0.30	
10	0	0	0	0.04	0.07	0.16	0.24	0.25	0.24

Table III. Long range coregionalization matrix, \mathbf{B}^2

Period (s)	0.01	0.1	0.2	0.5	1	2	5	7.5	10
0.01	0.31								
0.1	0.26	0.29							
0.2	0.27	0.22	0.29				sym.		
0.5	0.24	0.15	0.24	0.33					
1	0.17	0.07	0.15	0.27	0.38				
2	0.11	0	0.09	0.23	0.34	0.44			
5	0.08	0	0.03	0.17	0.23	0.33	0.45		
7.5	0.06	0	0.02	0.14	0.19	0.29	0.42	0.47	
10	0.05	-0.03	0	0.14	0.21	0.32	0.42	0.47	0.54

Table IV. Nugget effect coregionalization matrix, \mathbf{B}^3

Period (s)	0.01	0.1	0.2	0.5	1	2	5	7.5	10
0.01	0.38								
0.1	0.36	0.43							
0.2	0.35	0.35	0.45				sym.		
0.5	0.17	0.13	0.11	0.35					
1	0.04	0	-0.04	0.20	0.30				
2	0.04	0.02	-0.02	0.06	0.14	0.22			
5	0	0	-0.04	0.02	0.09	0.12	0.21		
7.5	0.03	0.02	-0.02	0.04	0.12	0.13	0.17	0.23	
10	0.08	0.08	0.03	0.02	0.04	0.09	0.13	0.10	0.22

which yields the following simple formulation:

$$\mathbf{C}(h) = \mathbf{B}^1 \exp\left(\frac{-3h}{20}\right) + \mathbf{B}^2 \exp\left(\frac{-3h}{70}\right) + \mathbf{B}^3 \mathcal{I}_{h=0} \quad (42)$$

where $\mathcal{I}_{h=0}$ is the indicator function equal to 1 at $h = 0$ and 0 otherwise.

Note that due to fitting approximations and the fact that the diagonal terms of Equation 41 are not exactly one, the resulting coregionalization matrices from the Goulard fitting algorithm have been standardized [†] such that:

$$B_{ij}^l \leftarrow \frac{B_{ij}^l}{\left(\sqrt{B_{ii}^1 + B_{ii}^2 + B_{ii}^3}\right) \times \left(\sqrt{B_{jj}^1 + B_{jj}^2 + B_{jj}^3}\right)} \quad (43)$$

The standardized \mathbf{B}^l matrices are provided in Tables II to IV for a set of nine periods.

[†]This standardization formula is simply obtained by dividing the covariance matrix coefficients by the product of the standard deviations at the two considered periods, since: $\sigma_i = \sqrt{C_{ii}(0)} = \sqrt{B_{ii}^1 + B_{ii}^2 + B_{ii}^3}$

5.2. Example

Consider the covariance between $\ln Sa(1s)$ at site A and $\ln Sa(2s)$ at site B in a given earthquake, where sites A and B are separated by a distance of $h = 10$ kilometers. One reads in Tables II and III that $B_{1s,2s}^1 = 0.22$ and $B_{1s,2s}^2 = 0.34$, and substitutes these values in Equation 42 to obtain:

$$C_{12}(10) = 0.22 \times \exp\left(\frac{-3 \times 10}{20}\right) + 0.34 \times \exp\left(\frac{-3 \times 10}{70}\right) = 0.27 \quad (44)$$

As stated by Equations 41 and 42, the nugget effect coregionalization matrix B^3 will only be used when evaluating cross-covariances at a single site.

This calculation is clearly rather simple, indicating that while the calibration of the model was complex, it is very easy to apply.

6. EVALUATION OF THE MARKOV-TYPE SCREENING HYPOTHESIS

In this section, we present an application of the use of the proposed covariance model. While showing the general principles of the construction of the spatial covariance matrix, we also evaluate the impact of accounting for different sets of other ground-motion intensities (e.g. spectral accelerations at different sites or different periods) in the variance of the final prediction of one ground-motion intensity at a given site. Models that involve conditioning on a smaller set of variables rather than the full considered set are called Markov models. Journal [26] introduced a Markov-type model to be used in the joint modeling of two random variables Z_1 and Z_2 , considering the “screening” hypothesis stated as follows:

$$\mathbb{E}[Z_2(u)|Z_1(u); Z_1(u+h)] = \mathbb{E}[Z_2(u)|Z_1(u)] \quad (45)$$

In words, this hypothesis assumes that the dependence of the variable Z_2 on the primary variable Z_1 is limited to the co-located primary variable. The primary variable Z_1 is the variable with the larger correlation range of the two. Under this hypothesis, the spatial correlation between the two variables can be shown to equal:

$$\rho_{12}(h) = \rho_{12}(0) \cdot \rho_1(h) \quad (46)$$

Goda and Hong [1] proposed such a model to characterize the spatial correlation between spectral accelerations at different periods ($Z_1 = \ln Sa(T_1)$, $Z_2 = \ln Sa(T_2)$, with $T_1 > T_2$). This is consistent with the definition of the primary variable above, since we often observe larger correlation ranges for higher periods. In the following, we evaluate the accuracy of this screening hypothesis by comparing predictions from the Markov dependence model to corresponding predictions from the full linear model of coregionalization derived above.

6.1. Accuracy of the correlation computed using Markov approximations

The model presented in Equation 46 is examined in this section. Figure 10a shows a comparison of the correlation coefficients obtained from the full linear model of coregionalization and from the Markov model of Equation 46, at $T_1 = 2s$ and $T_2 = 1s$. The latter model can be considered as a “reduced” coregionalization model, because it is still based on the previously developed LMC, but only one of the periods is involved in the spatial decay model. Also plotted is the Equation 46 result with $T_1 = 1s$ and $T_2 = 2s$, for which one observes a slightly greater difference with the full LMC: this is consistent with the rule that the primary period should be the larger one. One observes a very good match between the two approaches over all separation distances.

However, as can be seen on Figure 10b, this Markov approximation is not as good for periods more widely separated (plotted are the cross-correlations corresponding to $T_1 = 2s$ and $T_2 = 0.2s$). In such a case, using the full coregionalization model is the better option[‡].

[‡]It should also be noted that the low values of ρ on Figure 10b indicate that it might not be worth to consider any spatial correlation effect in this case.

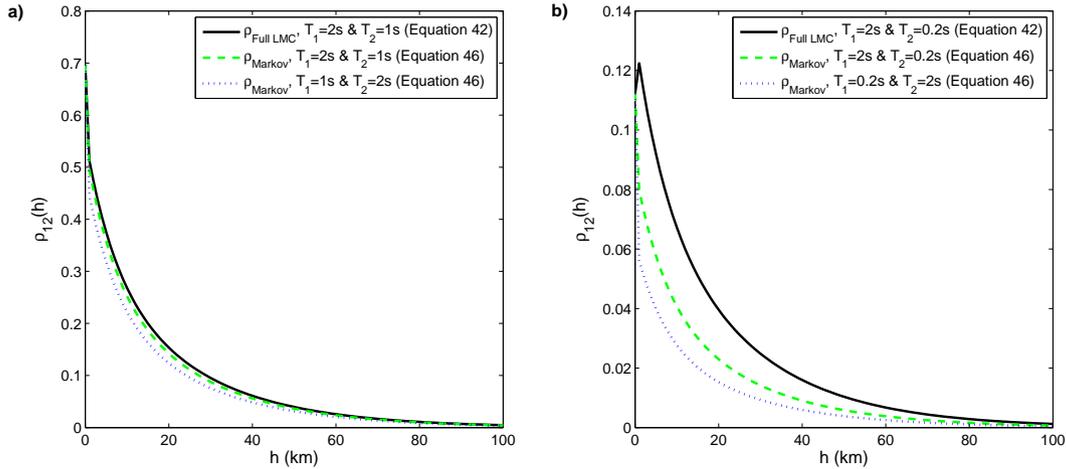


Figure 10. Comparison between the correlation coefficient obtained with the full linear model of coregonalization and the one computed from the reduced Markov-type model for: a) 2s and 1s; b) 2s and 0.2s.

6.2. Impact of the Markov approximations for computing conditional variances

While the accuracy of a Markov-type correlation approximation was discussed, it is also important to study the resulting variance implied by such estimation. The first case considered here is the situation where one observes spectral accelerations at various periods but at one site, and wants to predict the spectral acceleration at a single period at another site. In other words, one desires to know $\epsilon_{\text{site A}}(T^*)$ conditioned on the observations $\{\epsilon_{\text{site B}}(T_1), \dots, \epsilon_{\text{site B}}(T^*), \dots, \epsilon_{\text{site B}}(T_n)\}$. A problem of interest is how $\text{var}(\epsilon_{\text{site A}}(T^*)|\epsilon_{\text{site B}}(T^*))$ compares to $\text{var}(\epsilon_{\text{site A}}(T^*)|\epsilon_{\text{site B}}(T_1), \dots, \epsilon_{\text{site B}}(T^*), \dots, \epsilon_{\text{site B}}(T_n))$, which is an evaluation of the amount of extra information brought by incorporating additional conditioning periods at a remote site to assess the primary residual of interest. It can be theoretically shown that accounting for multiple conditioning periods rather than a single one will reduce the variance of $\epsilon_{\text{site A}}(T^*)$, thereby resulting in an increase in the accuracy of the intensity estimates [16]. Due to the multivariate normal distribution of the ϵ 's, one can easily compute the presented conditional variances; denoting ϵ_1 the set of residuals to be predicted, conditioned on the set of residuals ϵ_2 , one can express their joint distribution as follows:

$$\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \quad (47)$$

where $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The covariance matrix is obtained using a spatial correlation model described earlier. Given this model for the joint distribution, the distribution of ϵ_1 conditional on ϵ_2 is obtained as follows:

$$\epsilon_1|\epsilon_2 \sim \mathcal{N}(\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\mathbf{e}, \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}) \quad (48)$$

where \mathbf{e} is the vector of observed values of ϵ_2 at the recording stations. As a further application, the expected ground-motion intensities at all sites are then obtained by combining the median intensities with the expected value of the residuals obtained from Equation 48. Denoting h the separation distance between site A and site B, one can form the covariance matrices of interest to evaluate $\text{var}(\epsilon_{\text{site A}}(T^*)|\epsilon_{\text{site B}}(T^*)) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$ with:

$$\begin{cases} \Sigma_{11} = \Sigma_{22} = [C_{ss}(0)] = 1 \\ \Sigma_{12} = \Sigma_{21} = [C_{ss}(h)] \end{cases} \quad (49)$$

with $C_{ss}(h)$ the covariance matrix coefficient corresponding to the period T^* . Similarly, in order to estimate $\text{var}(\epsilon_{\text{site A}}(T^*)|\epsilon_{\text{site B}}(T_1), \dots, \epsilon_{\text{site B}}(T^*), \dots, \epsilon_{\text{site B}}(T_n))$, the corresponding submatrices will

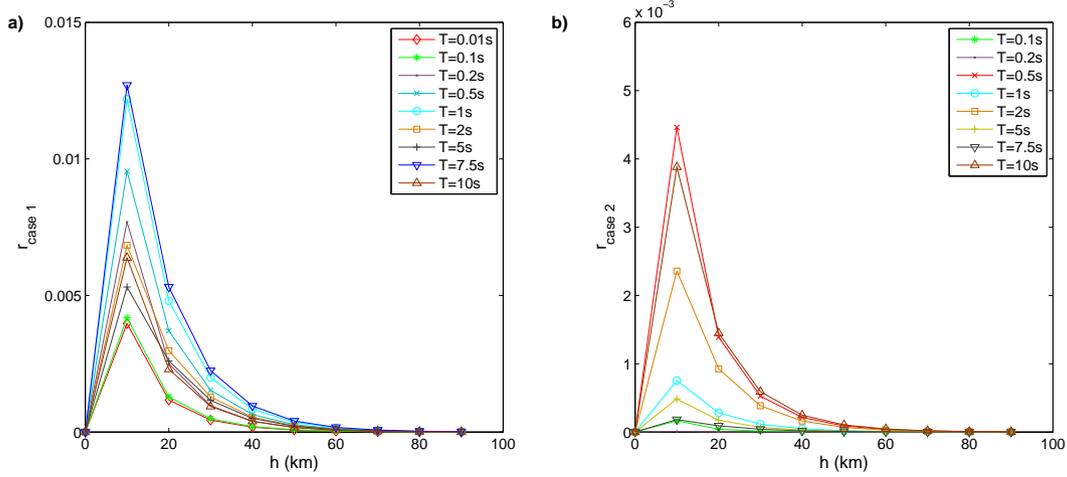


Figure 11. Reduction in variance obtained by considering additional spectral periods at a remote site when computing the variance at a given site of interest for: a) case 1; b) case 2.

be:

$$\left\{ \begin{array}{l} \Sigma_{11} = [C_{ss}(0)] = 1 \\ \Sigma_{22} = \begin{bmatrix} C_{11}(0) & \cdots & C_{1s}(0) & \cdots & C_{1n}(0) \\ & \ddots & & & \vdots \\ & & C_{ss}(0) & & C_{sn}(0) \\ & & & \ddots & \vdots \\ & & & & C_{nn}(0) \end{bmatrix} \\ \Sigma_{12} = \Sigma_{21}^T = [C_{11}(h), \dots, C_{1s}(h), \dots, C_{1n}(h)] \end{array} \right. \quad (50)$$

Figure 11a shows a plot of the relative variance reduction $r_{\text{case 1}}$ for different choices of the primary period T^* , over a varying separation distance h :

$$r_{\text{case 1}} = \text{var}(\epsilon_{\text{site A}}(T^*) | \epsilon_{\text{site B}}(T^*)) - \text{var}(\epsilon_{\text{site A}}(T^*) | \epsilon_{\text{site B}}(T_1), \dots, \epsilon_{\text{site B}}(T^*), \dots, \epsilon_{\text{site B}}(T_n)) \quad (51)$$

with T^* chosen among the sample $[0.1, 0.2, 0.5, 1, 2, 5, 7.5, 10]$, and T_1 to T_n the remaining periods of that sample. One observes that $r_{\text{case 1}}$ is equal to zero at $h = 0$, since site A and site B are at the same location, and the two variances are equal to 0. Also, $r_{\text{case 1}}$ tends to 0 as h tends to infinity, because the conditional set of observations at a far away site does not provide any information about the residual at the primary site. Overall, little variance reduction (less than 0.02) is achieved when incorporating multiple periods in the conditional set of observations. Note that the Markov model of Equation 45 would produce no variance reduction, which is a close approximation of the more precise result and thus indicates the reasonableness of a screening hypothesis in the joint modeling of spectral accelerations.

A second case was investigated in a similar manner. The problem is now to predict the residual $\epsilon_{\text{site A}}(T_1)$ conditioned on the residual at the same site but at a different period $\epsilon_{\text{site A}}(T_2)$, and then to quantify the variance reduction generated by additionally considering the residual at the conditioning period and at a remote site $\epsilon_{\text{site B}}(T_2)$. Equation 48 still applies, one will estimate $\text{var}(\epsilon_{\text{site A}}(T_1) | \epsilon_{\text{site A}}(T_2)) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ with:

$$\left\{ \begin{array}{l} \Sigma_{11} = [C_{11}(0)] = 1 \\ \Sigma_{22} = [C_{22}(0)] = 1 \\ \Sigma_{12} = \Sigma_{21}^T = [C_{12}(0)] \end{array} \right. \quad (52)$$

with $C_{12}(0)$ the covariance matrix coefficient corresponding to the periods T_1 and T_2 . Similarly, one can compute $\text{var}(\epsilon_{\text{site A}}(T_1) | \epsilon_{\text{site A}}(T_2), \epsilon_{\text{site B}}(T_2))$, denoting h the separation distance between site

A and site B:

$$\begin{cases} \Sigma_{11} = [C_{11}(0)] = 1 \\ \Sigma_{22} = \begin{bmatrix} C_{22}(0) & C_{22}(h) \\ C_{22}(h) & C_{22}(0) \end{bmatrix} \\ \Sigma_{12} = \Sigma_{21}^T = [C_{12}(0), C_{12}(h)] \end{cases} \quad (53)$$

A plot of the relative variance reduction $r_{\text{case } 2}$ for different choices of the primary period T_1 is shown on Figure 11b:

$$r_{\text{case } 2} = \text{var}(\epsilon_{\text{site A}}(T_1)|\epsilon_{\text{site A}}(T_2)) - \text{var}(\epsilon_{\text{site A}}(T_1)|\epsilon_{\text{site A}}(T_2), \epsilon_{\text{site B}}(T_2)) \quad (54)$$

T_1 was selected among the list of periods $[0.1, 0.2, 0.5, 1, 2, 5, 7.5, 10]$, and T_2 was chosen as the closest inferior period to T_1 in that same set (for the first period of the set $T_1 = 0.1s, T_2 = 0.01s$ was considered). Again, $r_{\text{case } 2}$ has the same properties as $r_{\text{case } 1}$ as h tends to 0 and to infinity. Even less variance reduction is achieved as compared to the previously investigated case ($r_{\text{case } 2} \ll r_{\text{case } 1}$), which means that the estimation variance is not affected by the incorporation of the extra information $\epsilon_{\text{site B}}(T_2)$. This again indicates the reasonableness of the screening hypothesis from Equation 45.

7. SUMMARY

This research has presented various techniques to model the spatial correlation of spectral accelerations at multiple periods. Quantifying this correlation was done with geostatistical tools involving semivariogram modeling, a measure of spatial dissimilarity. Ground motions recordings from eight earthquakes (Northridge, Chi-Chi, Tottori, Niigata, Parkfield, Chuetsu, Iwate, El Mayor Cucapah) were used to compute empirical semivariograms of spectral acceleration residuals at different periods.

Initial results were presented for independent fits of cross-semivariograms for each period pair, using an exponential function characterized by a sill (asymptotic value of the semivariogram) and a range (distance at which correlation is effectively zero). An automated least squares algorithm was developed, with a robust estimation of the sill using a kernel method. This approach allows evaluating a correlation coefficient between spectral accelerations at different periods and at different sites.

This first result is informative, but is not compatible with the generation of simulated ground motion maps, which requires a positive definite covariance matrix. Based on the direct fit results, three underlying structures were identified (short- and long-range functions both accounting for the spatial decay of the correlation as distance increases, as well as a nugget effect) that became inputs to a linear model of coregionalization, equivalent to the modeling of each cross-semivariogram with a linear combination of the three structures. Extending the simple framework of the separable model (in which only one range is used for all cross-semivariograms), the linear model of coregionalization proved to be a reliable technique to fit many cross-covariances at once. The Goulard algorithm, used to fit the model, is fast and easy to use, as it does not require any other input than the empirical semivariograms and the set of basic structures $g^l(h)$. The goodness of fit obtained with this new method is comparable to the results from the initial direct semivariogram fitting. This allowed generating a new admissible covariance model applicable for ground motion simulation purposes. It can be viewed as an extension of empirical ground motion prediction models, which generalizes the models to predict multivariate distributions of spectral accelerations at multiple periods and locations, rather than only univariate distributions of single spectral accelerations. Using this model, the correlation coefficient between any pair of spectral accelerations at different periods and at different sites may also be easily retrieved as shown in a simple example.

The robustness of the model calibration approach was evaluated using a novel approach, by simulating a synthetic set of ground motion data from the estimated cross-semivariogram model, and attempting to re-estimate the model from the synthetic data. The estimated cross-semivariograms obtained from the synthetic data were very similar to the model used to generate the data, indicating

that the algorithm is able to accurately detect spatial correlation features from observed ground motions.

The developed covariance model was then used to examine the validity of a Markovian screening hypothesis in the case of ground motion residuals. We focused on a Markov model formulating the cross-correlation coefficient as a product of the cross-correlation at a single site times the spatial correlation coefficient of the highest period. This approach proved to be compatible with the developed coregionalization model, and can therefore be considered as a possible simplification of the full linear model of coregionalization, as long as the two considered periods are relatively close to one another.

Even though the calibration of this model and investigation of its implications were somewhat complex, it should be emphasized that the model is very simple to use for making correlation predictions. A user of this model only needs to evaluate Equation 42, with the needed coefficients from Tables II to IV, to compute a correlation coefficient for spectral values at two periods at a specified separation distance. While this model is more general than most previous models that considered only single-period correlations or used a Markov-type assumption to compute multi-period correlations, the model proposed here is not significantly more complex to use than those earlier models, and so should be a useful resource for those interested in predicting correlations of spectral values at differing locations and periods.

ACKNOWLEDGEMENTS

We thank Nirmal Jayaram, Katsu Goda, Peter Stafford, Paolo Bazzurro, André Journal and Pierre Goovaerts for useful discussions on this research topic. This work was supported by the U.S. Geological Survey (USGS) via External Research Program award G10AP00046. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the USGS.

REFERENCES

1. Goda K, Hong HP. Spatial correlation of peak ground motions and response spectra. *Bulletin of the Seismological Society of America* 2008; **98**(1):354–365, doi:10.1785/0120070078.
2. Stahl DO. Rule learning in symmetric normal-form games: Theory and evidence. *Games and Economic Behavior* 2000; **32**(1):105 – 138, doi:10.1006/game.1999.0754. URL <http://www.sciencedirect.com/science/article/pii/S089982569907546>.
3. Boore DM, Atkinson GM. Ground motion prediction equations for the average horizontal component of PGA, PGV, and 5%-Damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthquake Spectra* 2008; **24**(1):99–138.
4. Abrahamson N, Silva W. Summary of the Abrahamson & Silva NGA ground motion relations. *Earthquake Spectra* 2008; **24**(1):67–97.
5. Chiou BSJ, Youngs RR. An NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra* 2008; **24**(1):173–215.
6. Campbell KW, Bozorgnia Y. Campbell-Bozorgnia NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. *Technical Report PEER 2007/02*, Pacific Earthquake Engineering Research Center, Berkeley, California 2007.
7. Jayaram N, Baker JW. Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of the Seismological Society of America* 2008; **98**(5):2231–2243, doi:10.1785/0120070208.
8. Foulser-Piggott R, Stafford PJ. A predictive model for Arias intensity at multiple sites and consideration of spatial correlations. *Earthquake Engineering & Structural Dynamics* 2011; n/a URL <http://dx.doi.org/10.1002/eqe.1137>.
9. Esposito S, Iervolino I. PGA and PGV spatial correlation models based on European multievent datasets. *Bulletin of the Seismological Society of America* 2011; **101**(5):2532–2541, doi:10.1785/0120110117. URL <http://www.bssaonline.org/cgi/content/abstract/101/5/2532>.
10. Boore DM, Gibbs JF, Joyner WB, Tinsley JC, Ponti DJ. Estimated ground motion from the 1994 Northridge, California, earthquake at the site of the Interstate 10 and La Cienega Boulevard bridge collapse, West Los Angeles, California. *Bulletin of the Seismological Society of America* 2003; **93**(6):2737–2751.
11. Wang M, Takada T. Macrospectral correlation model of seismic ground motions. *Earthquake Spectra* 2005; **21**(4):1137–1156.
12. Baker JW, Jayaram N. Correlation of spectral acceleration values from NGA ground motion models. *Earthquake Spectra* 2008; **24**(1):299–317, doi:10.1193/1.2857544.
13. Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics* 2009; **38**(15):1687–1708, doi:10.1002/eqe.922. URL <http://dx.doi.org/10.1002/eqe.922>.

14. Loth C, Baker JW. Spatial cross-correlation of spectral accelerations at multiple periods: model development and risk assessments considering secondary earthquake effects. *Technical Report USGS award G10AP00046*. 2011.
15. Journel AG, Huijbregts CJ. *Mining geostatistics*. Academic Press: London; New York, 1978.
16. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, USA, 1997.
17. Bazzurro P, Luco N. Effects of different sources of uncertainty and correlation on earthquake-generated losses. Stoos, Switzerland, 2004; 20.
18. Summary of scaling relations for spectral damping, peak velocity, and average spectral acceleration: Report for the PEGASOS project. *Personal communication*.
19. Inoue T, Cornell CA. Seismic hazard analysis of multi-degree-of-freedom structures. *Technical Report #RMS-8, Reliability of Marine Structures*, Stanford, CA 1990.
20. Wackernagel H. *Multivariate geostatistics: an introduction with applications*. Berlin: Springer. edn., 1995.
21. Jackel P. *Monte Carlo methods in finance*. J. Wiley & Sons: Chichester, West Sussex, England, 2002.
22. Majumdar A, Gelfand A. Multivariate spatial modeling for geostatistical data using convolved covariance functions. *Mathematical Geology* 2007; **39**(2):225–245. URL <http://dx.doi.org/10.1007/s11004-006-9072-6>.
23. Banerjee S, Gelfand A, Carlin B. Multivariate spatial modeling. *Hierarchical Modeling and Analysis for Spatial Data*. C&H/CRC Monographs on Statistics & Applied Probability, Chapman and Hall/CRC, 2004. 0.
24. Goulard M, Voltz M. Linear coregionalization model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology* 1992; **24**(3):269–286. URL <http://dx.doi.org/10.1007/BF00893750>.
25. Goda K. Interevent variability of spatial correlation of peak ground motions and response spectra. *Bulletin of the Seismological Society of America* 2011; **101**(5):2522–2531, doi:10.1785/0120110092. URL <http://www.bssaonline.org/content/101/5/2522.abstract>.
26. Journel AG. Markov models for cross-covariances. *Mathematical Geology* 1999; **31**(8):955–964. URL <http://dx.doi.org/10.1023/A:1007553013388>.