Bulletin of the Seismological Society of America, Vol. 103, No. 6, pp. 3189–3204, December 2013, doi: 10.1785/0120130061

Spatial Cross-Correlation Models for Vector Intensity Measures (PGA, I_a, PGV, and SAs) Considering Regional Site Conditions

by Gang Wang and Wenqi Du

Abstract This study investigates spatial cross-correlation models for two sets of vector intensity measures (IMs) considering the influence of regional site conditions. The first set of the vector IM consists of the peak ground acceleration, Arias intensity, and the peak ground velocity; the second set is for spectral accelerations at multiple periods. Geostatistics analyses are performed using 2686 strong-motion data from 11 recent earthquakes that occurred in California, Japan, Taiwan, and Mexico. The results indicate that the spatial cross correlations of the vector IMs are strongly influenced by the spatial distribution of regional site conditions, which can be quantified using $R_{V_{S30}}$, the correlation range of shear-wave velocity in the top 30 m. The linear model of coregionalization is proposed to construct a permissible spatial correlation model, and the short-range and long-range coregionalization matrices is specified to vary linearly with $R_{V_{S30}}$. The proposed model demonstrated excellent performance in quantifying the influence of regional site conditions on the spatial cross correlations for the vector IMs, meanwhile the model guarantees a positive-definite covariance matrix for any reasonable value of $R_{V_{S30}}$, a mathematical condition required for stochastic generation of the spatially correlated random fields. The spatial cross-correlation models proposed in this study can be conveniently used in regional-specific seismic risk analysis and loss estimation of spatially distributed infrastructure using vector IMs.

Introduction

Modeling spatial distribution of ground-motion intensity measures (IMs) is essential for the seismic-hazard analysis and risk assessment of spatially distributed infrastructure, such as lifelines, transportation systems, and structure portfolios (e.g., Jeon and O'Rourke, 2005; Wang and Takada, 2005; Lee and Kiremidjian, 2007; Sokolov and Wenzel, 2011). Over the years, ground-motion prediction equations (GMPEs) have been actively developed to estimate the probability distribution (e.g., the median and variance) of the IMs at a single location. However, GMPEs do not estimate the spatial correlation of ground motion IMs. In recent years, several spatial correlation models have been proposed for some IMs (e.g., Goda and Hong, 2008; Jayaram and Baker, 2009; Sokolov *et al.*, 2010; Esposito and Iervolino, 2011, 2012; Goda, 2011; Du and Wang, 2013a). In particular, Jayaram and Baker (2009) developed an intraevent spatial correlation model for spectral accelerations (SAs). Esposito and Iervolino (2011, 2012) studied the spatial correlations of peak ground acceleration (PGA), peak ground velocity (PGV), and SA using European earthquake data. The spatial correlation of the ground-motion IMs is usually strong if two sites are close together, and the correlation gradually decreases as the sites are separated father apart. More recently, it has been observed that the spatial correlations of some scalar IMs are closely related to regional site conditions. In general, the spatial correlations appear to be stronger if the regional site conditions are more homogeneous (Jayaram and Baker, 2009). The spatial homogeneity of a region can be indicated by the spatial correlation of shear-wave velocity in the top 30 m (V_{S30}), as a larger correlation range of V_{S30} implies a more homogeneous site condition (Du and Wang, 2013a). Relationships between the spatial correlations of PGA and V_{S30} were investigated using earthquake data in Taiwan (Sokolov *et al.*, 2012). Predictive models were also developed to quantify the spatial correlations of cumulative absolute velocity (Electrical Power Research Institute, 1988), Arias intensity (I_a; Arias, 1970), and SAs based on the correlation range of V_{S30} (Du and Wang, 2013a). These findings have significant implications in developing spatial correlation models for regional-specific applications.

Many earthquake engineering problems require the use of multiple IMs (termed as vector IM) to represent different attributes of the ground-motion characteristics. To date, only a few studies have been conducted regarding the spatial cross correlations of vector IMs. Among these few examples, Loth and Baker (2013) investigated the spatial cross correlations of SAs at multiple periods for the analysis of structural portfolios with various types of buildings. Yet, the influence of

regional site conditions on the spatial correlation structure of vector IMs has not been fully investigated.

Following the authors' previous work (Du and Wang, 2013a), the spatial cross correlation of two sets of vector IMs and their dependency on regional site conditions are studied in this paper. As have been widely used in earthquake engineering applications, PGA, I_a, and PGV are combined as the first set of vector IM. The PGA, I_a, and PGV represent the amplitude and energy attributes of the ground motions. A combination of these IMs can be effectively used to predict the earthquake-induced sliding displacements of earth structures (e.g., Rathje and Saygili, 2008) or cyclic structural response (Iervolino *et al.*, 2010). SAs at nine different periods are chosen as the second set of vector IM to characterize the frequency content of ground motions.

Mathematically, a permissible spatial correlation model for the vector IM must yield a positive semidefinite spatial correlation matrix (Goovaerts, 1997, p. 108). The requirement is challenging but necessary for the random realization of jointly distributed vector IMs in space, a process known as stochastic simulation. For this purpose, a geostatistical approach, the linear model of coregionalization (LMC; Journel and Huijbregts, 1978), is adopted in this study to derive the spatial cross correlations of these vector IMs using strong-motion recordings from 11 earthquakes that occurred in different regions. A simple permissible spatial correlation model for two sets of vector IMs will be proposed by taking into account the influence of regional site conditions, which is parameterized by the correlation range of V_{S30} of the recording stations. The proposed correlation matrices are guaranteed to be positive definite for any finite number of site locations and all reasonable correlation ranges of V_{S30}. Finally, an illustrative example is provided to highlight the model's capability in generating spatially distributed vector IMs and its possible applications in regional-specific hazard analysis of spatially distributed structures.

Strong-Motion Data

In this study, a large number of strong-motion data from different regions are systematically compiled and analyzed to develop the spatial cross-correlation models for PGA, I_a, PGV, and SAs. The selected earthquake events satisfy the following criteria: (1) ground-motion data are densely populated, a sizable number (greater than 30) of recordings must be available within each separation distance bin in order to get a statistically reliable sample size; (2) necessary seismological and geological information must be available for these events, this information includes source parameters, site-to-source distance and site conditions (V_{S30} values), etc., so that the median IMs and their residuals can be estimated from GMPEs; (3) the moment magnitudes of events are greater than 5, and the rupture distances of recordings are within 200 km. Following these criteria, 2686 records from 11 earthquake events are selected. These earthquakes occurred in California (1994 Northridge earthquake, 2004 Parkfield earthquake, 2005 Anza earthquake, 2007 Alum Rock earthquake, and 2008 Chino Hills earthquake), in Mexico (2010 El Mayor–Cucapah earthquake), in Japan (2000 Tottori earthquake, 2004 Niigata earthquake, 2007 Chuetsu earthquake, and 2008 Iwate earthquake), and in Taiwan (1999 Chi-Chi earthquake). The recorded time histories for these events are obtained from the Center for Engineering Strong-Motion Data (CESMD), Consortium of Organizations for Strong Motion Observation Systems (COSMOS) for U.S. earthquakes, and K-NET and KiK-net for Japan earthquakes. The raw ground-motion data are uniformly processed by the authors following the Pacific Earthquake Engineering Research Center (PEER) procedure (Darragh *et al.*, 2004). The seismic information and site conditions are obtained from the PEER's strong-motion database and the table S1 database provided by Kaklamanos and Baise (2011). For the Chi-Chi event, the V_{S30} data of recording stations are updated according to the Taiwan Strong Motion Instrumentation Program (TSMIP; Kuo *et al.*, 2012). The moment magnitude versus rupture-distance distribution of the data is illustrated in Figure 1. The detailed information for each event is summarized in Table 1.

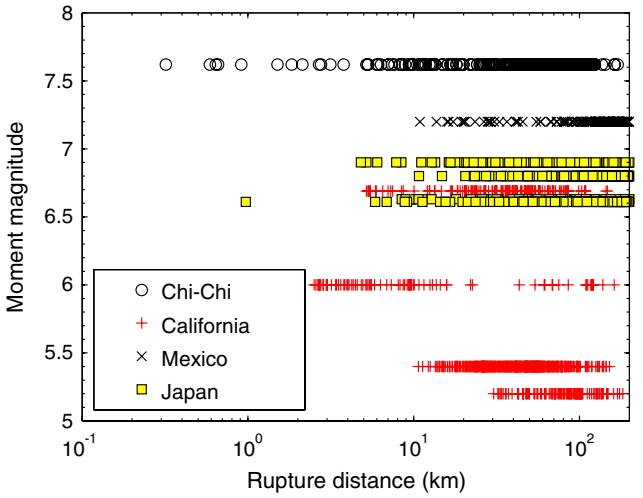

Figure 1. Magnitude and rupture-distance distribution of records in the database. The color version of this figure is available only in the electronic edition.

Geostatistical Analysis of Ground-Motion Residuals

Ground-Motion Residuals

Given an earthquake event i, a ground-motion IM at site j can be written as

$$\ln Y_{ij} = \overline{\ln Y_{ij}(M, R_{\text{rup}}, \theta)} + \eta_i + \varepsilon_{ij}, \qquad (1)$$

in which Y_{ij} is the recorded IM of earthquake event i at site j; $\overline{\ln Y_{ij}(M, R_{\text{rup}}, \theta)}$ is the predicted mean value by GMPE as a function of magnitude (M), rupture distance (R_{rup}), and other parameters (θ); η_i and ε_{ij} denote the interevent and intraevent residuals, respectively. The residual terms are assumed to

Table 1
Summary of Earthquake Events

Earthquake Name	Date (yyyy/mm/dd)	Moment Magnitude	Location	Fault Mechanism	Number of Recordings
Northridge	1994/01/17	6.69	California	Reverse	152
Chi-Chi	1999/09/20	7.62	Taiwan	Reverse-oblique	401
Tottori	2000/06/10	6.61	Japan	Strike-slip	235
Parkfield	2004/09/28	6	California	Strike-slip	90
Niigata	2004/10/23	6.63	Japan	Reverse	365
Anza	2005/06/12	5.2	California	Reverse-oblique	111
Chuetsu	2007/07/16	6.8	Japan	Reverse	401
Alum Rock	2007/10/30	5.4	California	Strike-slip	161
Iwate	2008/06/13	6.9	Japan	Reverse	279
Chino Hills	2008/07/29	5.4	California	Reverse-oblique	337
El Mayor–Cucapah	2010/04/04	7.2	Mexico	Strike-slip	154

Only recorded data within rupture distance of 200 km are included for the earthquakes in Japan.

follow normal distributions with a zero mean and a standard deviation of τ_i and σ_{ij}, respectively (Abrahamson and Youngs, 1992; Joyner and Boore, 1993; Jayaram and Baker, 2008), which are also provided by GMPEs. The total standard deviation is given by $\sigma_T = \sqrt{\sigma_{ij}^2 + \tau_i^2}$. Throughout the study, some recently developed GMPEs are used to calculate the ground-motion residuals for PGA, PGV, SAs (Campbell and Bozorgnia, 2008), and I_a (Campbell and Bozorgnia, 2012). It should be noted that most of these earthquake events (except for the Chi-Chi and the Northridge earthquakes) are not used in developing these GMPEs. Accordingly, it is not surprising that the ground-motion residuals are biased against the rupture distances (R_{rup}) and V_{S30} because the GMPEs are not calibrated by these data. For example, Figure 2 shows significant biases of the intraevent residuals of PGA with respect to R_{rup} for 11 earthquake events. Similar biases are also observed for other IMs. If not corrected, the biased residuals would inevitably result in inaccurate estimate of the spatial correlation (Sokolov *et al.*, 2010; Foulser-Piggott and Stafford, 2012). Therefore, the computed residuals need to be corrected to remove the bias as follows (Du and Wang, 2013a):

$$\varepsilon_{ij}^{corr} = \varepsilon_{ij} - [\varphi_1 + \varphi_2 \ln(R_{rup}) + \varphi_3 \ln(V_{S30})], \quad (2)$$

in which φ_1, φ_2, and φ_3 are the coefficients obtained by linear regression for each event. As expected, the corrected residuals are not biased against R_{rup} and V_{S30}, and they will be used in the following semivariogram analysis. For better comparison between different earthquakes, the corrected intraevent residuals are normalized as follows:

$$\varepsilon'_{ij} = \frac{\varepsilon_{ij}^{corr}}{\sigma_{ij}} \approx \frac{\ln Y_{ij} - \overline{\ln Y_{ij}(M, R_{rup}, \theta)} - [\varphi_1 + \varphi_2 \ln(R_{rup}) + \varphi_3 \ln(V_{S30})]}{\sigma_{ij}}, \quad (3)$$

in which ε'_{ij} is the normalized intraevent residual and σ_{ij} can be obtained either from sample variances or from GMPEs. Throughout this paper, estimates of σ_{ij} from the sample variances are adopted because the values provided by GMPEs may not be appropriate for all events considered. The above approximation equation (3) neglects the interevent residual term η_i because it is constant for each site during an earthquake and therefore does not contribute to the intraevent spatial correlation.

It is worth mentioning that there exists apparent inconsistency of using current GMPEs for the estimation of spatial correlation because they usually assume the IM residuals are uncorrelated. Considering spatial correlation of residuals in GMPEs, the interevent standard deviation would decrease and the intraevent standard deviation would increase; however, the predicted median IMs and their total residuals would not be significantly affected (Jayaram and Baker, 2010). By equation (3), the intraevent standard deviation will not affect the spatial correlation of the normalized residuals. The inconsistency in GMPEs is not expected to significantly influence the estimated spatial correlations presented in this study.

Univariate Analysis

Semivariogram analysis is a widely used geostatistical tool to analyze the spatial correlation of a univariate random field. If it is assumed that a random field Z is second-order stationary, the semivariogram formulation can be written as (Goovaerts, 1997)

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

in which $\mathbf{h}$ is the separation vector, $Z(\mathbf{u} + \mathbf{h})$ is the random variable at a position separated by $\mathbf{h}$ from the position $\mathbf{u}$. In this paper, the random variable refers to the normalized intraevent residual ε'_{ij}. The second-order stationary assumption implies that the mean value of the random field is constant over the entire domain, and the semivariogram $\gamma(\mathbf{h})$ (as well as the covariance function introduced later) does not depend

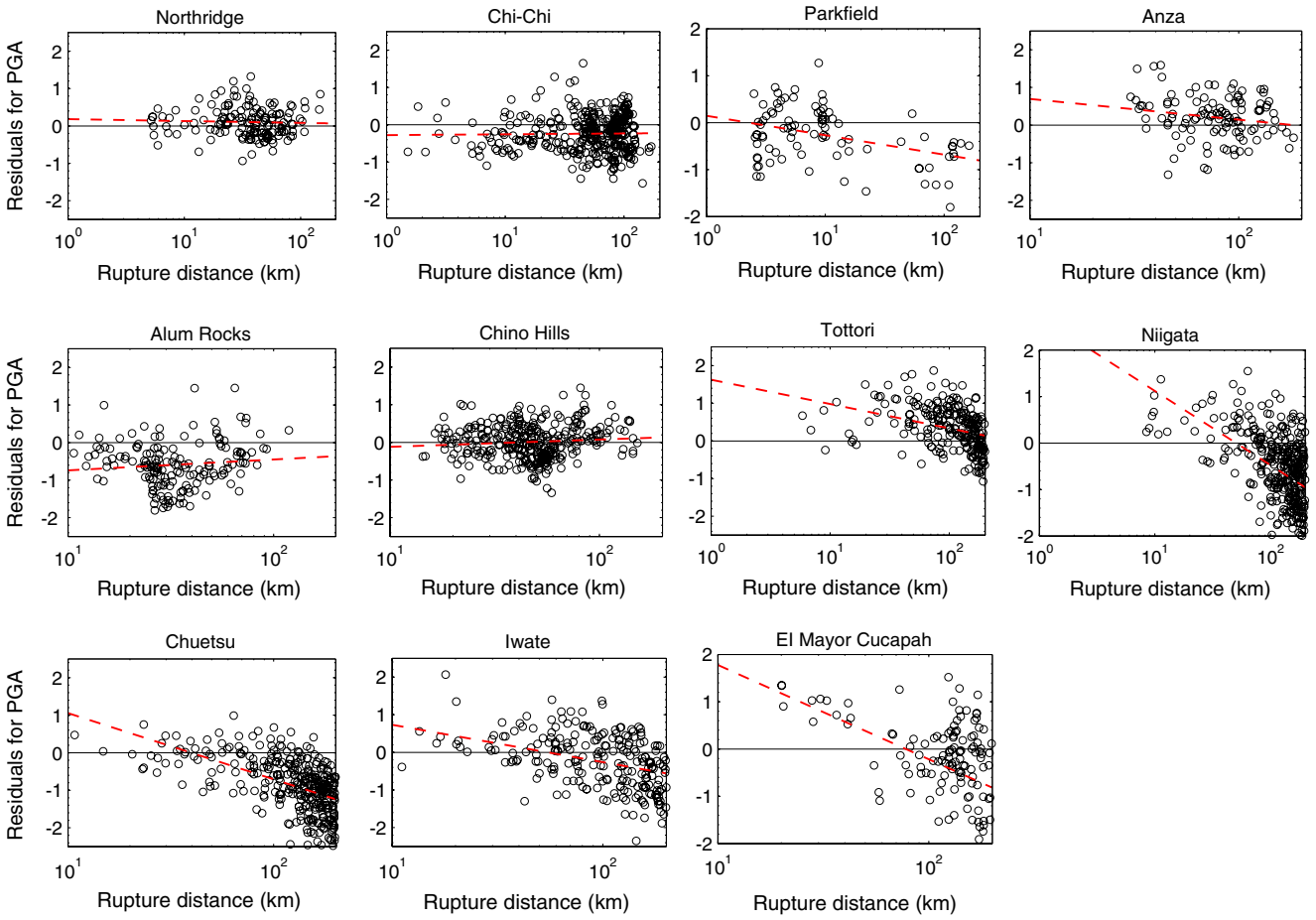

Figure 2. Distributions of intraevent residuals of PGA versus rupture distance for 11 earthquake events. The color version of this figure is available only in the electronic edition.

on the site location **u**, but only on the separation vector **h**. Furthermore, the stationary semivariogram is said to be isotropic if it is independent of direction such that the separation vector **h** in equation (4) can be replaced by the separation distance $h = \|\mathbf{h}\|$ (Goovaerts, 1997). It is worth pointing out that the isotropic assumption for the spatial distribution of the intraevent residuals have been verified for some events, such as the Northridge earthquake (Jayaram and Baker, 2009). However, comprehensive validation is still needed in future studies. Similar to the semivariogram function, the covariance function $C(h)$ is defined as

$$\begin{aligned} C(h) &= \mathrm{Cov}(Z(u), Z(u+h)) \\ &= E[(Z(u) - E[Z(u)])(Z(u+h) - E[Z(u)])]. \end{aligned} \quad (5)$$

Finally, the correlation function is defined as

$$\rho(h) = \frac{C(h)}{C(0)}. \quad (6)$$

Under the assumption that the random field is isotropic and second-order stationary, it is straightforward to relate the semivariogram function to the correlation function as follows:

$$\gamma(h) = C(0)[1 - \rho(h)] = \mathrm{Var}(Z)[1 - \rho(h)]. \quad (7)$$

Several basic functional forms can be used to approximate the empirical semivariogram data, such as the exponential, Gaussian, and spherical model (Goovaerts, 1997). In particular, the exponential model assumes

$$\begin{aligned} \gamma(h) &= a[1 - \exp(-3h/r)] \\ \text{and} \quad \rho(h) &= \exp(-3h/r), \end{aligned} \quad (8)$$

in which a is the sill of the semivariogram and also the population variance of empirical data and r is the range of the semivariogram, defined as the separation distance h at which $\gamma(h)$ equals 95% of the sill. Several approaches can be used to estimate the model parameters, such as the least square fit and weighted least square (WLS) fit. The weight for WLS is chosen as $1/h_k$ (in which h_k is the center of each distance bin) so that the empirical data in close separation distances can be better approximated. Figure 3 illustrates an example of the empirical semivariogram data and different fitting schemes. The exponential model fitted by WLS scheme can provide better estimate of the empirical data in the close

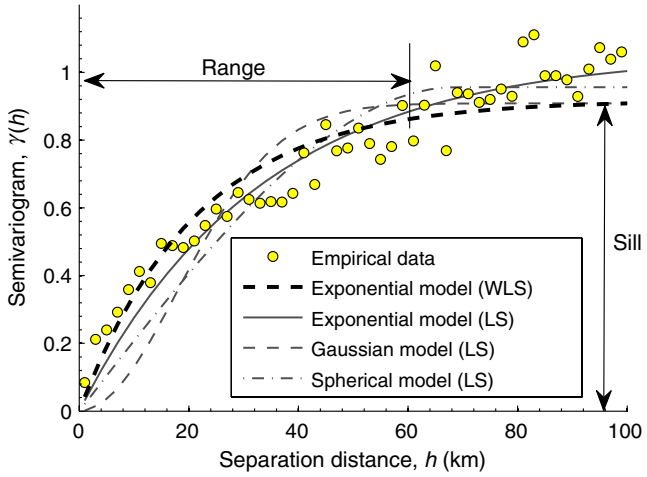

Figure 3. Fitting the empirical semivariogram data using different functional forms and fitting schemes. (The empirical data is obtained using the normalized intraevent residuals of PGA for the Chi-Chi earthquake; the range and sill are marked for the exponential model using WLS fit.) The color version of this figure is available only in the electronic edition.

separation distance range than other models. Therefore, the exponential model using WLS fit will be employed throughout this study. It is worth mentioning that the exponential, Gaussian, and spherical model are all permissible models for a univariate field by construction. The resulted correlation matrix will always be positive definite regardless of number of the sites and their locations.

Multivariate Analysis

Cross-semivariogram analysis is an extension of the above univariate case to the multivariate random field (Journel and Huijbregts, 1978; Goovaerts, 1997). Assuming that a multivariate random field consists of n isotropic second-order stationary random variables $Z_1, Z_2, \ldots, Z_n$, the cross-semivariogram $\gamma_{ij}(h)$ measures the average dissimilarity between two variables Z_i and Z_j separated by distance h via the equation

$$\gamma_{ij}(h) = \frac{1}{2}E[(Z_i(u+h) - Z_i(u))(Z_j(u+h) - Z_j(u))], \tag{9}$$

in which $Z_i(u)$ and $Z_i(u+h)$ refer to the values of Z_i evaluated at locations u and $u+h$, respectively. In this paper, Z_i refers to the normalized intraevent residual of the ith component of the vector IM. The empirical cross semivariogram $\tilde{\gamma}_{ij}(h)$ can be estimated as

$$\tilde{\gamma}_{ij}(h) = \frac{1}{2|N(h)|}\sum_{\alpha=1}^{N(h)}\{[z_i(u_\alpha + h) - z_i(u_\alpha)][z_j(u_\alpha + h) - z_j(u_\alpha)]\}, \tag{10}$$

in which $N(h)$ is the number of data pairs within this distance bin, $z_i(u_\alpha + h)$ and $z_i(u_\alpha)$ represent the αth data pair in this bin for ith component of the vector IM. On the other hand, the covariance function $C_{ij}(h)$ measures the similarity between data as follows:

$$\begin{aligned} C_{ij}(h) &= \mathrm{Cov}(Z_i(u), Z_j(u+h)) \\ &= E[(Z_i(u) - E[Z_i(u)])(Z_j(u+h) - E[Z_j(u)])]. \end{aligned} \tag{11}$$

For an isotropic and stationary field, $\lim_{h\to\infty}\gamma_{ij}(h) = \mathrm{Cov}(Z_i(u), Z_j(u)) = C_{ij}(0)$. Therefore, the relationship between covariance function and cross-semivariogram function can be established as (Goovaerts, 1997, pp. 72–74)

$$C_{ij}(h) = \lim_{h'\to\infty}\gamma_{ij}(h') - \gamma_{ij}(h) = C_{ij}(0) - \gamma_{ij}(h). \tag{12}$$

The spatial correlation coefficient between Z_i and Z_j is defined as

$$\begin{aligned} \rho_{ij}(h) &= \frac{C_{ij}(h)}{\sqrt{C_{ii}(0)\times C_{jj}(0)}} \\ &= \frac{C_{ij}(0)}{\sqrt{C_{ii}(0)\times C_{jj}(0)}} - \frac{\gamma_{ij}(h)}{\sqrt{C_{ii}(0)\times C_{jj}(0)}}, \end{aligned} \tag{13}$$

which is a multivariate analogy of equation (6). Accordingly, the cross-semivariogram matrix $\boldsymbol{\Gamma}(h)$, the covariance matrix $\mathbf{C}(h)$, and the correlation matrix $\mathbf{R}(h)$ for a vector IM with n components are $n \times n$ matrices defined as follows:

$$\begin{aligned} \boldsymbol{\Gamma}(h) &= [\gamma_{ij}(h)] = \begin{bmatrix} \gamma_{11}(h) & \cdots & \gamma_{1n}(h) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(h) & \cdots & \gamma_{nn}(h) \end{bmatrix}, \\ \mathbf{C}(h) &= [C_{ij}(h)], \\ \mathbf{R}(h) &= [\rho_{ij}(h)]. \end{aligned} \tag{14}$$

The correlation matrix $\mathbf{R}(h)$ contains the direct-correlation coefficients along its major diagonal ($i = j$) and the cross-correlation coefficients off that diagonal ($i \neq j$). As the normalized intraevent residuals of the vector IM can be reasonably assumed to follow a multivariate normal distribution, their spatial distribution for an earthquake event k can be fully characterized by the mean (zero vector in this case) and the covariance matrix. Given an n-component vector IM distributed at J sites from an earthquake event k, the total covariance matrix $\boldsymbol{\Sigma}$(event k) can be obtained by assembling the $n \times n$ submatrix $\mathbf{C}(h)$ as follows:

$$\boldsymbol{\Sigma}(\text{event } k) = \begin{bmatrix} \mathbf{C}(0) & \cdots & \mathbf{C}(h_{1J}) \\ \vdots & \ddots & \vdots \\ \mathbf{C}(h_{J1}) & \cdots & \mathbf{C}(0) \end{bmatrix}, \tag{15}$$

in which h_{ij} represents the separation distance between site i and site $j (i, j = 1, \ldots, J)$. The dimension of the total covariance matrix $\boldsymbol{\Sigma}$(event k) is $[J \times n, J \times n]$.

Linear Model of Coregionalization for the Vector IMs

Fitting the empirical cross-semivariogram data is not as straightforward as the univariate case. It may be tempting to apply the exponential model directly to fit the empirical cross-semivariogram data as

$$\gamma_{ij}(h) = a_{ij}[1 - \exp(-3h/r_{ij})], \tag{16}$$

in which a_{ij} and r_{ij} are fitting parameters $(i, j = 1, \ldots, n)$. However, the resulted covariance matrix $\mathbf{C}(h)$ usually cannot guarantee positive definiteness for all h values, and the model is not permissible.

The LMC is used in this study to build a permissible cross-semivariogram model for multivariate analysis (Goovaerts, 1997). The LMC model decomposes a cross-semivariogram matrix into a combination of L linearly independent components using

$$\mathbf{\Gamma}(h) = [\gamma_{ij}(h)] = \sum_{l=1}^{L} \mathbf{B}^l g_l(h), \tag{17}$$

in which $g_l(h)$ forms a set of permissible basic models $(l = 1, 2, \ldots, L)$, $\mathbf{B}^l = [b_{ij}^l]$ $(i, j = 1, \ldots, n;\ l = 1, 2, \ldots, L)$ is termed as the coregionalization matrix, which is an $n \times n$ matrix associated with each $g_l(h)$ and shall be determined by fitting the empirical semivariogram data. By the LMC construction, all direct and cross semivariograms share the same set of basic structures $g_l(h)$. For example, using two exponential functions with range values of r_1 and r_2 as the basic functions, the cross-semivariogram matrix equation (17) can be expressed as

$$\mathbf{\Gamma}(h) = \mathbf{B}^1\left[1 - \exp\left(\frac{-3h}{r_1}\right)\right] + \mathbf{B}^2\left[1 - \exp\left(\frac{-3h}{r_2}\right)\right]. \tag{18}$$

Following equation (12), the covariance matrix becomes

$$\begin{aligned}\mathbf{C}(h) &= \lim_{h\to\infty} \mathbf{\Gamma}(h) - \mathbf{\Gamma}(h) \\ &= \mathbf{B}^1\left[\exp\left(\frac{-3h}{r_1}\right)\right] + \mathbf{B}^2\left[\exp\left(\frac{-3h}{r_2}\right)\right].\end{aligned} \tag{19}$$

The total covariance matrix in equation (15) is guaranteed to be positive semidefinite regardless of the number of sites considered, if the coregionalization matrices $\mathbf{B}^l$ are all positive semidefinite (Goovaerts, 1997). The coregionalization matrices $\mathbf{B}^l$ can be obtained using the WLS method by minimizing the weighted sum of squares of the difference between the estimated and the empirical cross semivariograms between n IMs within K distance bins as follows (Goulard and Voltz, 1992):

$$WSS = \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{n} w(h_k)\frac{[\tilde{\gamma}_{ij}(h_k) - \gamma_{ij}(h_k)]^2}{\hat{\sigma}_i\hat{\sigma}_j}, \tag{20}$$

in which $\gamma_{ij}(h_k)$ is the estimated cross semivariogram by the LMC model, $\tilde{\gamma}_{ij}(h_k)$ is the empirical cross semivariogram, $\hat{\sigma}_i$ is the standard deviation of Z_i, and $w(h_k)$ refers to the weight for distance h_k, taking as $1/h_k$ to improve the goodness of fit to data at close separation distances. An efficient algorithm has been developed to achieve the positive semidefiniteness of $\mathbf{B}^l$ by eliminating their negative eigenvalues during a minimization process (e.g., Goulard and Voltz, 1992; Loth and Baker, 2013). The algorithm is implemented in this study to obtain $\mathbf{B}^l$ for each event.

Based on observation from selected earthquake events, a combination of a short-range and a long-range exponential function is proposed to approximate the cross semivariogram of the vector IM [PGA, I_a, PGV]:

$$\mathbf{\Gamma}(h) = \mathbf{B}^1\left[1 - \exp\left(\frac{-3h}{10}\right)\right] + \mathbf{B}^2\left[1 - \exp\left(\frac{-3h}{60}\right)\right], \tag{21}$$

in which h is the separation distance in kilometers, the correlation ranges of 10 and 60 km are specified for the short-range and long-range function, respectively. The correlation ranges 10 and 60 km are chosen because this combination provides best overall fit to the empirical data than other combinations of functions we have experimented. The fitted cross semivariograms by the above equation generally agree well with the empirical data for all earthquake events.

Furthermore, the covariance matrix $\mathbf{C}(h)$ and the cross-correlation matrix $\mathbf{R}(h)$ become

$$\mathbf{C}(h) = \mathbf{B}^1\left[\exp\left(\frac{-3h}{10}\right)\right] + \mathbf{B}^2\left[\exp\left(\frac{-3h}{60}\right)\right], \tag{22}$$

$$\mathbf{R}(h) = \mathbf{P}^1\left[\exp\left(\frac{-3h}{10}\right)\right] + \mathbf{P}^2\left[\exp\left(\frac{-3h}{60}\right)\right], \tag{23}$$

in which $\mathbf{P}^l = [p_{ij}^l]\,(l = 1, 2)$ is the coregionalization matrix, which can be obtained by standardizing the matrix $\mathbf{B}^l$ as follows:

$$p_{ij}^l = \frac{b_{ij}^l}{(\sqrt{b_{ii}^1 + b_{ii}^2}) \times (\sqrt{b_{jj}^1 + b_{jj}^2})}, \qquad l = 1, 2. \tag{24}$$

The above expression can be readily derived using $\rho_{ij}(h) = C_{ij}(h)/\sqrt{C_{ii}(0) \times C_{jj}(0)}$ (equation 13) and $C_{ii}(0) = b_{ii}^1 + b_{ii}^2$ (equation 22). Again, the coregionalization matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ must be positive semidefinite to guarantee the total covariance matrix equation (15) to be positive semidefinite.

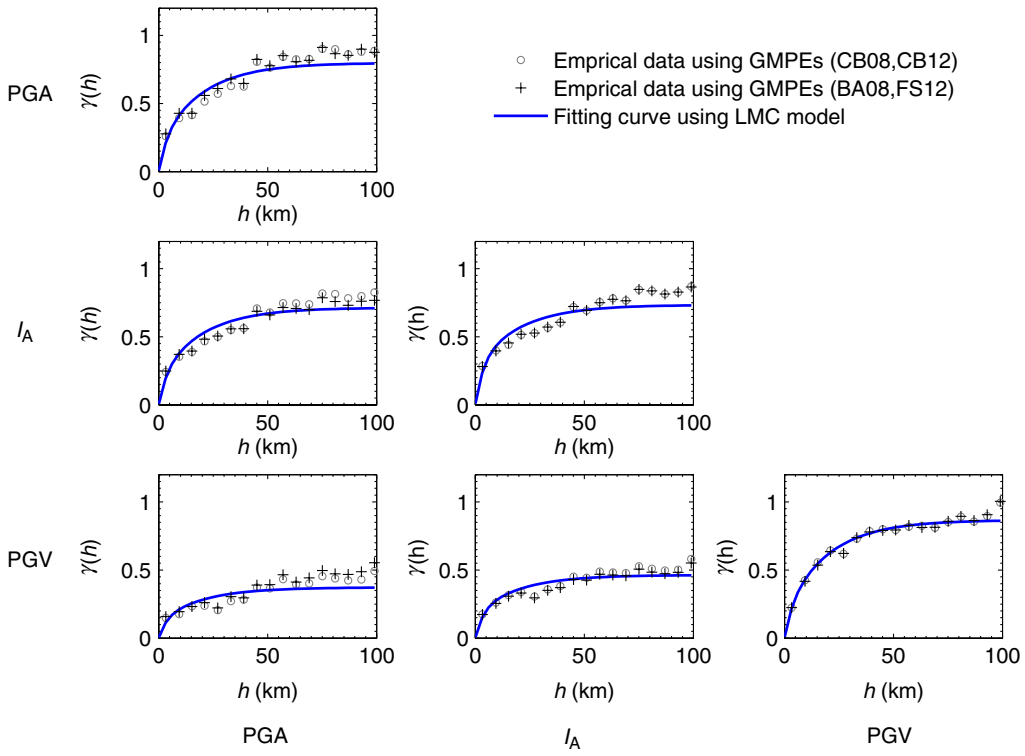

Figure 4. Cross semivariograms and fitted LMC curves for the Niigata earthquake. The color version of this figure is available only in the electronic edition.

Figure 4 shows an example of the LMC fitting for the Niigata event data. Two groups of GMPEs are used to generate the empirical cross-semivariogram data for intraevent residuals for PGA, I_a, and PGV. The first group uses Campbell and Borzorgnia (2008, 2012) GMPEs and the other group uses Boore and Atkinson (2008) model for PGA, PGV, and Foulser-Piggott and Stafford (2012) model for I_a. In general, very similar results can be derived by using different GMPEs, indicating that spatial correlations are not dependent on specific GMPEs. The influence of using different GMPEs is also insignificant for the spatial correlation of spatial accelerations (Du and Wang 2013a). Unless indicated otherwise, the Campbell and Borzorgnia (2008, 2012) GMPEs are used throughout this study.

LMC Model Considering Regional Site Conditions

Spatial Correlations of V_{S30} for Different Regions

The influence of regional geological conditions on the spatial correlation of IMs has been investigated recently (e.g., Jayaram and Baker, 2009; Sokolov *et al*, 2012; Du and Wang, 2013a). Because the regional geological conditions strongly affect the travel path and frequency contents of earthquake waves during propagation, the spatial correlations of the intraevent residuals of IMs are dependent on the spatial distribution of site conditions. For example, the spatial correlation of the intraevent residuals of IMs over a heterogeneous region usually decays faster than that of a homogeneous region. Because V_{S30} values are commonly used in GMPEs to represent local site conditions, it is chosen as an indicator to describe the regional site features. Figure 5 shows empirical semivariograms obtained using the normalized V_{S30} (the set is normalized to zero mean and unit standard deviation for easy comparison) of the Niigata and Northridge earthquake, respectively. An exponential model is used to fit the empirical semivariogram data, such that $\gamma(h) = a[1 - \exp(-3h/R_{V_{S30}})]$, in which $R_{V_{S30}}$ is the correlation range of V_{S30}. A larger value of $R_{V_{S30}}$ implies a more homogeneous regional site condition, such as the case for the Niigata earthquake ($R_{V_{S30}} = 21.8$ km). For the Northridge event, the empirical cross semivariograms for V_{S30} imply an almost independent distribution of V_{S30} values in this region ($R_{V_{S30}} = 0$ km). Based on our previous work (Du and Wang, 2013a), the ranges of V_{S30} are estimated for each event and are summarized in Table 2. The $R_{V_{S30}}$ value varies from 0 to 26 km from event to event, and the median V_{S30} value ranges from 348 to 425 m/s with a standard deviation from 101 to 218 m/s.

Coregionalization Matrices for the Vector IM [PGA, I_a, PGV]

In this study, coregionalization matrices for the vector IM [PGA, I_a, PGV] for 11 events are computed. The influence of regional site conditions on the correlation structure can be further revealed by plotting each entry of the coregionalization matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ against $R_{V_{S30}}$, as shown in Figure 6.

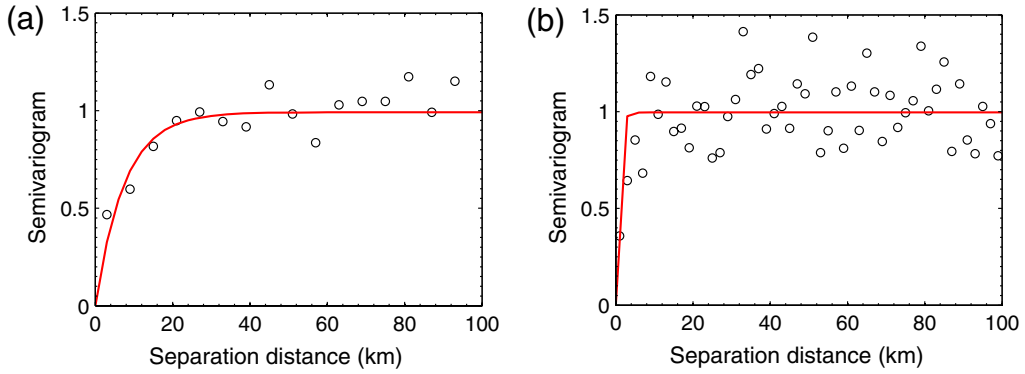

Figure 5. Empirical semivariograms for the normalized V_{S30}. (a) The Niigata earthquake and (b) the Northridge earthquake. The color version of this figure is available only in the electronic edition.

In general, each of the 3×3 components of the short-range matrix $\mathbf{P}^1$ decreases as $R_{V_{S30}}$ increases, while an opposite trend can be observed for the long-range matrix $\mathbf{P}^2$. This is not unexpected as the spatial correlation of IMs usually becomes stronger when $R_{V_{S30}}$ increases (the site condition becomes more homogeneous). Accordingly, the contribution from the short-range function diminishes while the contribution from the long-range function becomes more pronounced as $R_{V_{S30}}$ increases.

Based on Figure 6, the components of $\mathbf{P}^1$ and $\mathbf{P}^2$ matrices can be assumed to vary linearly with $R_{V_{S30}}$ through the following relationships:

$$\mathbf{P}^1 = \mathbf{P}^0 - \mathbf{K}\left(\frac{R_{V_{S30}}}{10}\right), \qquad \mathbf{P}^2 = \mathbf{K}\left(\frac{R_{V_{S30}}}{10}\right), \tag{25}$$

in which $R_{V_{S30}}$ is in the unit of km; $\mathbf{P}^0$ and $\mathbf{K}$ are fitting matrices to be determined. In order to construct a permissible correlation model, both $\mathbf{P}^0$ and $\mathbf{K}$ must be positive semidefinite, which is a necessary condition to guarantee the positive semidefiniteness of $\mathbf{P}^1$ and $\mathbf{P}^2$. First, linear regression analysis is performed to determine $\mathbf{P}^0$ and $\mathbf{K}$ from the empirical data. If the resulted matrices are not positive semidefinite, they can be adjusted manually by slightly increasing the diagonal terms and decreasing the off-diagonal terms.

Table 2
Summary of Regional Site Conditions

Earthquake Name	Median V_{S30} (m/s)	Standard Deviation V_{S30} (m/s)	Correlation Range $R_{V_{S30}}$ (km)
Northridge	422	218	0
Chi-Chi	384	178	26
Tottori	425	174	18.8
Parkfield	395	131	3.5
Niigata	404	168	21.8
Anza	348	118	20.3
Chuetsu	415	167	20.8
Alum Rock	386	144	14.2
Iwate	407	176	8.7
Chino Hills	342	101	14.5
El Mayor–Cucapah	422	180	20.3

The manual adjustment usually does not change the matrices significantly. Finally, the following positive-definite $\mathbf{P}^0$ and $\mathbf{K}$ matrices are obtained as follows:

$$\mathbf{P}^0 = \begin{bmatrix} 1 & 0.91 & 0.65 \\ 0.91 & 1 & 0.71 \\ 0.65 & 0.71 & 1 \end{bmatrix} \quad \text{and}$$

$$\mathbf{K} = \begin{bmatrix} 0.28 & 0.24 & 0.17 \\ 0.24 & 0.22 & 0.16 \\ 0.17 & 0.16 & 0.31 \end{bmatrix}. \tag{26}$$

It is obvious that $\mathbf{P}^2$ is always positive definite for any value of $R_{V_{S30}}$ because it is simply $\mathbf{K}$ multiplying by a positive factor. However, $\mathbf{P}^1$ may not be positive definite for all values of $R_{V_{S30}}$. It is easy to check that for $R_{V_{S30}} = 25$ km,

$$\mathbf{P}^1 = \begin{bmatrix} 0.30 & 0.31 & 0.225 \\ 0.31 & 0.45 & 0.31 \\ 0.225 & 0.31 & 0.225 \end{bmatrix}$$

is positive definite. In view of that $\mathbf{P}^1 = \mathbf{P}^1(R_{V_{S30}} = 25) + \mathbf{K}(25 - R_{V_{S30}})/10$, $\mathbf{P}^1$ is guaranteed to be positive definite for all $R_{V_{S30}} \leq 25$ km following the mathematical theorem that the summation of two positive-definite matrices is always positive definite. For the cases where $R_{V_{S30}}$ is greater than 25 km, it is recommended that $R_{V_{S30}} = 25$ km should be used to evaluate the matrices for practical purposes. As is also shown in Figure 6, the linear model equations (25) and (26) can reasonably approximate the general trend of each of the 3×3 components of $\mathbf{P}^1$ and $\mathbf{P}^2$. Finally, the spatial correlation matrix is expressed as a function of h and $R_{V_{S30}}$ as follows:

$$\mathbf{R}(h, R_{V_{S30}}) = \mathbf{P}^0 \exp\left(\frac{-3h}{10}\right) + \mathbf{K}\left(\frac{R_{V_{S30}}}{10}\right) \times \left[\exp\left(\frac{-3h}{60}\right) - \exp\left(\frac{-3h}{10}\right)\right]. \tag{27}$$

Because the coefficient $\exp(\frac{-3h}{60}) - \exp(\frac{-3h}{10})$ is always positive for $h > 0$, the correlation coefficients will increase

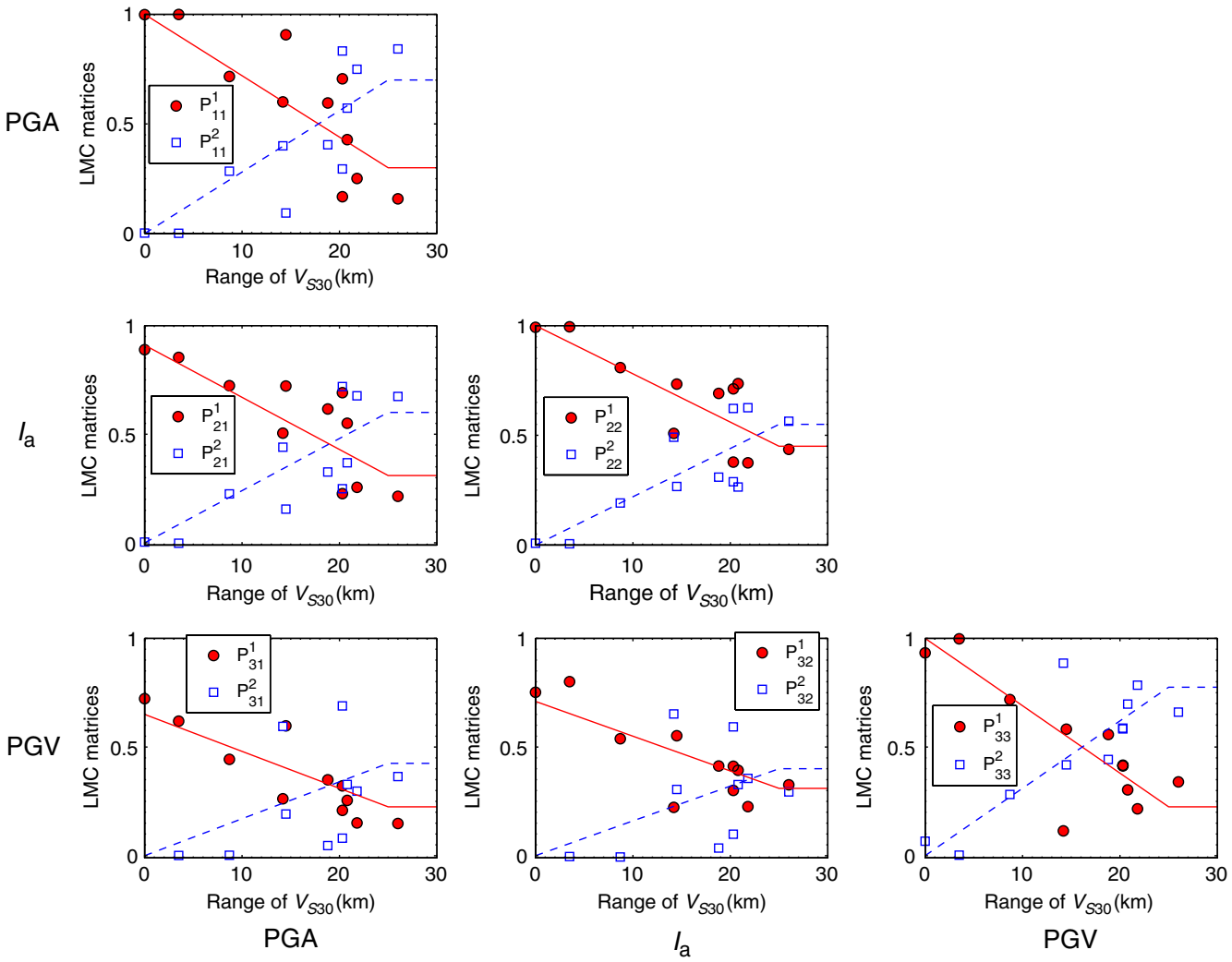

Figure 6. Distribution of coregionalization matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ versus the range of V_{S30} ($R_{V_{S30}}$) for 11 earthquakes (points, empirical data; lines, predicted curves using equation 25). The color version of this figure is available only in the electronic edition.

with an increasing $R_{V_{S30}}$. For example, considering $R_{V_{S30}} = 10$ and 20 km, the correlation matrix for the vector IM [PGA, I_a, PGV] at a separation distance $h = 5$ km can be evaluated as

$$\mathbf{R}(h=5, R_{V_{S30}}=20) = \begin{bmatrix} 0.53 & 0.47 & 0.33 \\ 0.47 & 0.47 & 0.34 \\ 0.33 & 0.34 & 0.57 \end{bmatrix} \quad \text{and}$$

$$\mathbf{R}(h=5, R_{V_{S30}}=10) = \begin{bmatrix} 0.38 & 0.34 & 0.24 \\ 0.34 & 0.35 & 0.25 \\ 0.24 & 0.25 & 0.40 \end{bmatrix}. \tag{28}$$

The intrinsic structure of the proposed model can be further explored by considering several special cases. If $h = 0$, the spatial correlation matrix is reduced to the local correlation matrix $\mathbf{R}(0)$, that is, the correlation matrix for the vector IM [PGA, I_a, PGV] at an individual site, as follows:

$$\mathbf{R}(h=0) \triangleq \mathbf{R}(0) = \mathbf{P}^1 + \mathbf{P}^2 = \mathbf{P}^{01} = \begin{bmatrix} 1 & 0.91 & 0.65 \\ 0.91 & 1 & 0.71 \\ 0.65 & 0.71 & 1 \end{bmatrix}. \tag{29}$$

The above local correlation does not depend on $R_{V_{S30}}$ by construction. In fact, the above matrix is consistent with that obtained by Campbell and Bozorgnia (2012) (the difference is within ± 0.04) using the PEER-Next Generation Attenuation (NGA) database, which reads

$$\mathbf{R}(0) = \begin{bmatrix} \rho_{\mathrm{PGA,PGA}} & \rho_{\mathrm{PGA,Ia}} & \rho_{\mathrm{PGA,PGV}} \\ \rho_{\mathrm{PGA,Ia}} & \rho_{\mathrm{Ia,Ia}} & \rho_{\mathrm{Ia,PGV}} \\ \rho_{\mathrm{PGA,PGV}} & \rho_{\mathrm{Ia,PGV}} & \rho_{\mathrm{PGV,PGV}} \end{bmatrix} = \begin{bmatrix} 1 & 0.88 & 0.69 \\ 0.88 & 1 & 0.74 \\ 0.69 & 0.74 & 1 \end{bmatrix}. \tag{30}$$

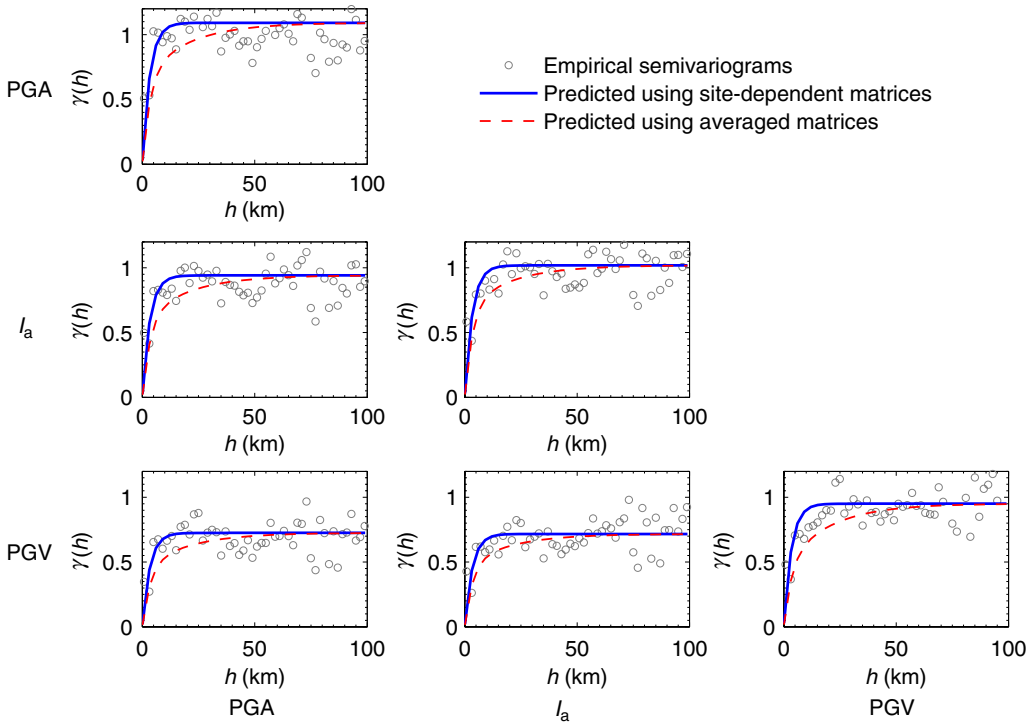

Figure 7. Comparison between the empirical cross semivariograms and model predictions for the Northridge earthquake. The color version of this figure is available only in the electronic edition.

Furthermore, for a heterogeneous site condition ($R_{V_{S30}}$ approaches zero), the spatial correlation matrix further reduces to a simple form:

$$\mathbf{R}(h) = \mathbf{R}(0)\exp\left(\frac{-3h}{10}\right), \quad \text{if } R_{V_{S30}} = 0 \text{ km}. \tag{31}$$

The above equation implies that all the components of the local correlation $\mathbf{R}(0)$ of the vector IM decays exponentially with increasing separation distance h at a constant rate, which can be described using an exponential model with a correlation range of 10 km. The results are consistent with our previous study (Du and Wang, 2013a).

Considering the influence of site conditions, equation (25) assumes that both $\mathbf{P}^1$ and $\mathbf{P}^2$ vary linearly with $R_{V_{S30}}$ by the same rate factors designated by $\mathbf{K}$. This assumption is proposed due to its mathematical simplicity and its consistency with empirical data shown in Figure 6. A larger $\mathbf{K}$ value implies the spatial correlation is more strongly influenced by the site conditions. In general, the site condition has the strongest influence on the direct correlations of PGA and PGV, followed by the direct and cross correlations of I_a. The site condition has less influence on the cross correlations related to PGV. In the cases that the local site information is not available, the following $\mathbf{P}^1$ and $\mathbf{P}^2$ matrices averaged over 11 events can be used in equation (23):

$$\mathbf{P}^1_{\text{avg}} = \begin{bmatrix} 0.61 & 0.57 & 0.38 \\ 0.57 & 0.67 & 0.45 \\ 0.38 & 0.45 & 0.50 \end{bmatrix},$$
$$\mathbf{P}^2_{\text{avg}} = \begin{bmatrix} 0.39 & 0.34 & 0.24 \\ 0.34 & 0.33 & 0.24 \\ 0.24 & 0.24 & 0.50 \end{bmatrix}. \tag{32}$$

Finally, the accuracy of the model prediction is examined using the empirical cross-semivariogram data for all 11 events. Figures 7 and 8 show the predicted cross semivariograms of the vector IM [PGA, I_a, PGV] using equations (25) and (26) for the Northridge and Chi-Chi earthquakes, respectively. These two events are selected to represent a heterogeneous and a homogeneous region, whose $R_{V_{S30}}$ values are estimated as 0 km (Northridge) and 26 km (Chi-Chi), respectively. Compared with the predicted curves using the averaged coregionalization matrices equation (32), the regional-dependent model using equations (25) and (26) better matches the empirical data for most cases in both events, demonstrating the model's capability in capturing the influence of regional site conditions on the spatial correlation of the vector IM. It is also to be noted that the site-dependant LMC model does not always provide the best fit to the empirical data, which can be observed in some plots (e.g., PGA–PGV cross semivariograms for the Chi-Chi earthquake in Fig. 8). Because of considerable scatter of data shown in

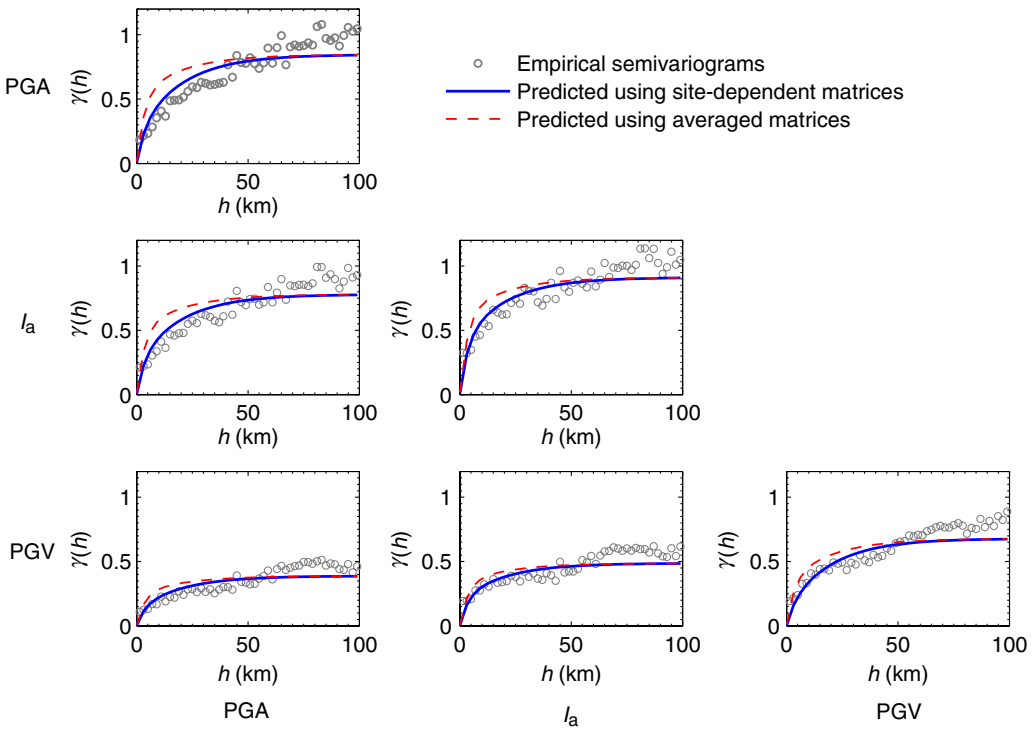

Figure 8. Comparison between the empirical cross semivariograms and model predictions for the Chi-Chi earthquake. The color version of this figure is available only in the electronic edition.

Figure 6, it is difficult to precisely fit the empirical data using the proposed model for each event.

Coregionalization Matrices for Spectral Acceleration at Multiple Periods

Similar analysis is also performed to study the influence of regional site conditions on the spatial cross correlations for SAs at multiple periods. First, empirical cross semivariograms of SAs at nine periods (namely, 0.01, 0.1, 0.2, 0.5, 1, 2, 5, 7.5, and 10 s) are obtained. Based on the empirical data, the following LMC structure is proposed:

$$\mathbf{R}(h) = \mathbf{P}^1 \exp\left(\frac{-3h}{10}\right) + \mathbf{P}^2 \exp\left(\frac{-3h}{70}\right), \tag{33}$$

in which h is in kilometers, the correlation ranges are chosen as 10 and 70 km, the combination provides an overall best fit to the empirical data. Considering the influence of regional site conditions, it is further postulated that the components of $\mathbf{P}^1$ and $\mathbf{P}^2$ matrices vary linearly with $R_{V_{S30}}$ through the following relationships:

$$\mathbf{P}^1 = \mathbf{P}^{01}_{\mathrm{SA}} - \mathbf{K}_{\mathrm{SA}}\left(\frac{R_{V_{S30}}}{10}\right),$$
$$\mathbf{P}^2 = \mathbf{P}^{02}_{\mathrm{SA}} + \mathbf{K}_{\mathrm{SA}}\left(\frac{R_{V_{S30}}}{10}\right), \tag{34}$$

in which $R_{V_{S30}}$ is in the unit of kilometers; the matrices $\mathbf{P}^{01}_{\mathrm{SA}}$, $\mathbf{P}^{02}_{\mathrm{SA}}$, $\mathbf{K}_{\mathrm{SA}}$ need to be determined from the empirical data. Similar to the previous discussion, matrices $\mathbf{P}^{01}_{\mathrm{SA}}$, $\mathbf{P}^{02}_{\mathrm{SA}}$, $\mathbf{K}_{\mathrm{SA}}$ are required to be positive semidefinite as they are the necessary condition to construct a permissible correlation model. Furthermore, equation (33) reduces to $\mathbf{R}(0) = \mathbf{P}^{01}_{\mathrm{SA}} + \mathbf{P}^{02}_{\mathrm{SA}}$ if $h = 0$. Therefore, $\mathbf{P}^{01}_{\mathrm{SA}}$ and $\mathbf{P}^{02}_{\mathrm{SA}}$ can be regarded as partition of the local-correlation matrix, and the summation of diagonal elements of $\mathbf{P}^{01}_{\mathrm{SA}}$, $\mathbf{P}^{02}_{\mathrm{SA}}$ must equal unity. Based on regression analysis from the empirical cross-semivariograms data followed by manual adjustment, $\mathbf{P}^{01}_{\mathrm{SA}}$, $\mathbf{P}^{02}_{\mathrm{SA}}$, and $\mathbf{K}_{\mathrm{SA}}$ matrices are obtained and listed in Tables 3–5. It is easy to verify that all these matrices are positive definite. Furthermore, $\mathbf{P}^1$ and $\mathbf{P}^2$ are guaranteed to be positive definite for $R_{V_{S30}} \le 25$ km. Similar to the previous case, it is recommended that if $R_{V_{S30}} > 25$ km, the results for $R_{V_{S30}} = 25$ km should be used. Finally, the regional-dependent LMC model becomes

$$\mathbf{R}(h, R_{V_{S30}}) = \mathbf{P}^{01}_{\mathrm{SA}} \exp\left(\frac{-3h}{10}\right) + \mathbf{P}^{02}_{\mathrm{SA}} \exp\left(\frac{-3h}{70}\right) + \mathbf{K}\left(\frac{R_{V_{S30}}}{10}\right)\left[\exp\left(\frac{-3h}{70}\right) - \exp\left(\frac{-3h}{10}\right)\right]. \tag{35}$$

It is worthwhile to discuss some interesting features of the proposed model. First, all spatial cross-correlation coefficients increase with an increased $R_{V_{S30}}$ value. The diagonal element in $\mathbf{P}^{01}_{\mathrm{SA}}$ matrix generally decreases as the period becomes longer and vice versa for $\mathbf{P}^{02}_{\mathrm{SA}}$ matrix. The correlations for short periods ($T \le 0.5$ s) are mostly controlled by the short-range basic function. Contribution from the long-range basic function becomes more pronounced when periods becomes longer. Relatively small values appear in the lower left

Table 3
$\mathbf{P}_{SA}^{01}$ Matrix

Period (s)	0.01	0.1	0.2	0.5	1	2	5	7.5	10
0.01	0.96								
0.1	0.9	0.96							
0.2	0.8	0.81	0.93			Symmetric			
0.5	0.5	0.36	0.44	0.76					
1	0.15	0.08	0.1	0.25	0.62				
2	0.09	0.04	0.05	0.17	0.45	0.54			
5	0.1	0.05	0.09	0.14	0.34	0.42	0.47		
7.5	0.09	0.05	0.08	0.13	0.37	0.42	0.46	0.57	
10	0.04	0.02	0.05	0.07	0.31	0.35	0.39	0.4	0.56

Table 4
$\mathbf{P}_{SA}^{02}$ Matrix

Period (s)	0.01	0.1	0.2	0.5	1	2	5	7.5	10
0.01	0.04								
0.1	0	0.04							
0.2	0.01	0.01	0.07			Symmetric			
0.5	0.04	0	0.08	0.24					
1	0.08	0.01	0.08	0.28	0.38				
2	0.02	0	0.01	0.2	0.22	0.46			
5	0.02	0	0	0.15	0.23	0.32	0.53		
7.5	0	0	0	0.13	0.18	0.25	0.43	0.43	
10	0.02	0	0	0.13	0.19	0.25	0.42	0.41	0.44

Table 5
$\mathbf{K}_{SA}$ Matrix

Period (s)	0.01	0.1	0.2	0.5	1	2	5	7.5	10
0.01	0.28								
0.1	0.26	0.27							
0.2	0.2	0.21	0.2			Symmetric			
0.5	0.13	0.1	0.1	0.11					
1	0	0	0	0	0.14				
2	0	0	0	0	0.11	0.11			
5	0	0	0	0	0.08	0.09	0.11		
7.5	0	0	0	0	0.1	0.11	0.12	0.14	
10	0	0	0	0	0.1	0.12	0.12	0.13	0.17

block of $\mathbf{P}_{SA}^{01}$ and $\mathbf{P}_{SA}^{02}$ matrices, implying that the cross correlations between SA at the short periods ($T \leq 0.5$ s) and long periods ($T > 1$ s) are relatively weak. $\mathbf{K}_{SA}$ matrix controls the rate of influence of the regional site conditions. In general, SA at the short periods is more influenced by the regional site condition than that at the long period range. As observed through empirical data, $R_{V_{S30}}$ will not significantly influence the cross correlations between SA at short periods and those at long periods. Therefore, the corresponding elements of $\mathbf{K}_{SA}$ are assigned to zeros for simplicity.

The current model is different from the previous study (Loth and Baker, 2013) in a couple of ways. First, the Loth and Baker model does not consider the influence of site conditions. Second, three basic exponential functions have been used to construct the LMC model for SA in Loth and Baker (2013), in which the correlation ranges of the basic functions are chosen as 0 km (i.e., the nugget effect), 20 km, and 70 km. In our study, a relatively small correlation range (i.e., 10 km) is specified for the short-range basic function based on trials of several combinations. We found that inclusion of the nugget effect component is not particularly advantageous. On the other hand, the relative nugget effect may be associated with lack of data. It may decrease as more and better data become available (Goovaerts, 1997). Therefore, the current LMC model is adopted due to its simplicity and ability to capture the major spatial features of the vector IMs. Finally, Loth and Baker (2013) used the GMPE proposed by Boore and Atkinson (2008) to compute the predicted SA values, while the Campbell and Bozorgnia (2008) model is adopted in this study. As stated before,

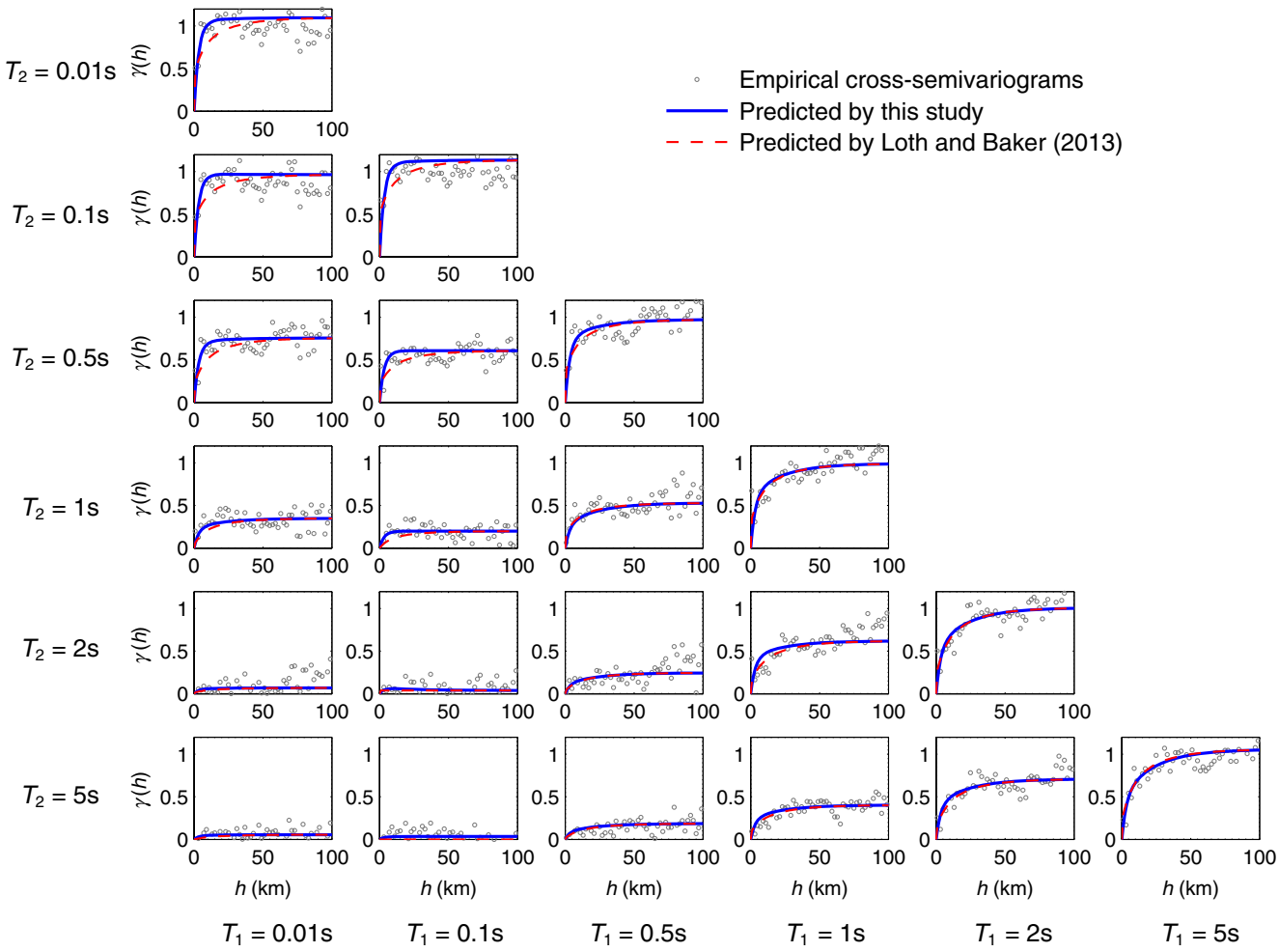

Figure 9. Predicted cross semivariograms using the LMC method for the Northridge earthquake. The color version of this figure is available only in the electronic edition.

the choice of GMPEs would not influence the spatial correlation significantly.

Figures 9 and 10 show the performance of model prediction for the Northridge and the Chi-Chi earthquake, respectively, compared with the empirical cross-semivariogram data and model prediction by Loth and Baker (2013). Because regional site conditions usually have relatively strong influence on SAs at short periods, the proposed model improves the cross-semivariogram prediction especially at the short-period range (0.01–0.5 s, the upper left subplots). However, the improvement is rather limited for SAs at long periods. For a number of cases, the model does not give improved performance over Loth and Baker model, especially for the correlations at the long-period range of the Chi-Chi earthquake.

Discussions and Conclusions

This study investigates a new spatial cross-correlation structure for two sets of vector IMs. The first set includes the PGA, I_a, and PGV, while the second set consists of the SA at nine different periods. Eleven recent earthquakes that occurred in California, Japan, Taiwan, and Mexico are used to obtain the empirical cross semivariograms of the vector IMs. The LMC model is utilized to construct a permissible spatial correlation model, which is the combination of a short-range and a long-range exponential function with different decaying rate. The correlation range of V_{S30}, $R_{V_{S30}}$, is a good indicator to characterize the regional geological conditions. Although some other geological features such as the azimuthal path may also influence the spatial correlation structure, they are reserved for a future study.

The study found that the coregionalization matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ vary significantly with $R_{V_{S30}}$. Specifically, a linear model has been proposed such that the short-range matrix $\mathbf{P}^1$ increases and the long-range matrix $\mathbf{P}^2$ decreases linearly with $R_{V_{S30}}$, as shown in equations (25) and (34). The proposed model has a distinctive advantage that it can quantitatively analyze the spatial cross correlations for the vector IMs based on the characteristics of regional geology, meanwhile the model guarantees a positive-definite covariance matrix for any reasonable value of $R_{V_{S30}}$ (no larger than

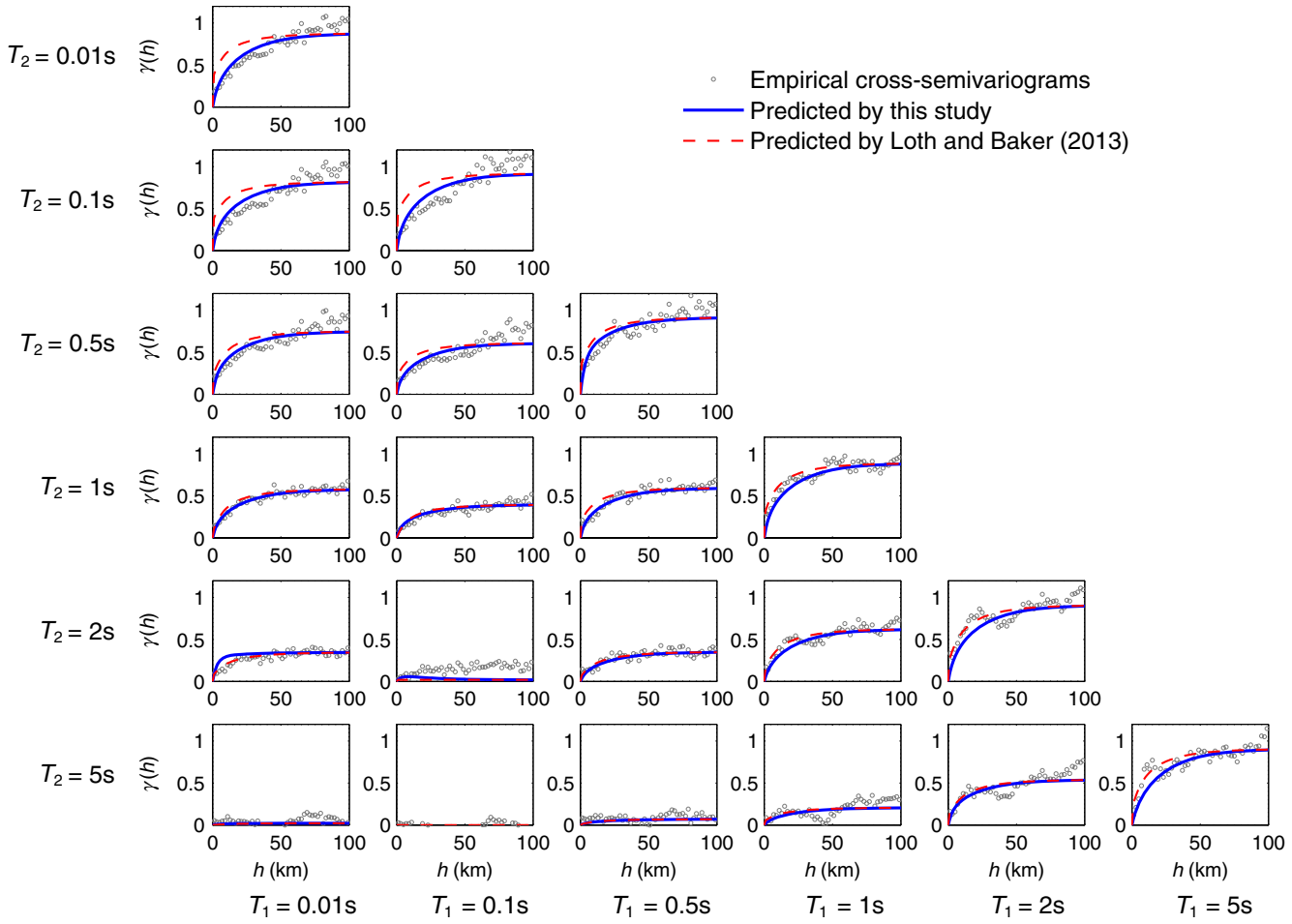

Figure 10. Predicted cross semivariograms using the LMC method for the Chi-Chi earthquake. The color version of this figure is available only in the electronic edition.

25 km)—a mathematical condition required for stochastic generation of the spatially correlated random fields. Compared with the averaged model, the site-dependent LMC model can better approximate the empirical cross semivariograms for the vector IM [PGA, I_a, PGV], although it does not always provide the best fit to the empirical data. On the other hand, the site-dependent LMC model only improves the cross-semivariogram prediction of SAs at the short-period range (0.01–0.5 s) compared with Loth and Baker model, and the improvement is rather limited for SAs at long periods.

Although the LMC matrices for SAs are provided only for nine periods, these matrices can be linearly interpolated to derive the spatial cross correlations for other periods. In the case that the interpolated coregionalization matrix are not positive definite, the negative eigenvalues of the coregionalization matrix should be set to zero in order to obtain an admissible model. One may refer to Loth and Baker (2013) for more details. The above modification usually does not change the coregionalization matrix significantly. For a majority of cases we have experimented using three or four interpolating periods, the maximum absolute change in the components of the coregionalization matrix is less than 0.05, which is insignificant for practical purposes.

Finally, an example is presented to illustrate the performance of the spatial correlation model. Considering a 30 km × 30 km hypothetical region divided into 900 1 km × 1 km cells, a point source is located at the origin. For simplicity, it is assumed that a scenario earthquake with moment magnitude 7 occurs at the point source. A heterogeneous and a homogeneous site conditions are considered for this region. The V_{S30} values follow a normal distribution with a mean value of 400 m/s and a standard deviation of 150 m/s over the region. The spatial correlation of V_{S30} follows the exponential model with correlation ranges assumed as 0 and 25 km, respectively, for each case. Campbell and Bozorgnia (2008, 2012) GMPEs are used to calculate the median predicted PGA, PGV, and I_a fields for this earthquake scenario. Then, the intraevent residuals are generated by Copula functions (Nelson, 2006) at each site given the mean (zero) and the covariance matrix.

Figure 11 compares two realizations of the jointly distributed fields of PGA, I_a, and PGV. Apparently, the spatial correlation of ground-motion residuals affects the distribution of

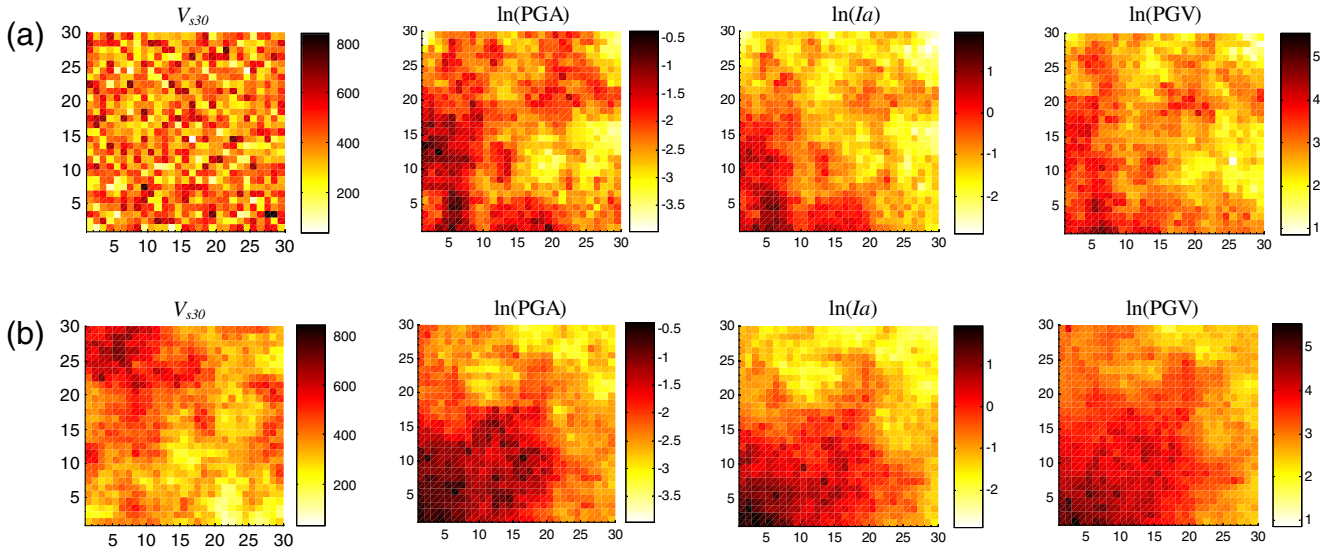

Figure 11. Examples of realized V_{S30} and vector-IM field over a region for different $R_{V_{S30}}$ values: (a) $R_{V_{S30}} = 0$ km, (b) $R_{V_{S30}} = 25$ km. PGA, I_a, and PGV are in the unit of g, m/s, and cm/s, respectively; V_{S30} is in the unit of m/s. The color version of this figure is available only in the electronic edition.

vector IM fields, as a larger $R_{V_{S30}}$ corresponds to more uniformly distributed IM fields. The spatial correlation model and realization of vector-IM fields such as Figure 11 can be conveniently used in regional-specific seismic-hazard analysis and loss estimation of spatially distributed infrastructure. Recently, many empirical models have been developed to predict the earthquake-induced slope displacements using a single or a combination of PGA, I_a, and PGV. Using the proposed spatial correlation model for the vector IM [PGA, I_a, PGV], a fully probabilistic seismic-hazard analysis framework has been developed for spatially distributed slopes (Du and Wang, 2013b). It is also interesting to mention that the vector IM [PGA, I_a, PGV] can be used to develop useful proxy for characterizing cyclic structural response. For example, the spatial correlation model will be useful in conditional hazard analysis of spatially distributed structures by extending the methodology developed by Iervolino *et al.* (2010). All these examples demonstrate that the spatial cross-correlation models proposed in this study can be conveniently used in regional-specific seismic-risk analysis and loss estimation of spatially distributed infrastructure using vector IMs.

Data and Resources

Strong-motion data used in this study are obtained from resources in the public domain, including the CESMD strong-motion database (http://strongmotioncenter.org/; last accessed September 2011), the COSMOS strong-motion database (http://www.cosmos-eq.org/; last accessed September 2011), the K-NET strong-motion database (http://www.kyoshin.bosai.go.jp/; last accessed September 2011), and the PEER NGA database (http://peer.berkeley.edu/nga/; last accessed September 2011). The V_{S30} values for Taiwan sites are obtained from TSMIP (http://egdt.ncree.org.tw/; last accessed August 2012).

Acknowledgments

The authors acknowledge financial support from Hong Kong Research Grants Council (RGC) Grant Number 620311 and Direct Allocation Grants DAG12EG07-3, FSGRF13EG09 (HKUST/RGC). Any opinions, findings, and conclusions or recommendations expressed in this material are ours and do not necessarily reflect those of the sponsor.

References

Abrahamson, N. A., and R. R. Youngs (1992). A stable algorithm for regression analyses using the random effects model, *Bull. Seismol. Soc. Am.* **82,** 505–510.

Arias, A. (1970). Measure of earthquake intensity, in R. J. Hansen (Editor), *Seismic Design for Nuclear Power Plants*, Massachusetts Institute of Technology Press, Cambrige, Massachusetts, 438–483.

Boore, D. M., and G. A. Atkinson (2008). 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, *Earthq. Spectra* **24,** no. 1, 99–138.

Campbell, K. W., and Y. Bozorgnia (2008). NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.1 to 10 s, *Earthq. Spectra* **24,** no. 1, 139–171.

Campbell, K. W., and Y. Bozorgnia (2012). A comparison of ground motion prediction equations for Arias intensity and cumulative absolute velocity developed using a consistent database and functional form, *Earthq. Spectra* **28,** no. 3, 931–941.

Darragh, B., W. Silva, and N. Gregor (2004). Strong motion record processing procedures for the PEER center, in *Proc. of COSMOS Workshop on Strong-Motion Record Processing*, Richmond, California, 26–27 May, 1–12.

Du, W., and G. Wang (2013a). Intra-event spatial correlations for cumulative absolute velocity, Arias intensity and spectral accelerations based on regional site conditions, *Bull. Seismol. Soc. Am.* **103,** no. 2A, 1117–1129.

Du, W., and G. Wang (2013b). Fully probabilistic seismic displacement analysis of spatially distributed slopes using spatially correlated vector intensity measures, *Earthq. Eng. Struct. Dynam.* (Published online), doi: 10.1002/eqe.2365.

Electrical Power Research Institute (EPRI) (1988). A criterion for determining exceedance of the operating basis earthquake, Report No. EPRI NP-5930, Palo Alto, California.

Esposito, S., and I. Iervolino (2011). PGA and PGV spatial correlation models based on European multievent datasets, *Bull. Seismol. Soc. Am.* **101,** no. 5, 2532–2541.

Esposito, S., and I. Iervolino (2012). Spatial correlation of spectral acceleration in European data, *Bull. Seismol. Soc. Am.* **102,** no. 6, 2781–2788.

Foulser-Piggott, R., and P. J. Stafford (2012). A predictive model for Arias intensity at multiple sites and consideration of spatial correlations, *Earthq. Eng. Struct. Dynam.* **41,** no. 3, 431–451.

Goda, K. (2011). Interevent variability of spatial correlation of peak ground motions and response spectra, *Bull. Seismol. Soc. Am.* **101,** no. 5, 2522–2531.

Goda, K., and H. P. Hong (2008). Spatial correlation of peak ground motions and response spectra, *Bull. Seismol. Soc. Am.* **98,** no. 1, 354–365.

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

Goulard, M., and M. Voltz (1992). Linear coregionalization model: Tools for estimation and choice of crossvariogram matrix, *Math. Geol.* **24,** no. 3, 269–286.

Iervolino, I., M. Giorgio, C. Galasso, and G. Manfredi (2010). Conditional hazard maps for secondary intensity measures, *Bull. Seismol. Soc. Am.* **100,** no. 6, 3312–3319.

Jayaram, N., and J. W. Baker (2008). Statistical tests of the joint distribution of spectral acceleration values, *Bull. Seismol. Soc. Am.* **98,** no. 5, 2231–2243.

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground-motion intensities, *Earthq. Eng. Struct. Dynam.* **38,** 1687–1708.

Jayaram, N., and J. W. Baker (2010). Considering spatial correlation in mixed-effects regression and the impact on ground-motion models, *Bull. Seismol. Soc. Am.* **100,** no. 6, 3295–3303.

Jeon, S., and T. D. O'Rourke (2005). Northridge earthquake effects on pipelines and residential buildings, *Bull. Seismol. Soc. Am.* **95,** no. 1, 294–318.

Journel, A. G., and C. J. Huijbregts (1978). *Mining Geostatistics*, Academic Press, London, 600 pp.

Joyner, W. B., and D. M. Boore (1993). Methods for regression analysis of strong-motion data, *Bull. Seismol. Soc. Am.* **83,** 469–487.

Kaklamanos, J., and L. G. Baise (2011). Model validations and comparisons of the next generation attenuation of ground motions (NGA-West) project, *Bull. Seismol. Soc. Am.* **101,** no. 1, 160–175, Table S1 in the electronic supplement.

Kuo, C. H., K. L. Wen, H. H. Hsieh, C. M. Lin, T. M. Chang, and K. W. Kuo (2012). Site classification and V_{S30} estimation of free-field TSMIP stations using the logging data of EGDT, *Eng. Geol.* **129–130,** 68–75.

Lee, R., and A. S. Kiremidjian (2007). Uncertainty and correlation for loss assessment of spatially distributed systems, *Earthq. Spectra* **23,** no. 4, 743–770.

Loth, C., and J. W. Baker (2013). A spatial cross-correlation model of ground motion spectral accelerations at multiple periods, *Earthq. Eng. Struct. Dynam.* **42,** no. 3, 319–475.

Nelson, R. B. (2006). *An Introduction to Copulas*, Springer Series in Statistics, Springer, New York, 269 pp.

Rathje, E. M., and G. Saygili (2008). Probabilistic seismic hazard analysis for the sliding displacement of slopes: scalar and vector approaches, *J. Geotech. Geoenvir. Eng.* **134,** no. 6, 804–814.

Sokolov, V., and F. Wenzel (2011). Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation, *Earthq. Eng. Struct. Dynam.* **40,** 993–1009.

Sokolov, V., F. Wenzel, W. Y. Jean, and K. L. Wen (2010). Uncertainty and spatial correlation of earthquakes ground motion in Taiwan, *Terr. Atmos. Ocean. Sci.* **21,** no. 6, 905–921.

Sokolov, V., F. Wenzel, W. Y. Jean, and K. L. Wen (2012). On the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: Analysis of PGA data from TSMIP (Taiwan) network, *Bull. Earthq. Eng.* **10,** 1402–1429.

Wang, M., and T. Takada (2005). Macrospatial correlation model of seismic ground motions, *Earthq. Spectra* **21,** no. 4, 1137–1156.

Department of Civil and Environmental Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon
Hong Kong SAR, China
gwang@ust.hk
dwq@ust.hk

Manuscript received 11 March 2013;
Published Online 5 November 2013