Earthquake Engineering & Structural Dynamics
The Journal of the International Association for Earthquake Engineering

EARTHQUAKE ENGINEERING & STRUCTURAL DYNAMICS
Earthquake Engng Struct. Dyn. 2015; **44**:775–794
Published online 16 October 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/eqe.2487

Stochastic simulation of regionalized ground motions using wavelet packets and cokriging analysis

Duruo Huang and Gang Wang*,†

Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong

SUMMARY

Performance-based earthquake engineering often requires ground-motion time-history analyses to be performed, but very often, ground motions are not recorded at the location being analyzed. The present study is among the first attempt to stochastically simulate spatially distributed ground motions over a region using wavelet packets and cokriging analysis. First, we characterize the time and frequency properties of ground motions using the wavelet packet analysis. The spatial cross-correlations of wavelet packet parameters are determined through geostatistical analysis of regionalized ground-motion data from the Northridge and Chi-Chi earthquakes. It is observed that the spatial cross-correlations of wavelet packet parameters are closely related to regional site conditions. Furthermore, using the developed spatial cross-correlation model and the cokriging technique, wavelet packet parameters at unmeasured locations can be best estimated, and regionalized ground-motion time histories can be synthesized. Case studies and blind tests using data from the Northridge and Chi-Chi earthquakes demonstrate that the simulated ground motions generally agree well with the actual recorded data. The proposed method can be used to stochastically simulate regionalized ground motions for time-history analyses of distributed infrastructure and has important applications in regional-scale hazard analysis and loss estimation.

Received 27 April 2014; Revised 1 September 2014; Accepted 2 September 2014

KEY WORDS: ground-motion simulation; spatial cross-correlation; cokriging estimate; wavelet packet analysis

1. INTRODUCTION

Studying the spatial distribution of ground motions is important in regional seismic hazard assessment and loss estimation [1, 2]. Intensity measures (IMs), such as peak ground acceleration, peak ground velocity, and spectral acceleration, are most commonly used to characterize ground motions. In recent years, the spatial correlations of various IMs at multiple sites have been actively researched. Models have been developed to quantify significant spatial correlations of various scalar IMs [2–6], as well as spatial cross-correlations of vector IMs [6, 7]. Studies on ground-motion data in California, Taiwan, and Japan have also indicated that the spatial correlations of IMs can be significantly influenced by regional site conditions [6, 8], implying that region-specific correlation models should be developed for applications. These spatial correlation models for IMs can be used to develop rigorous frameworks for seismic risk and regional loss assessment of spatially distributed structures provided that the structural responses and IMs are well correlated, and various sources of uncertainty and variability are incorporated in the analyses (e.g., [9, 10]).

*Correspondence to: Gang Wang, Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.
†E-mail: gwang@ust.hk

A ground-motion time history is a complex, transient time sequence, and an IM describes only one aspect of its many attributes. In the performance-based earthquake engineering, ground-motion time histories are often needed, but very often, they are not recorded at the location being analyzed. Motivated by the previous studies on the spatial correlation of IMs, a stochastic method to simulate spatially distributed time histories over a region is developed in this study. This method represents a big step toward more accurate performance-based hazard analysis and loss estimation on the regional scale using time-history analyses. It should be emphasized that the present study should not be confused with several previous studies on the spatial variation of strong ground motions (SVSGM), which has been used extensively in multiple-input analysis of structures [11]. First, SVSGM studies are mainly focused on quantifying the spatial variation in ground-motion time histories that are recorded within a short separation distance [12–14] using closely spaced seismograph arrays (e.g., the SMART 1 array in Taiwan, the USGS Parkfield Seismograph Array, and the Borrego Valley Differential Array in California). Data obtained from these closely spaced arrays allow empirical 'coherency functions' to be derived to measure the similarity among frequency characteristics of waves along the travel path and to quantify the wave-passage effects and site-response effects. Based on nearby earthquake recordings, spatially varying ground motions can be simulated using the coherency functions [15–20]. However, it should be noted that the coherency functions usually decay quickly with distance. The coherency relationship usually vanishes when the separation distance is greater than 5 km. On the other hand, ground-motion IMs can still be strongly correlated even at a separation distance greater than 10 km [8]. Therefore, the SVSGM technique is not suitable for simulating incoherent ground motions that are situated several kilometers apart.

The present study is among the first attempt to stochastically simulate ground motions on a regional scale. First, we characterize spatial correlations and spatial cross-correlations of important stochastic measures of ground motions, based on well-populated regionalized ground-motion data recorded from the 1994 Northridge earthquake in California and the 1999 Chi-Chi earthquake in Taiwan. For this purpose, we develop correlation models for 13 wavelet packet parameters used in a recently developed stochastic model by Yamamoto and Baker [21]. The model uses wavelet packet analysis to characterize the wave energy, and the time-domain and frequency-domain statistics of ground motions at an individual site, based on seismological parameters such as earthquake magnitude, distance, and local site conditions. The spatial correlation model that we develop in this study for the wavelet parameters extends the Yamamoto–Baker model to regional-scale applications, by simulating spatially correlated ground motions at multiple locations for a given scenario earthquake. It is interesting to mention that the previous scope of work has also been attempted in a recent study using stochastic point-source simulation [22], where the point-source model is modified to prescribe a spatial correlation and coherency structure.

The second scope of the present study is to conditionally simulate ground-motion time histories at unmeasured sites based on strong-motion data recorded in that region. Using the developed spatial cross-correlation model and the cokriging technique, wavelet parameters at unmeasured locations can be best estimated, from which ground-motion time histories can be synthesized. As part of this study, case studies are conducted using the Northridge and Chi-Chi earthquake data. The model capability is verified in blind tests by comparing simulated ground motions with the actual recorded data. It is also worth pointing out that, if not unlikely, it is not entirely clear how the conditional simulation could be conducted using the point-source model in [22].

2. WAVELET PACKET ANALYSIS OF GROUND MOTIONS

Most recently, a stochastic model was proposed by Yamamoto and Baker [21] to simulate nonstationary ground motions using the wavelet packet transform. The wavelet method decomposes an acceleration time history $x(t)$ into a collection of wavelet packets localized in time (t) and frequency (f) domain. The wavelet packet coefficients for each packet are defined as follows:

DOI: 10.1002/eqe

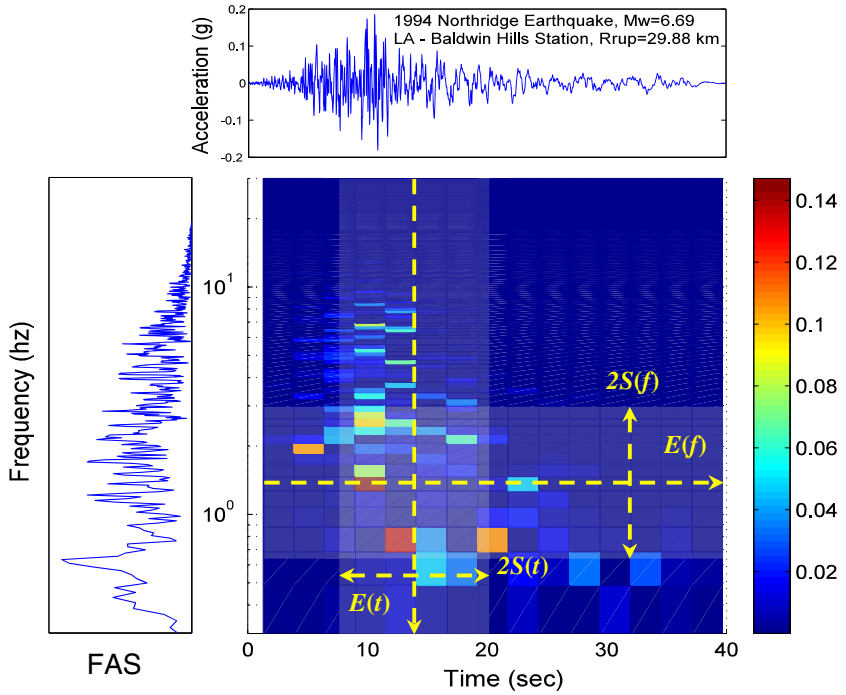

Figure 1. Wavelet packet spectrum, showing the distribution of the squared wavelet packet coefficients of the recorded acceleration time history at Los Angeles Baldwin Hills site in the 1994 Northridge earthquake.

$$c_{j,k}^{i} = \int_{-\infty}^{\infty} x(t)\,\psi_{j,k}^{i}(t)dt \tag{1}$$

where $c_{j,k}^{i}$ is the ith set of wavelet packets at the jth decomposition level (in frequency) and k is the translation parameter (in time), while $\psi_{j,k}^{i}(t)$ is the wavelet packet function. To standardize the process, the decomposition level of the discrete wavelet transform $j=8$ and the time interval of the time series $dt=0.01$ s are used in this study. Therefore, the time interval between the centers of adjacent wavelet packets is $dt_w = 2^j dt = 2.56$ s, and the frequency interval between the centers of adjacent wavelet packets is $\mathrm{d}f_w = 1/(2\mathrm{d}t_w) = 0.1953$ Hz. In total, there are 2^j frequency discretizations and 2^{N-j} time discretizations, if the number of data points in the time series is 2^N. Conversely, ground-motion time histories can be synthesized based on the time-frequency distribution of wavelet packet coefficients via the following reverse transform:

$$x(t) = \sum_{i=1}^{2^j} \sum_{k=1}^{2^{N-j}} c_{j,k}^{i}\,\psi_{j,k}^{i}(t) \tag{2}$$

Table I. Summary of wavelet parameters and parameter groups.

Energy parameter	Time-domain mean and standard deviation	Frequency-domain mean and standard deviation	Time-frequency correlation	Randomness
E_{acc} $E(a)_{major}$	$E(t)_{major}$ $E(t)_{minor}$ $S(t)_{major}$ $S(t)_{minor}$	$E(f)_{major}$ $E(f)_{minor}$ $S(f)_{major}$ $S(f)_{minor}$	$\rho(t,f)_{major}$ $\rho(t,f)_{minor}$	$S(\xi)$

DOI: 10.1002/eqe

Figure 1 illustrates a wavelet packet spectrum, showing the time and frequency-domain distribution of squared wavelet packet coefficients $\left|c_{j,k}^{i}\right|^2$ for an acceleration time history. The wavelet packet spectrum clearly indicates the nonstationarity of wave characteristics in the time and frequency domains. To further quantify the distribution of wavelet coefficients, they are separated into a major group containing 70% of the total energy and a minor group containing the remaining 30%. The stochastic model requires 13 parameters to characterize the wave energy and time-domain/frequency-domain distribution of the wavelet coefficients for the major and minor groups, as shown in Table I. One may refer to Yamamoto and Baker [21] for detailed definition of each parameter.

Of the 13 parameters, two are used to represent the wave energy. For example, the sum of squared wavelet packet coefficients represents the total energy contained in the ground motion, defined as E_{acc}:

$$E_{acc} = \sum_i \sum_k \left|c_{j,k}^{i}\right|^2 = \int_{-\infty}^{\infty} |x(t)|^2 dt \tag{3}$$

Clearly, E_{acc} is related to the well-known Arias intensity [23] by a constant factor. As 70% of energy is contained in the major group, the average of the squared wavelet coefficients for the major group can be defined as $E(a)_{major}$. The wavelet packet coefficients are assumed to follow a lognormal distribution in the time and frequency domains. Two parameters, $E(t)_{major}$ and $E(t)_{minor}$, are defined to describe the mean location of the distribution in the time domain, where the subscripts indicate the major group and the minor group. Similarly, $E(f)_{major}$ and $E(f)_{minor}$ quantify the mean of the major group and that of the minor group in the frequency domain, respectively. $S(t)_{major}$ and $S(t)_{minor}$ are defined to characterize the standard deviation of the time-domain distribution; $S(f)_{major}$ and $S(f)_{minor}$ are defined as the standard deviation of the frequency-domain distribution, for the major and minor groups, respectively. The time-frequency correlations for the major group and minor group distributions are characterized by $\rho(t,f)_{major}$ and $\rho(t,f)_{minor}$. Finally, a randomness parameter $S(\xi)$ is introduced to quantify the magnitude of wavelet coefficients in the minor group.

Given an earthquake event i, the wavelet parameters at site j can be written as follows:

$$Y_{ij} = \overline{Y}_{ij}\left(M_w, R_{hyp}, R_{rup}, V_{s30}\right) + \eta_i + \varepsilon_{ij} \tag{4}$$

where Y_{ij} represents a wavelet parameter as shown in Table I for the earthquake event i at site j, all in natural logarithm scale (e.g., $\ln(E_{acc})$, $\ln(E(t)_{major})$) except for $\rho(t,f)_{major}$ and $\rho(t,f)_{minor}$. The mean $\overline{Y}_{ij}\left(M_w, R_{hyp}, R_{rup}, V_{s30}\right)$ can be predicted using seismological variables, such as the moment magnitude, site-to-source distances, and site conditions, through regression analysis of strong-motion data [21]:

$$\begin{aligned}\overline{Y}_{ij}\left(M_w, R_{hyp}, R_{rup}, V_{s30}\right) &= \alpha + \beta_1 M_w + \beta_2 \ln(M_w) + \beta_3 \exp(M_w) \\ &+ \beta_4\left(R_{hyp} - R_{rup}\right) + \beta_5 \ln\sqrt{R_{rup}^2 + h^2} + \beta_6 \ln(V_{s30})\end{aligned} \tag{5}$$

where M_w is the moment magnitude, R_{rup} is the rupture distance, R_{hyp} is the hypocentral distance, and V_{s30} is the average shear-wave velocity in the top 30 m. The variability of simulated ground motions is introduced in Equation (4), as the predictive equations quantify not only the mean but also the inter-event residuals η_i and intra-event residuals ε_{ij}. The residual terms are assumed to follow a normal distribution with a zero mean, an inter-event standard deviation of τ_i and an intra-event standard deviation of σ_{ij}.

Note that the Yamamoto–Baker model can simulate nonstationary ground motions at a given site location for a given earthquake scenario but it does not provide the spatial correlation of wavelet parameters, and so it cannot be used to generate spatially correlated ground motions at multiple sites over a region. In the following section, we will develop a spatial correlation to extend the Yamamoto–Baker model to regionalized ground-motion simulation.

DOI: 10.1002/eqe

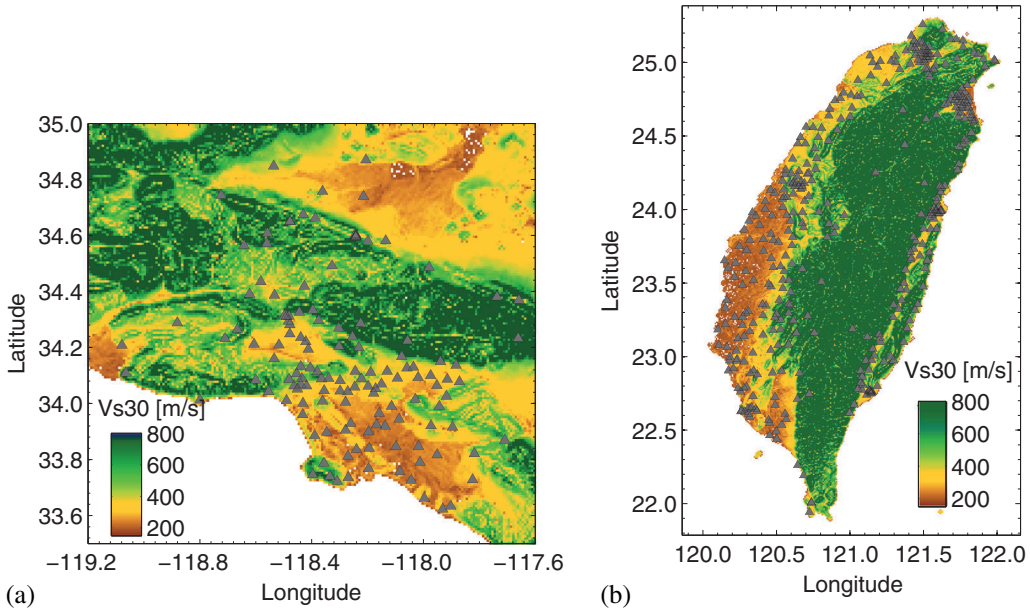

Figure 2. (a) V_{s30} map of Los Angeles area. Triangles indicate a total of 148 earthquake recording stations in this region; (b) V_{s30} map of Taiwan. Triangles indicate a total of 381 earthquake recording stations in this region.

3. SPATIAL CORRELATION OF WAVELET PACKET PARAMETERS

3.1. Semivariogram analysis

In this study, the normalized intra-event residuals of wavelet parameters are used in the spatial analysis. First, the intra-event residuals are corrected to remove their overall biased trend against the rupture distance, if there is any, as the bias could result in inaccurate estimates of the spatial correlation [2, 6, 8, 24]. For better comparisons, the corrected intra-event residuals are further normalized by the sample standard deviation:

$$\varepsilon'_{ij} = \frac{\varepsilon^{corr}_{ij}}{\sigma_{ij}} = \frac{\varepsilon_{ij} - \left[\varphi_1 + \varphi_2 \ln\left(R_{rup}\right)\right]}{\sigma_{ij}} \tag{6}$$

where φ_1 and φ_2 are correction factors obtained from linear regression for each event and σ_{ij} is the sample standard deviation. It should be emphasized that throughout the paper, the spatial correlation and cross-correlation analyses are performed using the normalized residuals of parameters instead of the values of the parameters themselves. For convenience, we may simplify the statement as 'the spatial correlation of parameters' if there is no confusion.

In this section, the spatial correlations of wavelet parameters are investigated using semivariogram analysis. The semivariogram is a widely used geostatistical tool for modeling regionalized variables, such as spatially distributed ground-motion IMs [5–8]. It characterizes the dissimilarity or decorrelation of a set of spatial data, which can be thought of as a stationary regionalized variable $\{Z(\mathbf{u}) : \mathbf{u} \in D\}$, in which the spatial index $\mathbf{u}$ varies continuously over the region D. For a data pair separated by a vector $\mathbf{h}$, the semivariogram is defined as follows [25]:

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathrm{E}\left[\left(Z(\mathbf{u}+\mathbf{h}) - Z(\mathbf{u})\right)^2\right] \tag{7}$$

Previous studies suggested that the vector lag distance $\mathbf{h}$ in Equation (7) can be replaced by a scalar variable $h = \|\mathbf{h}\|$ based on the assumption that the variable $Z(\mathbf{u})$ is spatially isotropic and second-order stationary. However, it is not always the case that two sites are separated by an exact lag distance h.

DOI: 10.1002/eqe

Therefore, this study employs a separation-distance bin $[h-\Delta h, h+\Delta h]$ with a bin size of 4 km to group all data pairs when computing semivariograms.

Recent advances in geostatistical studies of earthquake ground motions have led to the use of various estimators to estimate semivariograms, such as the method-of-moments estimator and the robust estimator [24]. To provide consistent results, the method-of-moments estimator is used to compute semivariograms throughout this study. It is formulated as follows:

$$\widetilde{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(u_i + h) - z(u_i)]^2 \tag{8}$$

where $N(h)$ represents the number of distinct data pairs in the separation-distance bin $[h-\Delta h, h+\Delta h]$.

Four basic continuous models are commonly used to fit the empirical semivariograms, namely, the exponential model, the spherical model, the Gaussian model, and the nugget effect model [25]. Among these models, the exponential model is found to provide the best fit and is thus adopted throughout this study. The exponential model is given by

$$\widetilde{\gamma}(h) = a[1 - \exp(-3h/b)] \tag{9}$$

where a and b are defined as the 'sill' and 'range' of the semivariogram, respectively. The exponential model specifies that 95% of the spatial correlation vanishes beyond the range b.

In the following section, semivariograms of 13 wavelet parameters are developed using ground-motion data from two well-recorded earthquakes, the 1994 Northridge earthquake and the 1999 Chi-Chi earthquake. These two earthquakes represent different regional geological conditions. The Northridge earthquake occurred in a heterogeneous region, whereas the Chi-Chi earthquake took place in a homogeneous region, based on their estimated ranges of V_{s30}. Previous studies reported that the range of V_{s30} for the Northridge earthquake is 0 km, indicating an independent spatial distribution of V_{s30} over the region. On the other hand, the range of V_{s30} for the Chi-Chi earthquake was estimated to be more than 30 km, representing a relatively homogeneous geological condition [6, 8]. Figure 2(a) and 2(b) shows V_{s30} maps of the Los Angeles area and Taiwan, respectively.

3.2. *Influence of regional site conditions*

The earthquake data used in developing empirical spatial correlations are a subset of the NGA database used by Boore and Atkinson [26], and only fault-normal components are adopted in the analyses. These criteria result in 148 ground-motion recordings available for the Northridge earthquake and 381 recordings for the Chi-Chi earthquake, as shown in Figure 2. Normalized residuals of wavelet parameters of each record are computed using the prediction model proposed by Yamamoto and

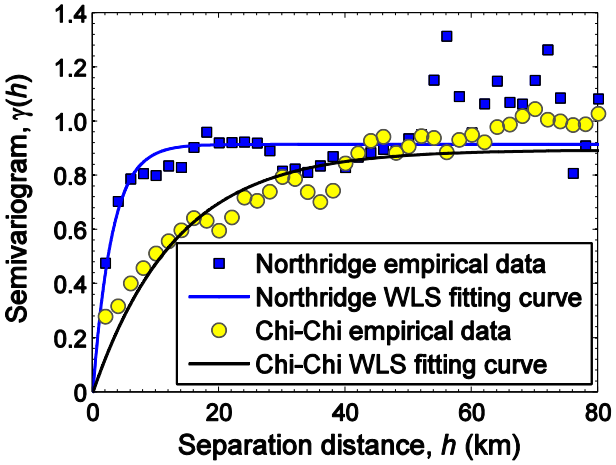

Figure 3. Semivariograms of E_{acc} residuals for the 1994 Northridge earthquake and the 1999 Chi-Chi earthquake using the weighted-least-square method to fit the exponential model.

DOI: 10.1002/eqe

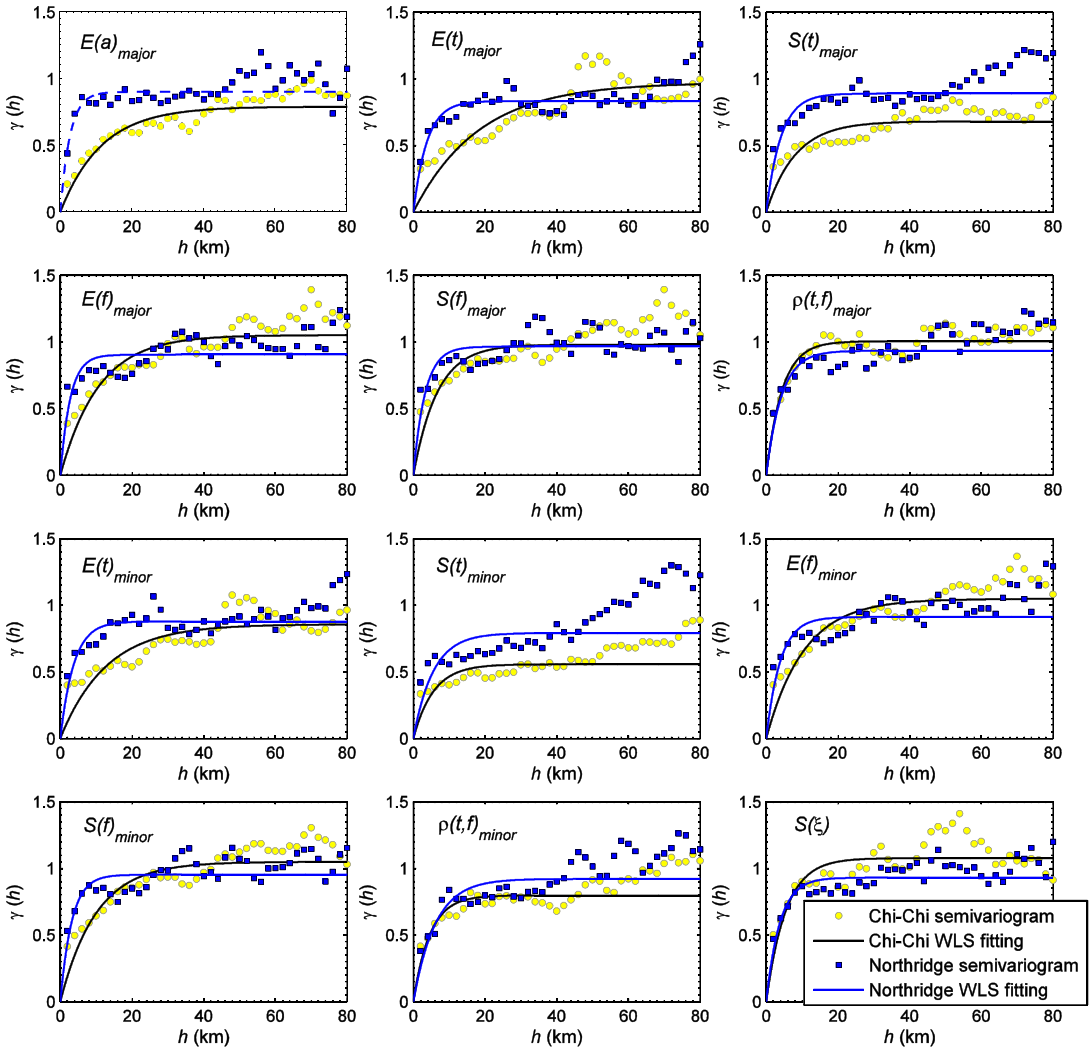

Figure 4. Empirical semivariograms of the normalized residuals of wavelet packet parameters for the Northridge and Chi-Chi earthquakes versus fitted curves using weighted least square.

Baker [21]. To construct semivariograms, the bin size is set to 4 km to compensate for the lack of seismic data within short separation distances.

Figure 3 presents semivariograms of the normalized residuals of E_{acc} for the Northridge and Chi-Chi earthquakes using the weighted-least-square method to fit an exponential model [8]. Ranges of semivariograms for E_{acc} are estimated to be 9.7 and 41.6 km for the Northridge and Chi-Chi earthquakes, respectively. These ranges are consistent with those obtained from Arias intensity (Ia) residuals in the same region [6], because E_{acc} and Ia both represent the integration of acceleration time histories and differ only by a constant multiplication factor.

Figure 4 shows semivariograms for the remaining wavelet parameters, whose range values are found to vary from 7.7 to 19.7 km for the Northridge earthquake and from 13.2 to 58.8 km for the Chi-Chi earthquake. It is observed that the spatial correlations of most wavelet parameters are closely related to regional site conditions. The correlation ranges for the Northridge earthquake are generally smaller than those obtained from the Chi-Chi earthquake. Therefore, it is important to investigate the region-specific spatial cross-correlations of these wavelet parameters when they are used for region-specific applications. A follow-up study has been recently conducted by the authors in [27] using strong-motion data from eight earthquake events in California, Japan, Mexico, and Taiwan.

DOI: 10.1002/eqe

The influence of regional site conditions is further indicated by the ratio of the correlation range from the Chi-Chi earthquake to that from the Northridge earthquake, as shown in Table II, where a larger ratio means that the spatial correlation is more significantly affected by the regional site conditions. Accordingly, the wavelet parameters are divided into four groups. Group I includes two wavelet parameters for the total energy and energy in the major group (E_{acc} and $E(a)_{major}$) as well as two wavelet parameters that define the centroids of the major group coefficients in the time and frequency domains ($E(t)_{major}$ and $E(f)_{major}$). They are found to be most strongly affected by regional site conditions.

In groups II and III, the influence of regional site conditions becomes slightly less pronounced. Group II represents the centroid of the minor group coefficient distribution; group III is the standard deviations of coefficients for both major and minor groups. The ratios of ranges are approximately 3 and 2 for groups II and III, respectively. On the other hand, group IV describes the time-frequency correlation, that is, $\rho(t,f)_{major}$ and $\rho(t,f)_{minor}$, and randomness in the minor group, that is, $S(\xi)$. The spatial correlations of these wavelet parameters are not strongly influenced by regional site conditions.

4. SPATIAL CROSS-CORRELATION OF WAVELET PACKET PARAMETERS

4.1. Linear model of coregionalization for multivariate spatial analysis

Cross-semivariogram analysis is an important extension of the univariate analysis particularly to the multivariate random field [25]. Considering n variables (i.e., the normalized residuals of wavelet parameters in this study) denoted by $Z_1, Z_2,\ldots,Z_n$, the cross-semivariogram $\gamma_{ij}(h)$ describes the average dissimilarity between two variables Z_i and Z_j separated by distance h. The formulation of an empirical cross-semivariogram is as follows:

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

where $N(h)$ represents the number of distinct data pairs in the separation-distance bin, and $z_i(u_\alpha + h)$ and $z_i(u_\alpha)$ represent the αth data pair in this bin for the ith wavelet parameters.

To build a permissible cross-semivariogram model, this study uses the linear model of coregionalization (LMC), which combines a set of basic structures $g_l(h)$ to fit the empirical cross-semivariograms as follows:

Table II. Estimated correlation ranges of wavelet parameters and the ratios for the Chi-Chi and Northridge earthquakes.

Group	Wavelet parameter	Range (km) for Northridge	Range (km) for Chi-Chi	Ratio of ranges	Average ratio in the group
I	E_{acc}	9.7	41.6	4.29	4.39
	$E(a)_{major}$	8.4	34.5	4.11	
	$E(t)_{major}$	12.1	58.8	4.86	
	$E(f)_{major}$	7.7	33.1	4.30	
II	$E(t)_{minor}$	11.3	38.1	3.37	3.07
	$E(f)_{minor}$	11.6	32.0	2.76	
III	$S(t)_{major}$	11.9	25.7	2.20	2.39
	$S(f)_{major}$	8.9	22.9	2.57	
	$S(t)_{minor}$	10.0	18.4	1.84	
	$S(f)_{minor}$	9.3	27.7	2.98	
IV	$\rho(t,f)_{major}$	11.5	13.2	1.15	1.08
	$\rho(t,f)_{minor}$	19.7	14.8	0.75	
	$S(\xi)$	12.3	16.4	1.33	

DOI: 10.1002/eqe

$$\boldsymbol{\Gamma}(h) = \left[\gamma_{ij}(h)\right] = \sum_{l=1}^{L} \mathbf{B}^l g_l(h) \tag{11}$$

where $\mathbf{B}^l = \left[b_{ij}^l\right]$ $(i,j = 1,\ldots,n;\ l = 1,2,\ldots,L)$ is an as-yet-unknown $n \times n$ coregionalization matrix associated with each of the specified basic models $g_l(h)$. Mathematically, $\mathbf{B}^l$ has to be positive semi-definite in order to guarantee a permissible cross-correlation model [25]. Therefore, an iterative procedure should be followed to minimize the weighted sum of differences between the empirical and estimated cross-semivariograms. Meanwhile, the positive semi-definiteness of $\mathbf{B}^l$ can be satisfied [7, 28].

Using the LMC scheme, the spatial cross-correlations of normalized residuals of different wavelet parameters are developed using ground-motion data recorded from the Northridge and Chi-Chi earthquakes. Based on previous univariate analyses, a short-range (5 km) exponential function and a long-range (50 km) one are chosen as basic models in the LMC model. The LMC structure is found to be effective in capturing major features of the empirical data and provides the best overall fit to the empirical semivariograms among all possible combinations we have experimented. Finally, the cross-semivariogram $\boldsymbol{\Gamma}(h)$ and the covariance matrix $\mathbf{C}(h)$ are written in the following form:

$$\boldsymbol{\Gamma}(h) = \mathbf{B}^1\left[1 - \exp\left(\frac{-3h}{5}\right)\right] + \mathbf{B}^2\left[1 - \exp\left(\frac{-3h}{50}\right)\right] \tag{12}$$

$$\mathbf{C}(h) = \mathbf{B}^1 \exp\left(\frac{-3h}{5}\right) + \mathbf{B}^2 \exp\left(\frac{-3h}{50}\right) \tag{13}$$

Note that the above LMC model is used to fit both the Northridge and Chi-Chi data. Further, the elements b_{ij}^l in the coregionalization matrices can be standardized as follows [7, 8]:

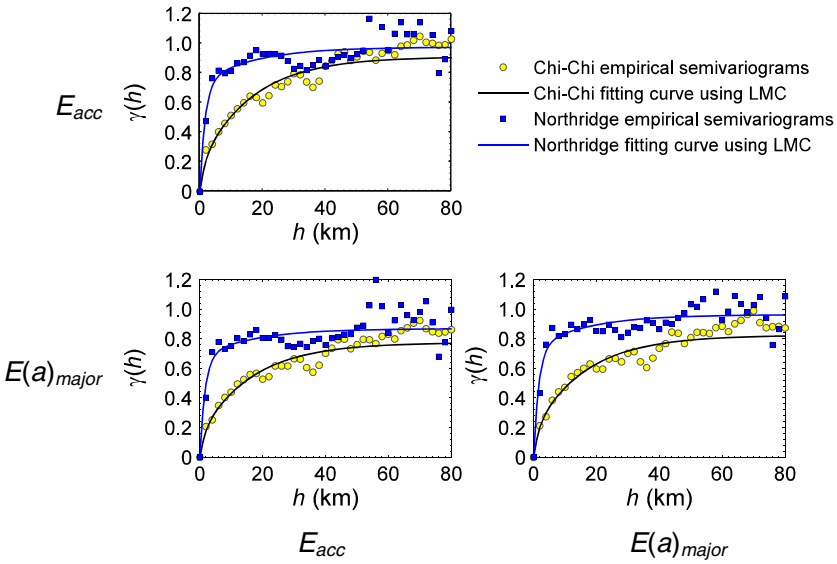

Figure 5. Cross-semivariograms of wavelet parameters $E(a)_{major}$ and E_{acc}, and fitted LMC curves for the Northridge and Chi-Chi earthquakes.

DOI: 10.1002/eqe

$$p_{ij}^{l} = \frac{b_{ij}^{l}}{\left(\sqrt{b_{ii}^{1} + b_{ii}^{2}}\right) \times \left(\sqrt{b_{jj}^{1} + b_{jj}^{2}}\right)} \tag{14}$$

Thus, the correlation matrix $\mathbf{R}(h) = [\rho_{ij}(h)]$ can be obtained using standardized coregionalization matrices $\mathbf{P}^{l} = \left[p_{ij}^{l}\right]$, as shown in the following equation:

$$\mathbf{R}(h) = \mathbf{P}^{1} \exp\left(\frac{-3h}{5}\right) + \mathbf{P}^{2} \exp\left(\frac{-3h}{50}\right) \tag{15}$$

4.2. *Cross-correlations of wavelet parameters*

In this section, spatial cross-correlations of wavelet packet parameters for the Northridge and Chi-Chi earthquakes are investigated. Recall that in Table I, the wavelet parameters are separated into five groups. Although it is not described in detail here, the cross-correlations of wavelet parameter (normalized residuals) at an individual site are usually strong, that is, $\rho_{ij}(h=0) > 0.6$, if they are from the same group; while the cross-correlations of parameters from two different groups are generally much weaker, that is, $\rho_{ij}(0) < 0.4$. For simplicity, the cross-group correlations can be neglected in spatial analysis. This significantly reduces the computational costs in the cokriging analyses presented in a later section.

In what follows, LMC models are determined for all of the wavelet parameters except for the randomness parameter $S(\xi)$. The LMC structure in Equation (12) is adopted for all analyses presented here because it provides a consistent and the best overall fit to the empirical cross-semivariogram data. Figure 5 shows cross-semivariograms and the LMC fitting curves for energy parameters (i.e., $E(a)_{major}$ and E_{acc}) using the Northridge and Chi-Chi earthquake data. The coregionalization matrices for the Northridge earthquake are as follows:

$$\mathbf{P}^{1} = \begin{bmatrix} 0.76 & 0.70 \\ 0.70 & 0.74 \end{bmatrix}, \quad \mathbf{P}^{2} = \begin{bmatrix} 0.24 & 0.20 \\ 0.20 & 0.26 \end{bmatrix} \tag{16}$$

and the coregionalization matrices for the Chi-Chi earthquake are as follows:

$$\mathbf{P}^{1} = \begin{bmatrix} 0.22 & 0.17 \\ 0.17 & 0.19 \end{bmatrix}, \quad \mathbf{P}^{2} = \begin{bmatrix} 0.78 & 0.72 \\ 0.72 & 0.81 \end{bmatrix} \tag{17}$$

Obviously, each entry in the short-range matrix $\mathbf{P}^{1}$ for the Northridge earthquake is larger than its corresponding entry in the short-range matrix $\mathbf{P}^{1}$ for the Chi-Chi earthquake, while the opposite can be observed for the two long-range matrices $\mathbf{P}^{2}$. This is not unexpected, as the Chi-Chi earthquake has a stronger spatial correlation (hence a larger correlation range) than the Northridge earthquake, because of the influence of local site conditions (refer to Section 3.2). In fact, $\mathbf{P}^{1}$ and $\mathbf{P}^{2}$ can be regarded as the weights for the short-range and long-range functions. When the site conditions become more homogeneous as in the case of Chi-Chi, the weights for the long-range function become dominant.

Similarly, cross-semivariogram analysis is conducted for the time-domain parameters (i.e., $E(t)_{minor}$, $S(t)_{minor}$, $E(t)_{major}$, and $S(t)_{major}$), the frequency-domain parameters (i.e., $E(f)_{minor}$, $S(f)_{minor}$, $E(f)_{major}$, and $S(f)_{major}$), and the time-frequency correlation parameters (i.e., $\rho(t,f)_{minor}$ and $\rho(t,f)_{major}$). Tables III–VIII present $\mathbf{P}^{1}$ and $\mathbf{P}^{2}$ matrices using the Northridge and Chi-Chi earthquake data. All $\mathbf{P}^{1}$ and $\mathbf{P}^{2}$ matrices are verified to be positive definite. For both the time-domain and frequency-domain parameters, it can be observed that each component in the short-range coregionalization

DOI: 10.1002/eqe

Table III. $\mathbf{P}^1$ matrix for time-domain mean and standard deviation.

	Northridge				Chi-Chi			
	$E(t)_{minor}$	$S(t)_{minor}$	$E(t)_{major}$	$S(t)_{major}$	$E(t)_{minor}$	$S(t)_{minor}$	$E(t)_{major}$	$S(t)_{major}$
$E(t)_{minor}$	0.67	0.43	0.59	0.47	0.28	0.35	0.19	0.24
$S(t)_{minor}$	0.43	0.47	0.33	0.43	0.35	0.47	0.25	0.37
$E(t)_{major}$	0.59	0.33	0.71	0.51	0.19	0.25	0.15	0.20
$S(t)_{major}$	0.47	0.43	0.51	0.61	0.24	0.37	0.20	0.39

Table IV. $\mathbf{P}^2$ matrix for time-domain mean and standard deviation.

	Northridge				Chi-Chi			
	$E(t)_{minor}$	$S(t)_{minor}$	$E(t)_{major}$	$S(t)_{major}$	$E(t)_{minor}$	$S(t)_{minor}$	$E(t)_{major}$	$S(t)_{major}$
$E(t)_{minor}$	0.33	0.20	0.31	0.29	0.72	0.16	0.76	0.42
$S(t)_{minor}$	0.20	0.53	0.15	0.36	0.16	0.53	0.18	0.43
$E(t)_{major}$	0.31	0.15	0.29	0.25	0.76	0.18	0.85	0.47
$S(t)_{major}$	0.29	0.36	0.25	0.39	0.42	0.43	0.47	0.61

Table V. $\mathbf{P}^1$ matrix for frequency-domain mean and standard deviation.

	Northridge				Chi-Chi			
	$E(f)_{minor}$	$S(f)_{minor}$	$E(f)_{major}$	$S(f)_{major}$	$E(f)_{minor}$	$S(f)_{minor}$	$E(f)_{major}$	$S(f)_{major}$
$E(f)_{minor}$	0.59	0.56	0.58	0.57	0.27	0.26	0.24	0.29
$S(f)_{minor}$	0.56	0.65	0.50	0.58	0.26	0.30	0.22	0.31
$E(f)_{major}$	0.58	0.50	0.68	0.62	0.24	0.22	0.27	0.29
$S(f)_{major}$	0.57	0.58	0.62	0.68	0.29	0.31	0.29	0.42

matrix $\mathbf{P}^1$ for the Northridge earthquake is larger than the corresponding component in $\mathbf{P}^1$ for the Chi-Chi earthquake. On the other hand, each component in the long-range coregionalization matrix $\mathbf{P}^2$ for the Northridge event is substantially smaller than its counterpart for the Chi-Chi event. This is consistent with previous observations regarding the energy parameters, indicating that these spatial cross-correlations are significantly influenced by regional site conditions.

Nonstationarity in time and frequency is another important characteristic of ground-motion time histories, which is quantitatively described in the stochastic model using $\rho(t,f)_{major}$ and $\rho(t,f)_{minor}$. Tables VII and VIII compare the coregionalization matrices of $\rho(t,f)_{major}$ and $\rho(t,f)_{minor}$ obtained from the Northridge and Chi-Chi event. Apparently, their spatial cross-correlations are not strongly influenced by regional site conditions, consistent with the univariate case in Section 3.2.

5. COKRIGING ESTIMATION OF WAVELET PACKET PARAMETERS AT UNMEASURED LOCATIONS

5.1. Ordinary cokriging estimation

Spatial cross-correlations developed in the previous section allow for estimation of wavelet parameters at unmeasured locations, using a spatial interpolation technique called ordinary cokriging. Ordinary cokriging provides the best linear unbiased estimate of variables at an unsampled location, given the covariance structure of variables $\mathbf{C}(h)$ [25, 29]. It is worth mentioning that cokriging estimates are not only derived from the data of the primary variable but also influenced by the secondary variables. In particular, the contribution of a secondary attribute to the primary attribute depends on the correlation coefficient $\rho_{ij}(0)$ between them, the sample density of each variable, as well as the

DOI: 10.1002/eqe

Table VI. $\mathbf{P}^2$ matrix for frequency-domain mean and standard deviation.

	Northridge				Chi-Chi			
	$E(f)_{minor}$	$S(f)_{minor}$	$E(f)_{major}$	$S(f)_{major}$	$E(f)_{minor}$	$S(f)_{minor}$	$E(f)_{major}$	$S(f)_{major}$
$E(f)_{minor}$	0.41	0.32	0.35	0.30	0.73	0.64	0.68	0.64
$S(f)_{minor}$	0.32	0.35	0.24	0.32	0.64	0.70	0.50	0.60
$E(f)_{major}$	0.35	0.24	0.32	0.23	0.68	0.50	0.73	0.58
$S(f)_{major}$	0.30	0.32	0.23	0.32	0.64	0.60	0.58	0.58

Table VII. $\mathbf{P}^1$ matrix for time-frequency correlation.

	Northridge		Chi-Chi	
	$\rho(t,f)_{minor}$	$\rho(t,f)_{major}$	$\rho(t,f)_{minor}$	$\rho(t,f)_{major}$
$\rho(t,f)_{minor}$	0.46	0.39	0.56	0.42
$\rho(t,f)_{major}$	0.39	0.63	0.42	0.62

Table VIII. $\mathbf{P}^2$ matrix for time-frequency correlation.

	Northridge		Chi-Chi	
	$\rho(t,f)_{minor}$	$\rho(t,f)_{major}$	$\rho(t,f)_{minor}$	$\rho(t,f)_{major}$
$\rho(t,f)_{minor}$	0.54	0.43	0.44	0.39
$\rho(t,f)_{major}$	0.43	0.37	0.39	0.38

locations of the primary and secondary data [29]. In this study, all 13 wavelet parameters are equally sampled in the study region. For a particular wavelet parameter, all other parameters in the same group (cf. Table I) are considered as secondary attributes. Taking the energy parameter group as an example, to estimate the total energy E_{acc} at an unmeasured location, both the measured E_{acc} (as a primary variable) and the measured $E(a)_{major}$ (as a secondary variable) are used in the ordinary cokriging approach, which accounts for the spatial cross-correlation between the primary and secondary variables. Similar cokriging analyses are conducted separately for the time-domain parameters, frequency-domain parameters, and time-frequency correlation parameters.

Consider a multivariate random field $\mathbf{Z}$ consisting of n isotropic, second-order stationary regionalized variables, that is, $\mathbf{Z}(\mathbf{u}) = [Z_1(\mathbf{u}), Z_2(\mathbf{u}), \cdots Z_n(\mathbf{u})]^T$, where $\mathbf{u}$ is the position vector, $\mathbf{Z}(\mathbf{u})$ at an unobserved location $\mathbf{u}_0$ can be estimated from a linear combination of the observed values at a total of J sites, located at $\mathbf{u}_\alpha$ ($\alpha = 1,2,\ldots,J$) [25]:

$$\hat{\mathbf{Z}}(\mathbf{u}_0) = \sum_{\alpha=1}^{J} \mathbf{\Lambda}^{(\alpha)} \mathbf{Z}(\mathbf{u}_\alpha) = \left[\mathbf{\Lambda}^{(1)}\mathbf{\Lambda}^{(2)}\cdots\mathbf{\Lambda}^{(J)}\right] \begin{bmatrix} \mathbf{Z}(\mathbf{u}_1) \\ \mathbf{Z}(\mathbf{u}_2) \\ \vdots \\ \mathbf{Z}(\mathbf{u}_J) \end{bmatrix} \tag{18}$$

where $\mathbf{\Lambda}^{(\alpha)} = \left[\lambda_{ij}^{(\alpha)}\right]$ is an as-yet-unknown $n \times n$ cokriging weight matrix associated with the αth observed site. Specifically, the component $\lambda_{ij}^{(\alpha)}$ is the weight assigned to the observed value Z_j ($\mathbf{u}_\alpha$) for the estimation of Z_i at the unobserved site $\mathbf{u}_0$. In indicial notation, the previous estimation can be written as $\hat{Z}_i(\mathbf{u}_0) = \sum_{\alpha=1}^{J}\sum_{j=1}^{n} \lambda_{ij}^{(\alpha)} Z_j(\mathbf{u}_I)$. The cokriging weights can be determined by satisfying the following conditions:

DOI: 10.1002/eqe

(1) The cokriging estimate is 'unbiased', that is, $E\left(\hat{\mathbf{Z}}(\mathbf{u}_0) - \mathbf{Z}(\mathbf{u}_0)\right) = 0$. The condition requires the sum of the cokriging weights $\sum_{\alpha=1}^{J} \mathbf{\Lambda}^{(\alpha)} = \mathbf{I}$, where $\mathbf{I} = [\delta_{ij}]$ is an $n \times n$ identity matrix. In indicial notation, $\sum_{\alpha=1}^{J} \lambda_{ij}^{(\alpha)} = \delta_{ij}$, for $i, j = 1, 2, \ldots, n$, which represents $n \times n$ constraints.

(2) The cokriging estimate is the 'best' estimate in the sense that it minimizes the variance of errors, that is, $\min_{\mathbf{\Lambda}} E\left(\left(\hat{\mathbf{Z}}(\mathbf{u}_0) - \mathbf{Z}(\mathbf{u}_0)\right)^2\right)$.

Using the method of Lagrangian multipliers, the previous two conditions lead to the following augmented system of equations:

$$\mathbf{K}\mathbf{L}(\mathbf{u}_0) = \mathbf{k}(\mathbf{u}_0) \tag{19}$$

where

$$\mathbf{K} = \begin{bmatrix} \boxed{\begin{matrix} \mathbf{C}(0) & \mathbf{C}(h_{12}) & \cdots & \mathbf{C}(h_{1J}) \\ \mathbf{C}(h_{12}) & \mathbf{C}(0) & \cdots & \mathbf{C}(h_{2J}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}(h_{1J}) & \mathbf{C}(h_{2J}) & \cdots & \mathbf{C}(0) \end{matrix}} & \begin{matrix} \mathbf{I} \\ \mathbf{I} \\ \vdots \\ \mathbf{I} \end{matrix} \\ \begin{matrix} \mathbf{I} & \mathbf{I} & \cdots & \mathbf{I} \end{matrix} & \mathbf{0} \end{bmatrix}, \quad \mathbf{L}(\mathbf{u}_0) = \begin{bmatrix} \mathbf{\Lambda}^{(1)} \\ \mathbf{\Lambda}^{(2)} \\ \vdots \\ \mathbf{\Lambda}^{(J)} \\ \mathbf{M} \end{bmatrix}, \quad \mathbf{k}(\mathbf{u}_0) = \begin{bmatrix} \mathbf{C}(h_{10}) \\ \mathbf{C}(h_{20}) \\ \vdots \\ \mathbf{C}(h_{J0}) \\ \mathbf{I} \end{bmatrix} \tag{20}$$

The cokriging weight $\mathbf{\Lambda}^{(\alpha)}$ can thus be solved using the previous equations. Note that $\mathbf{K}$ is an assemblage of $(J+1) \times (J+1)$ submatrices, where $\mathbf{C}(h_{\alpha\beta}) = [C_{ij}(h_{\alpha\beta})]$ is an $n \times n$ covariance submatrix derived in Section 4, $h_{\alpha\beta}$ represents the separation distance between site α and site β ($\alpha, \beta = 1,2,\ldots, J$), $\mathbf{I} = [\delta_{ij}]$ is an $n \times n$ identity matrix, and $\mathbf{M} = [\mu_{ij}]$ is an $n \times n$ matrix for Lagrangian multipliers. The boxed term in $\mathbf{K}$ highlights the total covariance matrix for all J observed sites.

Finally, $\mathbf{Z}(\mathbf{u})$ at an unobserved location $\mathbf{u}_0$ can be estimated using Equation (18). The covariance matrix for the cokriging estimate is

$$\mathbf{C}(\mathbf{u}_0) = \operatorname{cov}\left(Z_i(\mathbf{u}_0), Z_j(\mathbf{u}_0)\right) = \mathbf{C}(0) - \mathbf{k}^T\mathbf{K}^{-1}\mathbf{k} \tag{21}$$

It is worth pointing out that $\mathbf{C}(\mathbf{u}_0)$ is only dependent on the spatial correlation structure but not on the observed values. $\mathbf{C}(\mathbf{u}_0)$ should be understood as a conditional covariance matrix (conditioned on observed values in the neighborhood), which is smaller than the unconditional covariance matrix $\mathbf{C}(0)$ at a location.

5.2. An illustrative example

In this section, an illustrative example is presented to estimate wavelet parameters using strong-motion data from the Chi-Chi earthquake. First, intra-event residuals of wavelet parameters at recorded stations are corrected to remove their biased trend against the rupture distance following Equation (6). The corrected intra-event residuals are then normalized by the sample variance for better comparison of variables. Second, the study region is evenly discretized into a total of 74,000 interpolated locations, with 1 km × 1 km separation in longitude and latitude. Finally, wavelet parameters at interpolated locations are computed using ordinary cokriging estimates.

Figure 6 shows six estimated wavelet parameters throughout the whole region with a spatial resolution of 1 km × 1 km. The earthquake epicenter, surface trace of the Chelungpu fault, and horizontal projection of the fault plane are also plotted. Seismology investigation reveals that the fault is reverse-oblique and the rupture propagated to the north [30, 31]. Figure 6(a) clearly shows the pattern of energy radiation, which is concentrated on the hanging wall side of the rupture plane and attenuates over distances. In the near-fault and forward-directivity region, $E(t)_{major}$ is less than

DOI: 10.1002/eqe

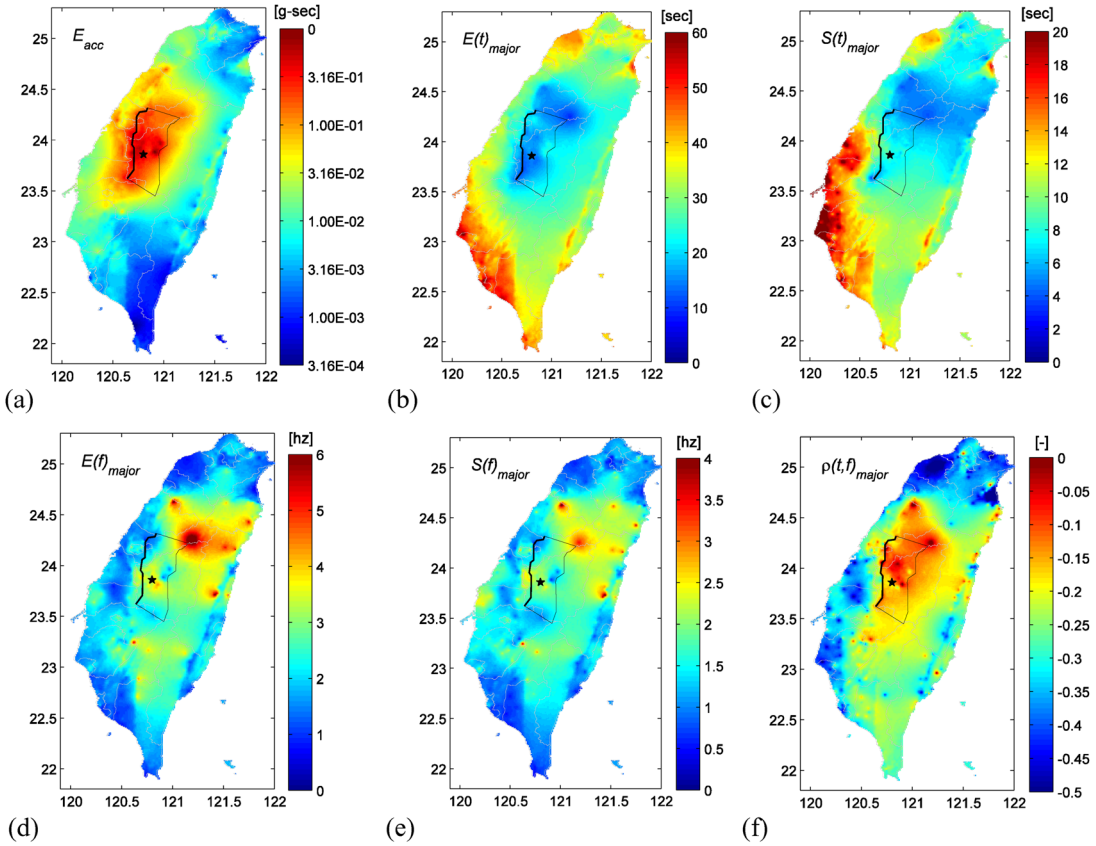

Figure 6. Cokriged wavelet packet parameters for the Chi-Chi earthquake (spatial resolution of 1 km × 1 km).

10 s. In the far field, the seismic waves usually have higher energy (larger E_{acc}) and prolonged duration (larger $E(t)_{major}$ and $S(t)_{major}$) on soil sites than on rock sites. Interestingly, the time-frequency correlation $\rho(t,f)_{major}$ is very weak near the fault and becomes more pronounced at the far field.

Figure 7 compares cokriged and observed wavelet parameters over rupture distance. To facilitate visual inspection, cokriged values are sampled at a separation distance of 5 km × 5 km. Evidently, the cokriged and observed data share a high degree of similarity in distance scaling.

6. STOCHASTIC SIMULATION OF REGIONALIZED GROUND MOTIONS

6.1. Ground-motion simulation using the Northridge earthquake data

Cokriging estimation of wavelet parameters provides a viable approach to simulate ground-motion time histories at any unmeasured location in a region. In this section, regionalized ground motions are simulated stochastically using strong-motion data from the Northridge and Chi-Chi earthquakes. For validation, blind tests are conducted. In a blind test, one recording station is completely removed from the ground-motion database throughout the spatial cross-correlation analysis and cokriging estimation. The simulated ground motion at that station is then compared with the actual recorded ground motion. It should be noted that in the blind test, the accuracy of cokriging estimation can be greatly undermined in some areas if a particular station is removed or its neighboring stations are sparsely or unevenly distributed in that area. Therefore, it is expected that the actual performance should be better than that indicated in the blind tests.

Figure 8(a) is a map of Los Angeles county, showing the epicenter (red star) and projection of the fault plane (dashed line) for the 1994 Northridge earthquake. Six representative recording stations

DOI: 10.1002/eqe

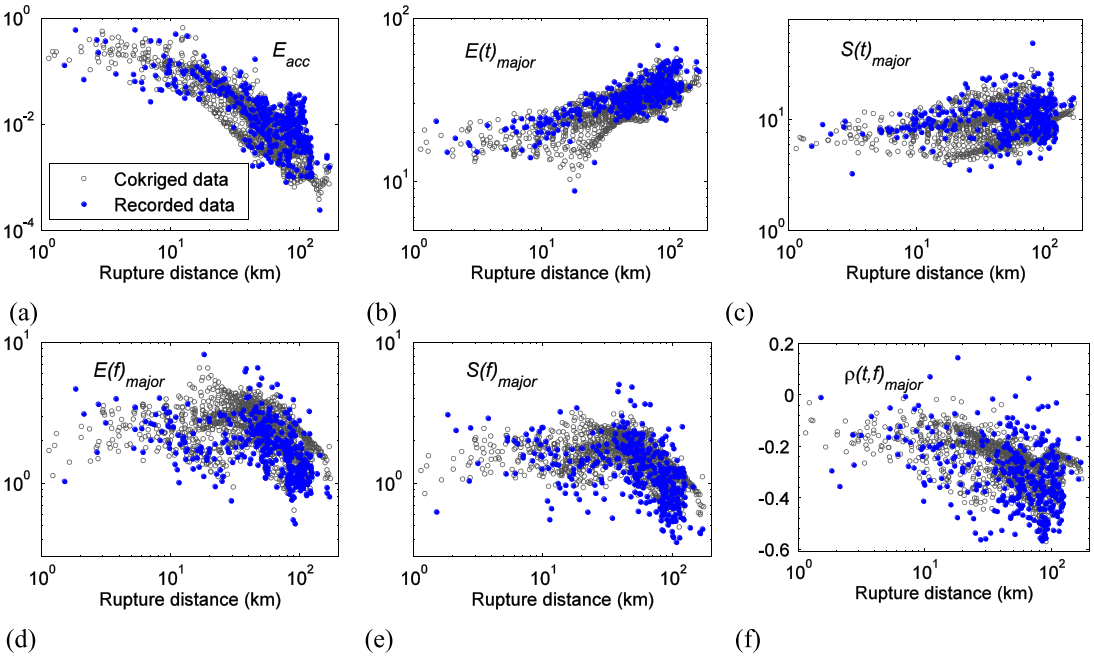

Figure 7. Cokriged (spatial resolution of 5 km × 5 km) versus recorded wavelet packet parameters for the Chi-Chi earthquake over rupture distance.

(solid triangles) are chosen for the blind test. They are the Tarzana-Cedar Hill station, the Sun Valley-Roscoe Blvd station, the UCLA Grounds station, the Hollywood Store station, the Baldwin Hills station, and the Inglewood-Union Oil station. The rupture distance of these stations ranges from 10

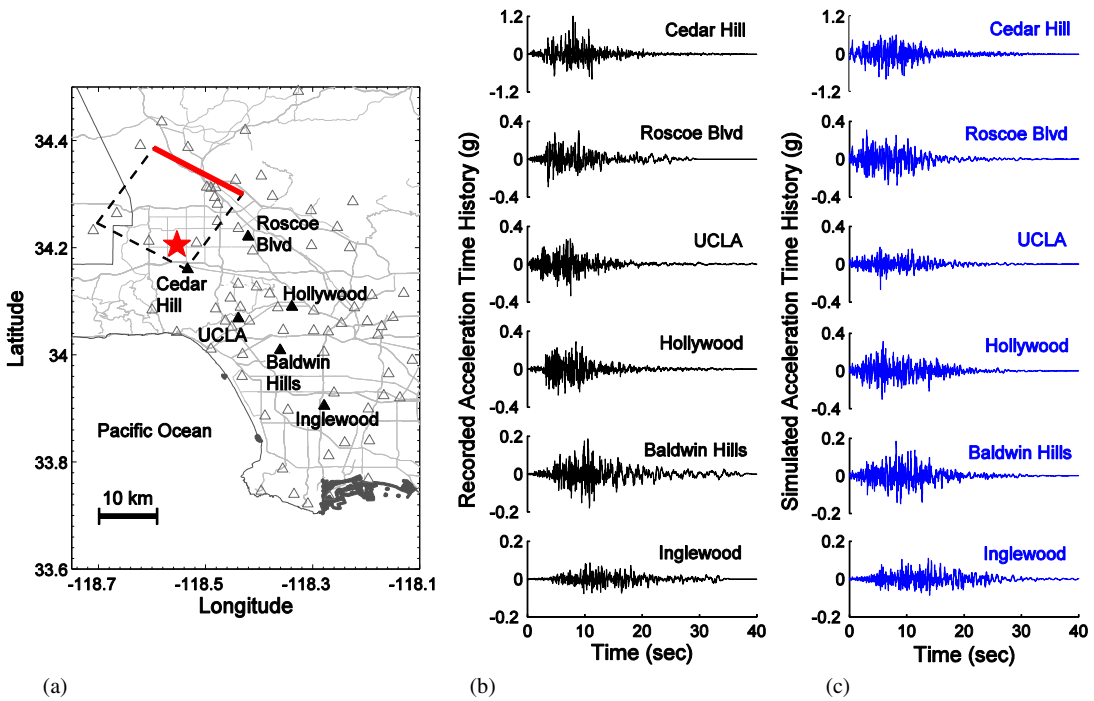

Figure 8. (a) A map of Los Angeles area, showing the six recording stations used in the blind test (solid triangles), all other recording stations in this region (open triangles), the epicenter of the 1994 Northridge earthquake (red star), and surface projection of the Northridge blind thrust fault plane (dashed line). (b) Recorded acceleration time histories at the six stations. (c) Simulated acceleration time histories at the six stations.

DOI: 10.1002/eqe

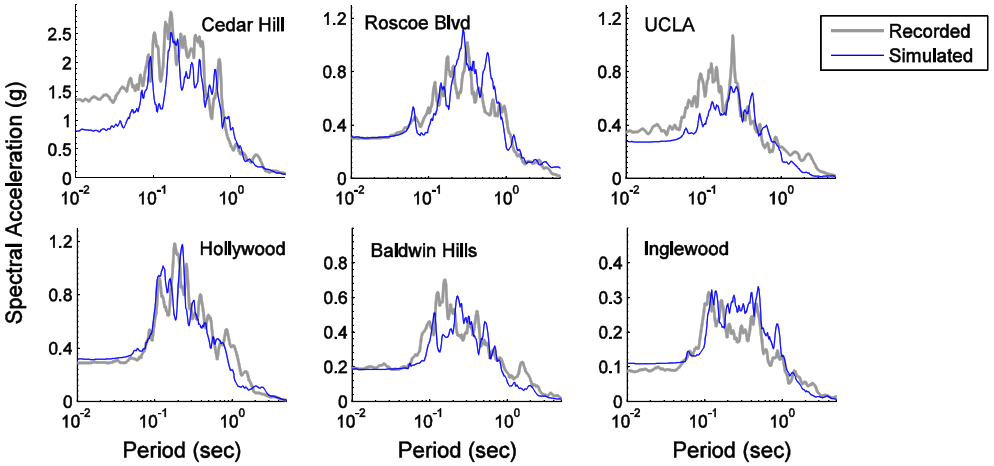

Figure 9. Response spectra for the recorded and simulated ground motions at the six stations used in the blind test during the Northridge earthquake.

to 42 km, while the recorded peak ground acceleration ranges from 0.09 to 1.33 g. Other recording stations used in the analysis are indicated with open triangles.

Figure 8(b) and 8(c) provides a side-by-side comparison of the actual recorded and simulated acceleration time histories in the blind test. Response spectra for the recorded and simulated ground motions are compared in Figure 9. By visual inspection, the simulated and recorded acceleration time histories are found to be similar in terms of both time and frequency characteristics. In general, the simulation captured the overall spectral shapes of the recorded motions reasonably well. To further quantify the similarity, five important ground-motion IMs, namely, peak ground acceleration (PGA), peak ground velocity, significant duration $T_{5\text{-}95}$, Ia, and cumulative absolute velocity ([32]) from the recorded and simulated ground motions are compared in Table IX. On average, the absolute relative errors are in the range of 15–25% for all IMs.

6.2. *Ground-motion simulation using the Chi-Chi earthquake data*

Similar blind tests are conducted using the Chi-Chi earthquake data. Figure 10(a) shows a map of the 1999 Chi-Chi earthquake with the red star indicating the epicenter. The surface traces of the Chelungpu fault and the horizontal projection of the fault plane are also shown. Seven seismic stations are chosen for the blind tests. They are TCU074 and TCU079 on the hanging wall side, TCU063 and TCU113 on

Table IX. Comparison of intensity measures of recorded (the first rows) and simulated (the second rows) ground motions for the Northridge earthquake.

Station	PGA (g)	PGV (cm/s)	$T_{5\text{-}95}$ (s)	Ia (g × s)	CAV (g × s)
Cedar Hill	1.33	65.8	10.44	1.09	2.65
	0.80	67.7	11.94	1.06	2.89
Roscoe Blvd	0.30	25.7	15.57	0.14	1.04
	0.31	42.8	11.41	0.19	1.21
UCLA	0.34	19.7	11.90	0.11	0.94
	0.27	20.3	14.02	0.07	0.75
Hollywood	0.29	23.7	11.14	0.16	1.07
	0.31	21.5	14.32	0.15	1.13
Baldwin Hills	0.19	15.4	17.71	0.06	0.78
	0.18	12.7	13.09	0.05	0.65
Inglewood	0.09	7.71	20.34	0.02	0.43
	0.11	7.76	18.77	0.03	0.57
Avg. relative absolute error	16%	17%	20%	24%	17%

PGA, peak ground acceleration; PGV, peak ground velocity; $T_{5\text{-}95}$, significant duration; Ia, Arias intensity; CAV, cumulative absolute velocity.

DOI: 10.1002/eqe

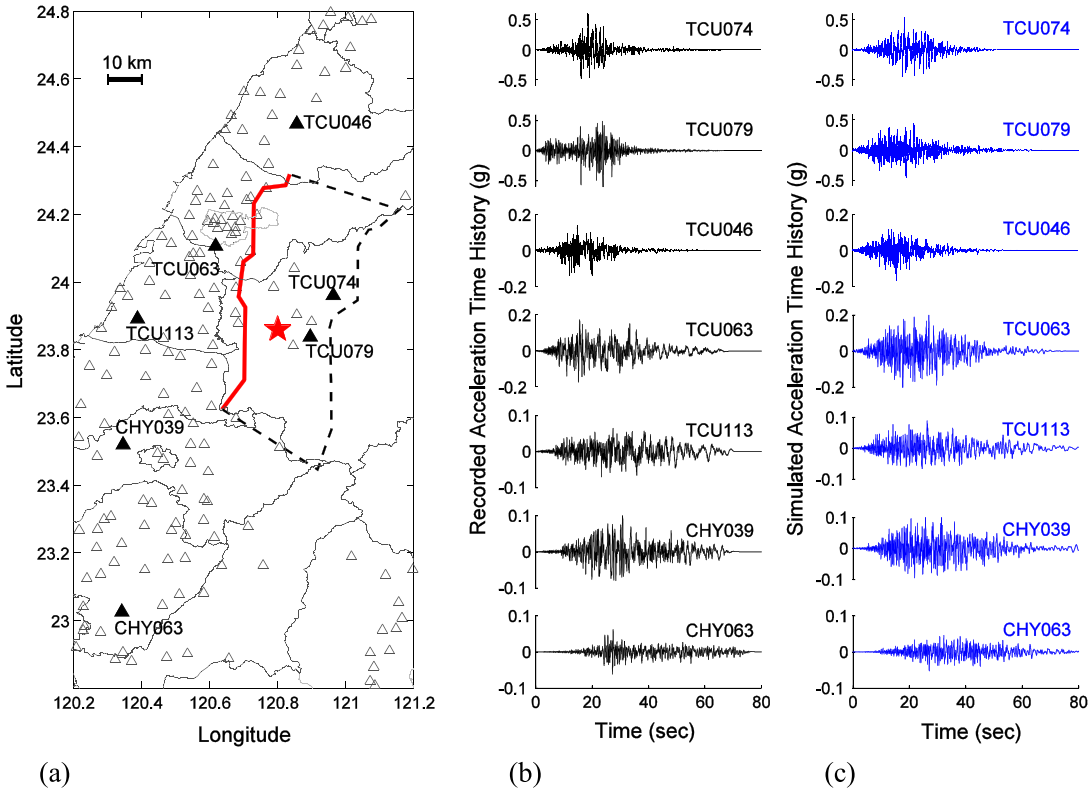

Figure 10. (a) A map of the location of the Chi-Chi earthquake, showing the seven recording stations chosen for the blind tests (solid triangles), other recording stations in this region (open triangles), the epicenter of the earthquake (red star), and surface projection of the Chelungpu fault (dashed line). (b) Recorded acceleration time histories at the seven stations. (c) Simulated acceleration time histories at the seven stations.

the footwall side, TCU046 in the region of forward directivity, and CHY039 and CHY063 in the region of backward directivity. Rupture distances of these stations range from 9.8 to 72 km, while recorded PGAs range from 0.06 to 0.73 g. Figure 10(b) and 10(c) compares simulated time-history traces with recorded ones at these stations. Figure 11 presents comparisons of response spectra. Comparisons of PGA, peak ground velocity, $T_{5\text{-}95}$, Ia, and cumulative absolute velocity are reported in Table X. On average, the absolute relative errors of simulated versus recorded IMs are in the range of 10–30%.

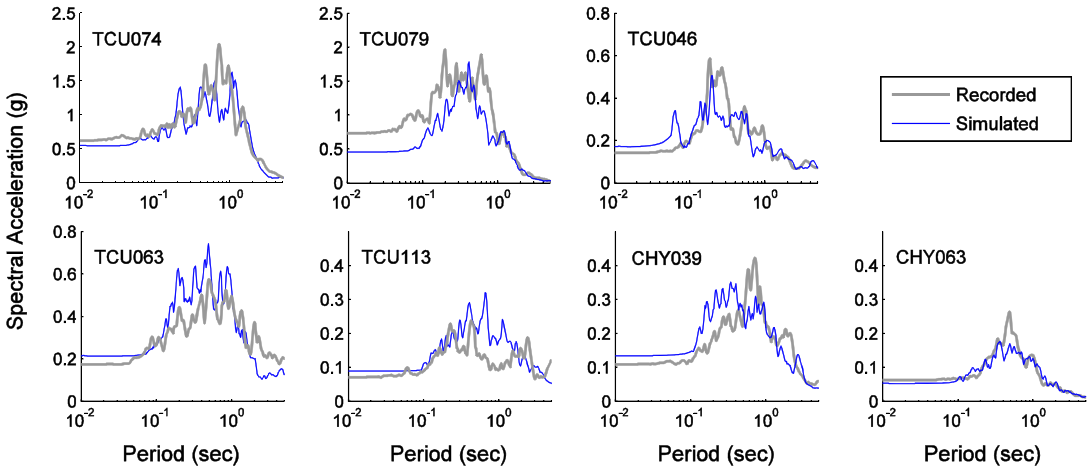

Figure 11. Response spectra for recorded and simulated ground motions at the seven recording stations used in the blind tests during the Chi-Chi earthquake.

DOI: 10.1002/eqe

Table X. Comparison of intensity measures of recorded (the first rows) and simulated (the second rows) ground motions for the Chi-Chi earthquake.

Station	PGA (g)	PGV (cm/s)	$T_{5\text{-}95}$ (s)	Ia (g × s)	CAV (g × s)
TCU074	0.61	76.1	11.8	0.73	2.80
	0.54	82.4	18.8	0.74	2.96
TCU079	0.73	62.6	23.4	0.79	3.15
	0.45	34.3	23.4	0.63	3.14
TCU046	0.14	44.0	18.5	0.05	0.76
	0.17	16.1	24.0	0.05	0.81
TCU063	0.17	61.3	33.6	0.15	1.71
	0.21	45.9	30.5	0.22	2.07
TCU113	0.07	27.8	44.2	0.03	0.92
	0.09	18.2	44.3	0.05	1.08
CHY039	0.11	26.1	37.5	0.05	1.01
	0.13	23.5	37.0	0.07	1.23
CHY063	0.06	8.17	40.6	0.01	0.52
	0.05	5.81	40.0	0.01	0.59
Avg. relative absolute error	22%	30%	14%	24%	12%

PGA, peak ground acceleration; PGV, peak ground velocity; $T_{5\text{-}95}$, significant duration; Ia, Arias intensity; CAV, cumulative absolute velocity.

7. CONCLUSIONS

In this study, ground-motion time histories are characterized using the wavelet packet transform, following a model proposed by Yamamoto and Baker. As ground-motion time histories are transient, complex, nonstationary time sequences, wavelet packet analyses provide useful parameters to quantify the wave energy, the distribution of wavelet amplitudes in the time and frequency domains, their time-frequency correlation, and so on. On the other hand, ground-motion time histories can be stochastically simulated if their wavelet packet parameters are available.

Based on geostatistical analyses, spatial correlations of individual wavelet packet parameters are developed and compared using regionalized ground-motion data from the Northridge and Chi-Chi earthquakes. The spatial cross-correlations of wavelet packet parameters within four wavelet-parameter groups are developed using an LMC. The spatial cross-correlation model enables wavelet packet parameters at an unsampled location to be best estimated using cokriging analysis. Accordingly, ground-motion time histories can be simulated at unmeasured sites.

Case studies and blind tests using data from the Northridge and Chi-Chi earthquakes demonstrate that the simulated ground motions generally agree well with the actual recorded data, by checking the simulated time sequences, response spectra, and various IMs. As mentioned before, the blind tests are performed by completely removing the recording stations from the analysis, which could significantly reduce the accuracy of simulation in some cases. For example, the simulated PGAs on the hanging wall side (e.g., Cedar Hill, TCU079) are largely underestimated in the blind test. This is due to the uneven and sparse distribution of recording stations in that area after removal of blind-test stations. On the other hand, the model performance can be improved with more accurate empirical prediction of wavelet packet parameters in the near-fault region.

Regional site conditions are found to influence the spatial correlations of wavelet parameters to varying degrees. In general, ground motions in the Chi-Chi event have a stronger spatial correlation than those in the Northridge event in terms of the energy parameters, and the time-domain and frequency-domain parameters, while the spatial correlations of the time-frequency correlation parameters and the randomness parameter are not much influenced by regional site conditions. In a follow-up study, we further investigate the region-specific cross-correlation structure of wavelet packet parameters using eight earthquake events from different regions [27]. A simple model for predicting the cross-correlation structure has also been proposed based on site conditions.

In summary, the proposed method is an innovative approach to stochastically simulate regionalized ground motions. The simulated ground motions can enrich the strong-motion database in ground motion selection and modification process [33, 34]. More importantly, the regionalized ground

DOI: 10.1002/eqe

motions can find important applications in time-history analyses of distributed infrastructure systems and regional-scale hazard analysis and loss estimation.

ACKNOWLEDGEMENTS

The study was supported by the Hong Kong Research Grants Council (RCG) through the Collaborative Research Fund grant no. CityU8/CRF/13G and Direct Allocation Grant DAG12EG07-3, FSGRF13EG09 (HKUST/RGC). The authors would like to acknowledge Jack Baker for providing constructive comments and for sharing the source code for ground-motion simulation at http://www.stanford.edu/~bakerjw/gm_simulation.html

REFERENCES

1. Jeon S, O'Rourke TD. Northridge earthquake effects on pipelines and residential buildings. *Bulletin of the Seismological Society of America* 2005; **95**(1):294–318.
2. Sokolov V, Wenzel F. Influence of spatial correlation of strong ground motion on uncertainty in earthquake loss estimation. *Earthquake Engineering & Structural Dynamics* 2011; **40**:993–1009.
3. 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.
4. 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.
5. Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering & Structural Dynamics* 2009; **38**:1687–1708.
6. Du W, Wang G. Intra-event spatial correlations for cumulative absolute velocity, Arias intensity, and spectral accelerations based on regional site conditions. *Bulletin of the Seismological Society of America* 2013; **103**(2A):1117–1129.
7. Loth C, Baker JW. A spatial cross-correlation model of ground motion spectral accelerations at multiple periods. *Earthquake Engineering & Structural Dynamics* 2013; **42**:397–417.
8. Wang G, Du W. Spatial cross-correlation models for vector intensity measures (PGA, Ia, PGV and SAs) considering regional site conditions. *Bulletin of the Seismological Society of America* 2013; **103**(6):3189–3204.
9. DeBock DJ, Garrison JW, Kim KY, Liel AB. Incorporation of spatial correlations between building response parameters in regional seismic loss assessment. *Bulletin of the Seismological Society of America* 2014; **104**(1):214–228.
10. Du W, Wang G. Fully probabilistic seismic displacement analysis of spatially distributed slopes using spatially correlated vector intensity measures. *Earthquake Engineering & Structural Dynamics* 2014; **43**(5):661–679.
11. Lou L, Zerva A. Effects of spatially variable ground motions on the seismic response of a skewed, multi-span bridge. *Soil Dynamics and Earthquake Engineering* 2005; **25**:729–740.
12. Abrahamson NA, Schneider JF, Stepp JC. Empirical spatial coherency functions for application to soil-structure interaction analyses. *Earthquake Spectra* 1991; **7**(1):1–27.
13. Ancheta TD, Stewart JP, Abrahamson NA. Engineering characterization of earthquake ground motion coherency and amplitude variability. Proc. 4th IASPEI / IAEE Int. Sym. on Effects of Surface Geology on Seismic Motion, August 23–26, 2011, University of California, Santa Barbara.
14. Konakli K, Der Kiureghian A, Dreger D. Coherency analysis of accelerograms recorded by the UPSAR array during the 2004 Parkfield earthquake. *Earthquake Engineering & Structural Dynamics* 2014; **43**(5):641–659.
15. Loh CH, Yeh YT. Spatial variation and stochastic modeling of seismic differential ground movement. *Earthquake Engineering & Structural Dynamics* 1988; **16**(4):583–596.
16. Hao H, Olivieria CS, Penzien J. Multiple-station ground motion progressing and simulation based on SMART-1 array data. *Nuclear Engineering and Design* 1989; **111**:293–310.
17. Vanmarcke EH, Heredia-Zavoni E, Fenton GA. Conditional simulation of spatially correlated earthquake ground motion. *Journal of Engineering Mechanics* 1993; **119**(11):2333–2352.
18. Deodatis G. Nonstationary stochastic vector processes: seismic ground motion applications. *Probabilistic Engineering Mechanics* 1996; **11**(3):149–167.
19. Zerva A, Zervas V. Spatial variation of seismic ground motions: an overview. *Applied Mechanics Reviews* 2002; **55**(3):272–297.
20. Konakli K, Der Kiureghian A. Simulation of spatially varying ground motions including incoherence, wave-passage and differential site-response effects. *Earthquake Engineering & Structural Dynamics* 2012; **41**(3):495–513.
21. Yamamoto Y, Baker JW. Stochastic model for earthquake ground motion using wavelet packets. *Bulletin of the Seismological Society of America* 2013; **103**(6):3044–3056.
22. Liu TJ, Hong HP. Simulation of multiple-station ground motions using stochastic point-source method with spatial coherency and correlation characteristics. *Bulletin of the Seismological Society of America* 2013; **103**(3):1912–1921.
23. Arias A. *Measure of Earthquake Intensity, in Seismic Design for Nuclear Power Plants*, Hansen RJ (ed). Massachusetts Institute of Technology Press: Cambridge, Massachusetts, 1970; 438–483.
24. Foulser-Piggott R, Stafford PJ. A predictive model for Arias intensity at multiple sites and consideration of spatial correlations. *Earthquake Engineering & Structural Dynamics* 2012; **41**(3):431–451.
25. Goovaerts P. *Geostatistics for Natural Resources Evaluation*. Oxford University Press: Oxford, New York, 1997.

DOI: 10.1002/eqe

26. 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.
27. Huang D, Wang G. Region-specific spatial cross-correlation model for stochastic simulation of regionalized ground-motion time histories. *Bulletin of the Seismological Society of America* 2014, in press.
28. Goulard M, Voltz M. Linear coregionalization model: tools for estimation and choice of crossvariogram matrix. *Mathematical Geology* 1992; **24**(3):269–286.
29. Issaks EH, Srivastava RM. *Applied Geostatistics*. Oxford University Press: New York, 1989: 561 pp.
30. Shin T-C, Teng T-L. An overview of the 1999 Chi-Chi, Taiwan, Earthquake. *Bulletin of the Seismological Society of America* 2001; **91**(5):895–913.
31. Zeng Y, Chen C-H. Fault rupture process of the 20 September 1999 Chi-Chi, Taiwan, Earthquake. *Bulletin of the Seismological Society of America* 2001; **91**(5):1088–1098.
32. Du W, Wang G. A simple ground-motion prediction model for cumulative absolute velocity and model validation. *Earthquake Engineering & Structural Dynamics* 2013; **42**(8):1189–1202.
33. Wang G. A ground motion selection and modification method capturing response spectrum characteristics and variability of scenario earthquakes. *Soil Dynamics and Earthquake Engineering* 2011; **31**:611–625.
34. Wang G, Youngs R, Power M, Li Z. Design ground motion library (DGML): an interactive tool for selecting earthquake ground motions. *Earthquake Spectra* 2015, **31**(1). DOI: 10.1193/090612EQS283M