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

EARTHQUAKE ENGINEERING & STRUCTURAL DYNAMICS
Earthquake Engng Struct. Dyn. 2014; **43**:661–679
Published online 16 September 2013 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/eqe.2365

Fully probabilistic seismic displacement analysis of spatially distributed slopes using spatially correlated vector intensity measures

Wenqi Du and Gang Wang*,†

[1]*Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong SAR, China*

SUMMARY

Earthquake-induced slope displacement is an important parameter for safety evaluation and earthquake design of slope systems. Traditional probabilistic seismic hazard analysis usually focuses on evaluating slope displacement at a particular location, and it is not suitable for spatially distributed slopes over a large region. This study proposes a computationally efficient framework for fully probabilistic seismic displacement analysis of spatially distributed slope systems using spatially correlated vector intensity measures (IMs). First, a spatial cross-correlation model for three key ground motion IMs, that is, peak ground acceleration (PGA), Arias intensity, and peak ground velocity, is developed using 2686 ground motion recordings from 11 recent earthquakes. To reduce the computational cost, Monte Carlo simulation and data reduction techniques are utilized to generate spatially correlated random fields for the vector IMs. The slope displacement hazards over the region are further quantified using empirical predictive equations. Finally, an illustrative example is presented to highlight the importance of the spatial correlation and the advantage of using spatially correlated vector IMs in seismic hazard analysis of spatially distributed slopes. Copyright © 2013 John Wiley & Sons, Ltd.

Received 2 February 2013; Revised 20 July 2013; Accepted 12 August 2013

KEY WORDS: spatial cross-correlation; vector intensity measures; probabilistic seismic hazard analysis; seismic slope displacement; data reduction technique

1. INTRODUCTION

Estimating seismic displacement of natural slopes and earth structures is important for risk assessment of earthquake-induced landslides and performance-based evaluation of key infrastructures. In practice, the seismic slope displacement at a single site can be evaluated using a pseudoprobabilistic or a fully probabilistic approach based on hazard information derived from probabilistic seismic hazard analysis (PSHA) [1, 2]. However, quantifying the seismic performance of a slope system over a spatially distributed region rather than at just a single site is critical for a variety of applications, including regional risk assessment of landslide and landslide-related damage to lifelines, road systems, and portfolios of infrastructures in this region. Rigorous seismic analysis over a spatially distributed region is less straightforward than that for an individual site. It is not feasible to extend the conventional PSHA-based method directly to the spatially distributed systems because the conventional PSHA does not incorporate the spatial correlation of ground motion intensity measures (IMs) over a region. While it is tempting to obtain the regional seismic loss estimation using the

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

equal hazard maps obtained from the conventional PSHA, this approach implicitly assumes the IMs at different sites are perfectly correlated. Some recent studies show that this approach would result in significantly biased results when used for loss estimation of spatially distributed structures (e.g., [3, 4]).

Two major issues need to be addressed in developing a rational analytical scheme for predicting earthquake-induced slope displacements in a regional scale. First, the spatial cross-correlations between important ground motion IMs related to the estimation of seismic slope displacement have to be systematically studied. One of the major challenges in developing spatial correlation models for IMs is the availability of well-populated strong motion data. A couple of independent earthquakes with densely populated recording stations are often used in previous studies, such as the Northridge (1994) and Chi-Chi (1999) earthquakes. Recently, the spatial correlations of some important IMs, such as the peak ground acceleration (PGA) and Arias intensity (Ia) [5], have been developed by several researchers [6–10]. In addition, the cross-correlation between spectral accelerations at multiple periods is investigated [11]. Currently, there is no spatial cross-correlation study between PGA and other important parameters (Ia and peak ground velocity (PGV)) available in the literature.

The second major issue is related with computational efficiency. For a fully probabilistic analysis of spatially distributed slopes, Monte Carlo-based simulation (MCS) is the only feasible approach to rigorously treat all sources of uncertainties. Because spatial correlations have to be considered between each pair of IMs at each pair of sites, it is not possible to incorporate all these correlations analytically in a vector PSHA-based approach. MCS becomes an inevitable solution, which has attracted considerable attention recently in performing seismic hazard assessment (e.g., [3, 12, 13]). In this study, we propose the following MCS-based framework: First, multiple magnitude-location earthquake scenarios are simulated with frequencies assigned according to recurrence relationships on the basis of the seismicity of the source; second, for each earthquake magnitude-location scenario, vector IMs at all sites (called IM maps) will be randomly generated using ground motion prediction equations (GMPEs) by incorporating inter-event variability and intra-event spatial variability (using the derived spatial cross-correlation matrix) in the process; third, for each vector IM map, multiple sliding displacement maps will be randomly generated using empirical displacement prediction equations and considering their corresponding uncertainties. In the end, the displacement hazard curve for the whole region can be further computed. The MCS approach would result in increasing computational demand downstream in this process. For example, simulating a 30×30 km area (divided into 900 sites separated by 1 km) using 20 earthquake scenarios, 200 sets of vector IM maps for each scenario, and 20 displacement maps for each IM map would generate a total of 80,000 displacement maps. The efforts will be computationally prohibitive when all possible magnitude-distance distributions of the earthquake scenarios need to be considered for practical applications. The inefficiency of the conventional brute-force MCS method calls for a new computationally efficient approach for the proposed stochastic simulation. Several recent studies have been devoted to providing some suitable techniques to reduce the computational cost (e.g., [14, 15]).

This paper aims at developing a framework for a fully probabilistic analysis of spatially distributed slopes. The spatial correlation between several key IMs most relevant to the prediction of seismic slope displacement, that is, PGA, Ia, and PGV, is studied using 11 recent well-recorded earthquakes. A MCS-based computational framework is also developed to rigorously account for all sources of uncertainties. Three data reduction techniques are explored to reduce the computational cost. Following this framework, an illustrative example is also provided in the end.

2. SCALAR AND VECTOR INTENSITY MEASURES FOR PREDICTING SEISMIC SLOPE DISPLACEMENTS

A suitable prediction model is necessary to predict seismic slope displacement on the basis of ground motion IMs. Since Newmark's pioneering work on the rigid sliding block method [16], the Newmark sliding displacement has important applications in evaluating natural slopes or earthquake-induced landslides [17]. The Newmark displacement analysis assumes that the slope behaves as a rigid-plastic material, and the slope displacement is calculated by double integrating the part of the input

DOI: 10.1002/eqe

acceleration that exceeds a critical value (a_c, which can be determined by the properties of slopes). It provides a simple index of seismic slope performance. In the past, a large number of empirical prediction equations have been proposed to predict the Newmark displacement based on a single (scalar) or multiple IMs (a vector IM). The PGA, Ia, and PGV were often used as predictors for the Newmark displacement, for example, [17–20]. As earthquake records are complex, transient time series, multiple ground motion IMs are necessary to represent different aspects of ground motion characteristics. The predictive models using a vector IM usually result in reduced aleatory variability (i.e., improved efficiency) and unbiased results for a wide range of earthquake scenarios.

In this study, four recently developed Newmark displacement prediction equations are chosen on the basis of a scalar IM (termed as PGA model), two-IMs (termed as (PGA, Ia) model and (PGA, PGV) model), as well as three-IMs (termed as (PGA, Ia, PGV) model) as follows [19]:

(1) PGA model:

$$\ln D = 5.52 - 4.43\left(\frac{a_c}{PGA}\right) - 20.39\left(\frac{a_c}{PGA}\right)^2 + 42.61\left(\frac{a_c}{PGA}\right)^3 - 28.74\left(\frac{a_c}{PGA}\right)^4 + 0.72\ln(PGA) \tag{1a}$$

$$\sigma_{\ln D} = 1.13 \tag{1b}$$

(2) (PGA, Ia) model:

$$\ln D = 2.39 - 5.24\left(\frac{a_c}{PGA}\right) - 18.78\left(\frac{a_c}{PGA}\right)^2 + 42.01\left(\frac{a_c}{PGA}\right)^3 - 29.15\left(\frac{a_c}{PGA}\right)^4 - 1.56\ln(PGA) + 1.38\ln(\mathrm{Ia}) \tag{2a}$$

$$\sigma_{\ln D} = 0.46 + 0.56(a_c/PGA) \tag{2b}$$

(3) (PGA, PGV) model:

$$\ln D = -1.56 - 4.58\left(\frac{a_c}{PGA}\right) - 20.84\left(\frac{a_c}{PGA}\right)^2 + 44.75\left(\frac{a_c}{PGA}\right)^3 - 30.5\left(\frac{a_c}{PGA}\right)^4 - 0.64\ln(PGA) + 1.55\ln(PGV) \tag{3a}$$

$$\sigma_{\ln D} = 0.41 + 0.52(a_c/PGA) \tag{3b}$$

(4) (PGA, Ia, PGV) model:

$$\ln D = -0.74 - 4.93\left(\frac{a_c}{PGA}\right) - 19.91\left(\frac{a_c}{PGA}\right)^2 + 43.75\left(\frac{a_c}{PGA}\right)^3 - 30.12\left(\frac{a_c}{PGA}\right)^4 - 1.3\ln(PGA) + 1.04\ln(PGV) + 0.67\ln(\mathrm{Ia}) \tag{4a}$$

$$\sigma_{\ln D} = 0.2 + 0.79(a_c/PGA) \tag{4b}$$

where D is the predicted sliding displacement in cm, a_c and PGA are in the unit of g, PGV is the peak ground velocity in cm/s, and Ia is the Arias intensity in the unit of m/s.

In general, the prediction models using vector IMs can significantly reduce the aleatory variability, and the models show less bias with respect to rupture distance and moment magnitude compared with the scalar (PGA) model [19]. One of the challenges associated with using the vector IM models is to develop spatial cross-correlations between these IMs, which will be studied in the following sections.

DOI: 10.1002/eqe

3. SPATIAL CROSS-CORRELATION FOR VECTOR IM [PGA, IA, PGV]

3.1. Ground motion database for the spatial correlation

A total of 2686 ground motion recordings from 11 earthquakes are compiled to develop the spatial cross-correlation models for PGA, Ia, and PGV in this study. These earthquakes occurred in California (1994 Northridge, 2004 Parkfield, 2005 Anza, 2007 Alum Rock, and 2008 Chino Hills earthquakes), in Mexico (2010 EI Mayor Cucapah earthquake), in Japan (2000 Tottori, 2004 Niigata, 2007 Chuetsu, and 2008 Iwate earthquakes), and in Taiwan region (1999 Chi-Chi earthquake). The recorded time histories for these events are obtained from CESMD, CESMOS for US earthquakes and K-NET, Kik-Net for Japan earthquakes. The seismic information and site conditions are obtained from the Pacific Earthquake Engineering Research Center's Next Generation Attenuation strong motion database (http://peer.berkeley.edu/products/strong_ground_motion_db.html) and the Table S1 database provided by [21]. This detailed information of the earthquakes is summarized in Table I. The moment magnitude and rupture distance distribution of the data in the database are illustrated in Figure 1.

3.2. Geostatistical analysis of intra-event residuals

Based on ground motion prediction equations, the observed logarithmic IM, denoted as $\ln Y_{ij}$, at site j for an earthquake event i, can be written as follows:

$$\ln Y_{ij} = \overline{\ln Y_{ij}(M,R,\theta)} + \eta_i + \varepsilon_{ij} \tag{5}$$

where $\overline{\ln Y_{ij}(M,R,\theta)}$ is the predicted mean IM based on magnitude (M), rupture distance (R), and other variables (θ); η_i is the inter-event residuals with zero means and standard deviations of τ_i; and ε_{ij} denotes the intra-event residuals with zero means and standard deviations of σ_{ij}. Both η_i and ε_{ij} are assumed to be normally distributed independent random variables [22, 23]. The standard deviation of the total residuals is given by $\sigma_T = \sqrt{\sigma_{ij}{}^2 + \tau_i{}^2}$.

In this study, GMPEs developed by Campbell and Bozorgnia [24, 25] are used for PGA, PGV, and Ia, respectively. These GMPEs are developed from the PEER-NGA database using the same functional form. The intra-event residuals of recorded ground motions from each event are computed and corrected to remove their bias against the rupture distance and shear wave velocity in the top 30 m (V_{s30}) values for each event as follows:

$$\varepsilon_{ij}^{corr} = \varepsilon_{ij} - (\varphi_1 + \varphi_2 \ln(R) + \varphi_3 \ln(V_{s30})) \tag{6}$$

where φ_1, φ_2, and φ_3 are the coefficients obtained by linear regression. The correction is necessary to avoid artificial spatial correlations because of systematical predictive biases [9, 10].

Table I. Earthquake events used in this study.

Earthquake name	Date (mm/dd/yyyy)	Moment magnitude	Hypocenter latitude (°)	Hypocenter longitude (°)	Fault mechanism	Number of recordings
Northridge	01/17/1994	6.69	34.206	−118.554	Reverse	152
Chi-Chi	09/20/1999	7.62	23.860	120.800	Reverse-oblique	401
Tottori	10/06/2000	6.61	35.275	133.350	Strike-slip	235
Parkfield	09/28/2004	6	35.817	−120.365	Strike-slip	90
Niigata	10/23/2004	6.63	37.307	138.839	Reverse	365
Anza	06/12/2005	5.2	35.533	−116.578	Reverse-oblique	111
Chuetsu	07/16/2007	6.8	37.538	138.617	Reverse	401
Alum Rock	10/30/2007	5.4	37.432	−121.776	Strike-slip	161
Iwate	06/13/2008	6.9	39.027	140.878	Reverse	279
Chino Hills	07/29/2008	5.4	33.955	−117.765	Reverse-oblique	337
EI Mayor Cucapah	04/04/2010	7.2	32.128	−115.303	Strike-slip	154

Note: only recorded data within the rupture distance of 200 km is included.

DOI: 10.1002/eqe

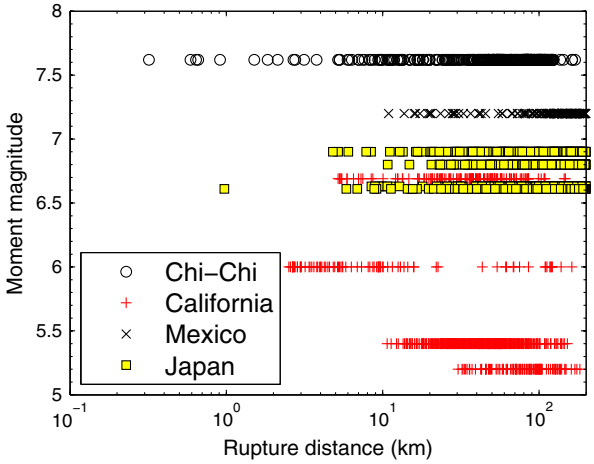


Figure 1. Magnitude and rupture distance distribution of records in the database.

For a vector **IM** = [PGA, Ia, PGV], the intra-event residuals can be assumed to follow a multivariate normal distribution, for example, [26]. Under this assumption, the intra-event residuals $\boldsymbol{\varepsilon}_{ij} = \left(\varepsilon_{ij}^1, \ldots, \varepsilon_{ij}^n\right)$ for n IMs for earthquake event i at site j can be fully obtained by their mean (zero vector in this case) and covariance matrix. For event i, the covariance matrix can be shown as

$$\boldsymbol{\Sigma}(\text{event } i) = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \cdots & \boldsymbol{\Sigma}_{1,J} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{J,1} & \cdots & \boldsymbol{\Sigma}_{J,J} \end{bmatrix} \tag{7}$$

where J is the total number of sites. The submatrix $\boldsymbol{\Sigma}_{k,l}$ ($k,l = 1,\ldots J$, representing different site) can be written as

$$\boldsymbol{\Sigma}_{k,l} = \begin{bmatrix} \operatorname{cov}\left(\varepsilon_{ik}^1, \varepsilon_{il}^1\right) & \cdots & \operatorname{cov}\left(\varepsilon_{ik}^1, \varepsilon_{il}^n\right) \\ \vdots & \ddots & \vdots \\ \operatorname{cov}\left(\varepsilon_{ik}^n, \varepsilon_{il}^1\right) & \cdots & \operatorname{cov}\left(\varepsilon_{ik}^n, \varepsilon_{il}^n\right) \end{bmatrix} \tag{8}$$

where n is the number of IMs considered and $\operatorname{cov}\left(\varepsilon_{ik}^p, \varepsilon_{il}^q\right)$ denotes the covariance between ε_{ik}^p (the residual of the p-th *IM* at site k) and ε_{il}^q (the residual of the q-th *IM* at site l) for event i. The dimension of the total covariance matrix Σ(event i) is $[J \times n, J \times n]$.

Cross-semivariogram is a widely used statistical tool to estimate the spatial correlation of random variables. The cross-semivariogram can be defined as measuring the average dissimilarity between two second-order stationary random variables Z_i and Z_j separated by a distance vector $\boldsymbol{h}$ as follows [27]:

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

where $Z_i(\boldsymbol{u})$ and $Z_i(\boldsymbol{u}+\boldsymbol{h})$ are variable Z_i evaluated at position u and at a position separated by a distance vector $\boldsymbol{h}$, respectively. In this study, Z_i and Z_j refer to the intra-event residuals of the i-th and the j-th component of the vector **IM**. Under the assumptions that the spatial field is isotropic and second-order stationary, a scalar variable $h = \|\boldsymbol{h}\|$ can be used in the formulation. As an estimate of the theoretical Equation (9), an empirical cross-semivariogram can be calculated from a sample

DOI: 10.1002/eqe

dataset. For an earthquake event, all empirical data pairs are gathered if the separation distance of the data pair falls into a distance bin h. A classical estimator can be defined as

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

where $\widetilde{\gamma}_{ij}(h)$ represents empirical cross-semivariograms, $N(h)$ is the number of data pairs within this distance bin h, and $z_i(u_\alpha + h)$ and $z_i(u_\alpha)$ represent the α-th data pair in this bin for i-th component of the vector **IM**. An exponential functional form can be used to fit the earlier empirical cross-semivariogram data:

$$\widetilde{\gamma}_{ij}(h) = a_{ij}\left[1 - \exp\left(-3h/r_{ij}\right)\right] \tag{11}$$

where a_{ij} is the *sill* of the cross-semivariogram and r_{ij} is the *range* of the cross-semivariogram, defined as the separation distance h at which $\widetilde{\gamma}_{ij}(h)$ equals 95% of the sill. On the other hand, the covariance function $C_{ij}(h)$ is defined as

$$C_{ij}(h) = \operatorname{cov}\left(Z_i(u), Z_j(u+h)\right) = E\left[\left(Z_i(u) - m_i\right)\left(Z_j(u+h) - m_j\right)\right] \tag{12}$$

where $m_i = E[Z_i(u)] = 0$ and $i,j = 1,\ldots n$, represent different IM residuals. At an individual site ($h \to 0$), $C_{ij}(0) = \operatorname{cov}(Z_i(u), Z_j(u))$ is the covariance between the i-th and the j-th component of the vector **IM**. It is straightforward to show that the following relationship holds between the covariance function and the cross-semivariogram function (c.f., pp. 72–74 in [27]):

$$C_{ij}(h) = \lim_{h\to\infty}\gamma_{ij}(h) - \gamma_{ij}(h) = C_{ij}(0) - \gamma_{ij}(h) \tag{13}$$

The unit-free correlation coefficient between two variables Z_i and Z_j is

$$\rho_{ij}(h) = \frac{C_{ij}(h)}{\sqrt{C_{ii}(0)\cdot C_{jj}(0)}} \tag{14}$$

Accordingly, the covariance matrix $\boldsymbol{C}(h)$ for the n-component vector **IM** is defined as follows:

$$\mathbf{C}(h) = \left[C_{ij}(h)\right] = \begin{bmatrix} C_{11}(h) & \ldots & C_{1n}(h) \\ \vdots & \ddots & \vdots \\ C_{n1}(h) & \cdots & C_{nn}(h) \end{bmatrix} \tag{15}$$

Therefore, the total covariance matrix Σ(event i) in Equation (7) can be implemented by submatrix $\boldsymbol{C}(h)$ in Equation (15) as follows:

$$\boldsymbol{\Sigma}(\text{event } i) = \begin{bmatrix} \mathbf{C}(0) & \ldots & \mathbf{C}(h_{1J}) \\ \vdots & \ddots & \vdots \\ \mathbf{C}(h_{J1}) & \cdots & \mathbf{C}(0) \end{bmatrix} \tag{16}$$

where h_{ij} represents the specific separation distance between site i and site j among a total of J sites (the separation distance is always zero for diagonal elements).

In summary, the total covariance matrix for the n-component vector **IM** at J sites can be obtained once the correlation range in Equation (11) is obtained by regression. The procedure

DOI: 10.1002/eqe

(called 'direct fit method') is straightforward and efficient. However, the total covariance matrix obtained by the direct fit method cannot guarantee the positive definiteness, making it difficult to generate spatially correlated random field in application. Hence, a statistical approach termed as the linear model of coregionalization (LMC, [28]) will be subsequently adopted in this study to overcome the earlier limitation, such that the resulted total covariance matrix will guarantee to be positive definite.

3.3. Coregionalization matrix for vector IM [PGA, Ia, PGV]

The LMC can be used to decompose the cross-semivariograms $\gamma_{ij}(h)$ as a linear combination of independent random functions as follows [28]:

$$\gamma_{ij}(h) = \sum_{l=1}^{L} b_{ij}^{l} g_l(h) \qquad \forall i,j \tag{17}$$

where $g_l(h)$ is a given permissible basic function, b_{ij}^{l} represents the sill, and L is the number of basic functions. Accordingly, the cross-semivariogram matrix $\mathbf{\Gamma}(h)$ can be decomposed as ([28], pp. 171–173):

$$\mathbf{\Gamma}(h) = \sum_{l=1}^{L} \mathbf{B}^{l} g_l(h) \tag{18}$$

where $\mathbf{B}^{l} = \left[b_{ij}^{l}\right]$ is called the coregionalization matrix. It is to be noted that as long as the positive definiteness of matrix $\mathbf{B}^{l}$ is satisfied, the total covariance matrix $\mathbf{\Sigma}$(event i) in Equation (16) is guaranteed to be positive definite regardless of the number of sites located in this region. This condition can be easily satisfied because $\mathbf{B}^{l}$ is just a $n \times n$ matrix (n is the number of IMs considered).

In this study, a short range (10 km) and a long range (60 km) exponential functions are used as the basic functions of $\mathbf{\Gamma}(h)$, h is in the unit of km:

$$\mathbf{\Gamma}(h) = \mathbf{B}^{1}\left(1 - \exp\left(\frac{-3h}{10}\right)\right) + \mathbf{B}^{2}\left(1 - \exp\left(\frac{-3h}{60}\right)\right) \tag{19}$$

Accordingly, the spatial correlation coefficient matrix can be obtained as

$$\mathbf{R}(h) = \mathbf{P}^{1}\left(\exp\left(\frac{-3h}{10}\right)\right) + \mathbf{P}^{2}\left(\exp\left(\frac{-3h}{60}\right)\right) \tag{20}$$

where $\mathbf{R}(h)$ is the correlation matrix and $\mathbf{P}^{1}$ and $\mathbf{P}^{2}$ are standardized versions of $\mathbf{B}^{1}$ and $\mathbf{B}^{2}$, defined as

$$p_{ij}^{l} = \frac{b_{ij}^{l}}{\left(\sqrt{b_{ii}^{1} + b_{ii}^{2}}\right) \times \left(\sqrt{b_{jj}^{1} + b_{jj}^{2}}\right)} \qquad l = 1,2 \tag{21}$$

An iterative algorithm can be used to obtain a positive definite matrix $\mathbf{B}^{l}$ efficiently for the LMC model. This method minimizes the weighted sum of squares of differences between the empirical cross-semivariograms and the one estimated by the LMC model:

$$WSS = \sum_{k=1}^{K} \sum_{i=1}^{n} \sum_{j=1}^{n} w(h_k) \; \frac{\left[\widetilde{\gamma}_{ij}(h_k) - \widehat{\gamma}_{ij}(h_k)\right]^2}{\hat{\sigma}_i \cdot \hat{\sigma}_j} \tag{22}$$

where $\widehat{\gamma}_{ij}(h_k)$ is the estimated cross-semivariogram by LMC, $\widetilde{\gamma}_{ij}(h_k)$ is the empirical cross-semivariogram data, $\hat{\sigma}_i$ is the standard deviation of Z_i ($i = 1,..,n$), and $w(h_k)$ is the weight for separation distances h_k

DOI: 10.1002/eqe

($k = 1,..,K$). The algorithm checks and ensures the positive definiteness of $\mathbf{B}^l$ matrix during iterations. One can refer to [29] for the details.

Following the earlier procedure, the cross-semivariograms with respect to separation distances for PGA, Ia, and PGV, along with the fitted curves by the LMC method are illustrated in Figure 2 for the Chi-Chi and Northridge earthquakes, respectively. All fitting LMC curves approximate the empirical data reasonably well for each case. Similar results can be observed for other earthquakes considered in this study. It is worth pointing out significant difference in the cross-semivariograms that can be observed for the Chi-Chi and the Northridge data. In a most recent study [30], the spatial correlations of [PGA, Ia, PGV] are found to be closely related to the regional site conditions. In general, an IM recorded from a relatively homogenous region (e.g., the Chi-Chi event) tends to have a stronger spatial correlation than that from a heterogeneous region (e.g., the Northridge event). One can refer to [30] for predictive equations to estimate the coregionalization matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ for the vector **IM** [PGA, Ia, PGV] if the regional specific application is desired. For most practical purposes, the following averaged coregionalization matrices $\mathbf{P}^1$ and $\mathbf{P}^2$ can be used, which are obtained by averaging the coregionalization matrices from all 11 earthquake events considered in this study:

$$\mathbf{P}^1 = \begin{bmatrix} 0.61 & 0.57 & 0.39 \\ 0.57 & 0.67 & 0.46 \\ 0.39 & 0.46 & 0.50 \end{bmatrix}, \qquad \mathbf{P}^2 = \begin{bmatrix} 0.39 & 0.34 & 0.22 \\ 0.34 & 0.33 & 0.22 \\ 0.22 & 0.22 & 0.50 \end{bmatrix} \tag{23}$$

Both $\mathbf{P}^1$ and $\mathbf{P}^2$ are positive definite, which as we introduced earlier, is vitally important for stochastic simulation of spatially correlated fields. Besides, any submatrices of $\mathbf{P}^1$ and $\mathbf{P}^2$ remain positive definite. Considering the spatial correlation between PGA and Ia (i.e., $p_{12}^1 = 0.57$ and

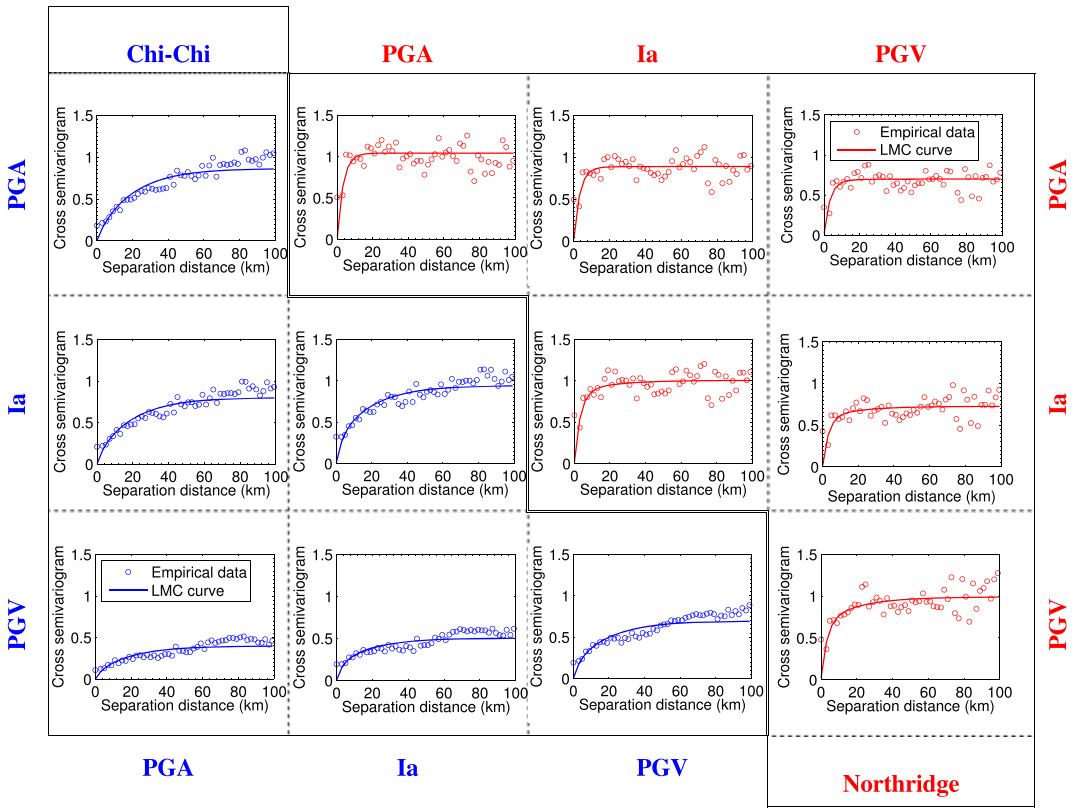


Figure 2. Cross-semivariograms and fitted least median square curves for the Chi-Chi earthquake (lower triangle, blue) and for the Northridge earthquake (upper triangle, red).

DOI: 10.1002/eqe

$p_{12}^2 = 0.34$ in Equation (23)), the following spatial correlation coefficient can be obtained for two sites separated by 10 km:

$$\rho_{\mathrm{PGA,Ia}}(10) = 0.57\exp\left(\frac{-3\times 10}{10}\right) + 0.34\exp\left(\frac{-3\times 10}{60}\right) = 0.24 \tag{24}$$

At each individual site ($h = 0$ km), the correlations between PGA, Ia, and PGV can be obtained as $\rho_{PGA,Ia} = 0.91$, $\rho_{PGA,PGV} = 0.61$, and $\rho_{Ia,PGV} = 0.66$ from this study. These values are in good agreement with previous studies. For example, it was reported the correlations between these IMs being 0.83, 0.58, and 0.64, respectively [1]. These proposed correlation matrices will be used in the following example to generate spatially correlated fields of PGA, Ia, and PGV for calculating the Newmark displacement.

4. STOCHASTIC SIMULATION USING DATA REDUCTION TECHNIQUES

For a fully probabilistic seismic displacement analysis, three levels of variabilities must be rigorously accounted for in the analysis: (i) generation of all possible earthquake scenarios; (ii) generation of the vector IM maps considering inter-event variability, and their intra-event spatial correlation between all sites; and (iii) uncertainties of empirical displacement prediction equations. For this purpose, the MCS-based stochastic simulation is the only viable solution. However, using conventional MCS will result in prohibitive computational cost. A computationally efficient framework must be developed to rigorously account for all sources of uncertainties.

4.1. Importance sampling technique

Importance sampling (IS) is a widely used data reduction technique to sample earthquake scenarios (e.g., [14, 31]). Generally speaking, earthquake magnitudes follow some recurrence relationships (e.g., Gutenberg–Richter law). Random sampling of earthquake scenarios is inefficient because large magnitude events are infrequently sampled although they are more important in hazard analysis. Instead, the IS technique preferentially samples the rare large events. The effects of IS technique are accounted for through assigning suitable weights to each sampling so that the occurrence rate of the earthquake scenarios can still be correctly represented. The procedure is introduced as follows:

(1) Rupture location is assumed to be uniformly distributed within each earthquake source zone, and a magnitude density function $f(m)$ is used to characterize each earthquake source. The range of magnitude (between a lower bound $M_{\min}$ and an upper bound $M_{\max}$) is divided into n_m intervals. The interval can vary with magnitude (i.e., a smaller interval for larger magnitudes).
(2) A magnitude can be randomly selected within each interval $[m_k, m_{k+1}]$, with an actual probability of $\int_{m_k}^{m_{k+1}} f(m)dm$. So, a total of n_m earthquake events can be sampled for each rupture location. The sampling probability is $1/n_m$ for each event.
(3) The IS weight for each sampled scenario k is computed as

$$\mathrm{IS}_k = \frac{\int_{m_k}^{m_{k+1}} f(m)dm}{1/n_m} \tag{25}$$

(4) If a total of N_M magnitude-location scenarios are generated from all earthquake sources, the actual annual occurrence probability for scenario j is assigned as

$$P_j = \frac{\mathrm{IS}_j}{\sum_{i=1}^{N_M}\mathrm{IS}_i} \tag{26}$$

4.2. Stratified sampling method

In conventional MCS, a constant number of ground motion maps are generated for each earthquake scenario. Yet, the number of intensity maps can be reduced for some unimportant events (e.g., a

DOI: 10.1002/eqe

small-magnitude far-distance event). Using the stratified sampling (SS) technique [32], an optimal number of intensity maps can be assigned for each scenario.

Let N_j denote the number of corresponding intensity maps for the j-th event. The total number of intensity maps is $N = \sum_{j=1}^{J} N_j$ for all events. Define $\boldsymbol{X}_{ir} = \left\{X_{ir}^{(1)}, \ldots, X_{ir}^{(N)}\right\}$, where $X_{ir}^{(n)}$ is a binary variable for the generated n-th IM map at site i for a given return period r years. Its value is defined as

$$X_{ir}^{(n)} = \begin{cases} 0, & \text{if } y_{in} < Y_{ir} \\ 1, & \text{if } y_{in} \geqslant Y_{ir} \end{cases} \tag{27}$$

where y_{in} is the generated IM value at site i for n-th IM map, $n = 1, \ldots, N$ is the total number of maps, and Y_{ir} is an IM value on the analytical hazard curve corresponding to a return period r at site i (Figure 3). Y_{ir} can be computed from the conventional PSHA procedure. Define $\overline{X}_{ijr} = \frac{1}{N_j} \sum_{n=1}^{N_j} X_{ijr}^{(n)}$, where $\{X_{ijr}^{(1)}, \ldots X_{ijr}^{(N_j)}\} \equiv \boldsymbol{X}_{ijr}$ represents a subset of $\boldsymbol{X}_{ir}$ corresponding to j-th event. Then, we can define $\overline{X}_{ir} = \sum_{j=1}^{J} P_j \overline{X}_{ijr}$, where P_j is the annual occurrence probability for the j-th scenario calculated by Equation (26). In fact, $\overline{X}_{ir}$ is the estimated value of Y_{ir}. The variance of $\overline{X}_{ir}$ can be calculated as [32]:

$$var\left(\overline{X}_{ir}\right) = \sum_{j=1}^{J} \frac{P_j^2 \varphi_{ijr}^2}{N_j} - \sum_{j=1}^{J} \frac{P_j^2 \varphi_{ijr}^2}{N} \tag{28}$$

where N is the total number of intensity maps and φ_{ijr}^2 is the population variation of $\boldsymbol{X}_{ijr}$ (for the j-th scenario at site i and return period r). Because $\boldsymbol{X}_{ijr}$ follows Bernoulli distribution, φ_{ijr}^2 can be calculate as

$$\varphi_{ijr}^2 = P\left(y_{ij} \geqslant Y_{ir}\right) \cdot \left[1 - P\left(y_{ij} \geqslant Y_{ir}\right)\right] \tag{29}$$

where y_{ij} is an IM variable for scenario j at site i, and $P(y_{ij} \geq Y_{ir})$ is the theoretical probability of exceeding Y_{ir} at site i for event j.

Under the constraint that $\sum_{j=1}^{J} N_j = N$, the optimal N_j that minimizes $var\left(\overline{X}_{ir}\right)$, called Neyman allocation [33], would be

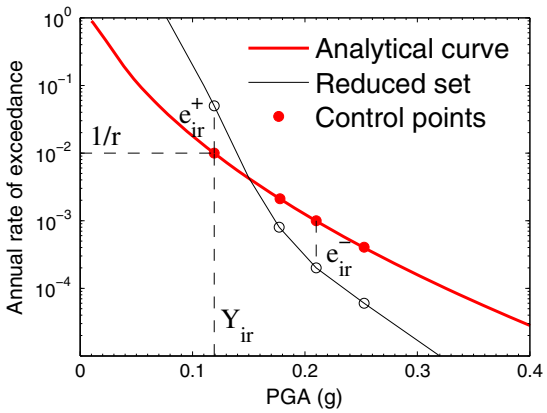


Figure 3. Errors between the analytical curve and reduced set hazard curve for PGA.

DOI: 10.1002/eqe

$$N_j = N \cdot \left[\frac{P_j \varphi_{ijr}}{\sum_{j=1}^{J} P_j \varphi_{ijr}} \right] \tag{30}$$

Then, the value of N_j for each event can be given by combining Equations (29) and (30):

$$N_j = N \cdot \left[\frac{P_j \sqrt{P(y_{ij} \geqslant Y_{ir}) \cdot [1 - P(y_{ij} \geqslant Y_{ir})]}}{\sum_{j=1}^{J} P_j \sqrt{P(y_{ij} \geqslant Y_{ir}) \cdot [1 - P(y_{ij} \geqslant Y_{ir})]}} \right] \tag{31}$$

This method results in an optimal number of ground motion IM maps, N_j, for each scenario j for each site i that minimizes the generated and the analytical hazard curves at a given return period r. We used four return periods (r =100, 475, 1000, and 2475 years) in this study. An averaged value of N_j over all sites, all return periods and all IMs are chosen as the final number of ground motion maps to be generated for each scenario. Finally, the annual occurrence probability for each ground motion intensity map is

$$P_n = \frac{P_j}{N_j} \qquad (n = 1, \ldots, N) \tag{32}$$

4.3. Optimization-based probabilistic scenario method

The optimization-based probabilistic scenario (OPT) method uses a mixed-integer linear programming technique to further reduce the candidate set of ground motion maps obtained by the SS method [34, 35]. The key feature of OPT method is to select the subset of candidate maps so as to minimize the difference between the hazard curves based on the stochastically simulated IM maps and the analytical curves (or 'true' hazard curves) obtained by conventional PSHA at all sites. Considering a set of R controlling points on the true hazard curve at i-th ($i = 1,\ldots,I$) site with r-year return periods, the objective function is

$$\min\left\{ \sum_{i=1}^{I} \sum_{r}^{R} w_r \left(e_{ir}^{+} + e_{ir}^{-} \right) \right\} \tag{33}$$

where e_{ir}^{+} and e_{ir}^{-} are the (absolute) errors for overestimation and underestimation of the hazard curve and w_r is the weight for the return period r, as demonstrated in Figure 3. Given that it is more important to consider the larger return periods, the weight w_r is assigned as the same as the corresponding return period r in this study (i.e., $w_r = r$). If the reduced hazard curve overestimates the true hazard curve, e_{ir}^{+} and e_{ir}^{-} are a positive value and 0, respectively, and they equal to 0 and a positive value when the reduced curve underestimates the true hazard curve.

Assume P_n is the annual occurrence probability for each ground motion map ($n = 1, \ldots, N$); Y_{ir} is the intensity value from analytical hazard curve for return period r-year at site i, and y_{in} is the intensity value at site i on the n-th ground motion map. Then, the constrained equilibrium function reads:

$$\sum_{n=1}^{N} \{P_n \cdot I(y_{in} \geqslant Y_{ir})\} - e_{ir}^{+} + e_{ir}^{-} = \frac{1}{r} \tag{34}$$

where $I(y_{in} \geqslant Y_{ir})$ is a binary function (equals 1 when $y_{in} \geqslant Y_{ir}$ and 0 otherwise). In fact, $\sum_{n=1}^{N} \{P_n \cdot I(y_{in} \geqslant Y_{ir})\}$ is the simulated exceedance probability of IM at site i for a return period of r (i.e., $1/r$ is the analytical exceedance probability). Two additional constraints are applied to restrict the total number of reduced IM maps being less than a user-defined number N_{set}:

DOI: 10.1002/eqe

$$P_n \leqslant Z_n, \quad \forall n \text{ and } \sum_{n=1}^{N} Z_n \leqslant N_{set} \tag{35}$$

where Z_n is an indicator variable ($Z_n = 1$ if the n-th IM map is retained in the reduced set, 0 otherwise). Each individual variable is constrained as follows:

$$Z_n \in \{0, 1\} \quad \forall n; \quad e_{ir}^{+}, e_{ir}^{-} \geqslant 0 \quad \forall i, r; \quad 0 \leqslant P_n \leqslant 1 \quad \forall n \tag{36}$$

The optimization process is initiated by a mixed-integer linear program to solve for Z_n and P_n through the objective function Equation (33) under the constraints given by Equations (34–36). If the calculated $P_n = 0$, the n-th IM map is removed from the reduced set. It is worth pointing out that the obtained P_n for each IM map is updated from that computed from SS technique by Equation (32).

These data reduction techniques can greatly reduce the computational cost without compromising the accuracy. The method is compared in details with the conventional MCS method in the next section.

4.4. Summary of the fully probabilistic analysis procedure

A fully probabilistic analysis procedure is summarized in the following steps:

Step 1: A set of earthquake magnitude-location scenarios can be simulated using stochastic method following magnitude-recurrence relationships. Important sampling technique is used in this step to reduce the number of samplings.

Step 2: The median predicted values of IMs and their corresponding standard deviation (τ_i and σ_{ij}) of the inter/intra-event residuals for each site are computed using Equation (5).

Step 3: The inter-event residuals (η_{ij}) are randomly generated following univariate normal distribution $\eta_{ij} = N(0,\tau_i)$. The spatially correlated intra-event residuals (ε_{ij}) are randomly generated for n IMs at all sites following multivariate normal distribution with a zero mean, standard deviation (σ_{ij}), and the total spatial correlation matrix $\mathbf{R}$.

Step 4: Ground motion IM maps are calculated by combining the median, inter-event, and intra-event residuals for each scenario. The SS and OPT methods are used to determine the number of ground motion IM maps need to be generated, and P_n is calculated via Equation (34) for each IM map.

Step 5: The median predicted Newmark displacement and the standard deviation σ_D are computed for each IM map using predictive models. The displacement residuals are randomly generated at all sites following a univariate normal distribution with a zero mean, standard deviation (σ_D), and a site-to-site correlation coefficient of $\rho_D(h)$, which will be reported in a separate study. In the following example, the range of spatial correlation of displacement residuals is assumed to be 10 km, that is, $\rho_D(h) = \exp(-3h/10)$. Our preliminary study shows that $\rho_D(h)$ does not significantly influence the regional hazard analysis.

Step 6: A total number of N_D displacement maps are retained using the SS method. For j-th displacement map, the corresponding probability is calculated as $P_{Dj} = \frac{P_n}{N_n}$, where N_n is the assigned number of displacement maps for the n-th IM map by SS method. Finally, the annual probability of exceedance λ_{D*} for specific value D^* for site i can be computed as

$$\lambda_{D^*} = \sum_{j=1}^{N_D} P_{Dj} \cdot I\left(D_{ij} \geqslant D^*\right) \tag{37}$$

where $I(D_{ij} \geq D^*)$ is again a binary function (equals 1 if the argument is true, and 0 else).

5. AN ILLUSTRATIVE EXAMPLE

5.1. Problem description

In this section, a hypothetical area is investigated by the proposed fully probabilistic approach. The 30×30 km area is divided into 900 sites separated by 1×1 km in distance. A constant critical

DOI: 10.1002/eqe

acceleration $a_c = 0.1$ g is assigned to all sites. It is assumed the epicenters of all rupture scenarios are located along the 30-km long linear fault trace close to this area, shown in Figure 4(a). The 30-km long linear source is further discretized into five 6-km long segments, and the epicenter of rupture scenarios is randomly distributed within each segment. The Wells and Coppersmith empirical equation [36] is used to estimate the fault rupture length for each scenario:

$$\log_{10}(L) = -3.22 + 0.69M_w \tag{38}$$

where M_w is the moment magnitude and L represents fault rupture length.

The following Gutenberg–Richter relationship is assumed to describe the seismicity of the source:

$$\log_{10}\lambda_m = 4.4 - 1.0M_w \tag{39}$$

where λ_m is the mean annual rate of exceedance of the moment magnitude M_w. The minimum and maximum magnitudes are set as 4.4 and 7.5, respectively. The GMPEs [24, 25] are used to estimate the predicted median values for PGA, PGV, and Ia, respectively. The four aforementioned models (Equations (1)–(4)) are adopted to compute the Newmark displacement.

The IS method is applied to stratify the range of magnitudes. The partition interval is 0.3 for $4.4 \leqslant M_w < 5.6$, 0.2 for $5.6 \leqslant M_w \leqslant 6.6$, and 0.1 for $M_w > 6.6$. In this example, 18 scenarios are sampled within each fault segment, resulting in a total of 90 scenarios considered. After applying the SS and OPT method, 400 maps are generated for each IM (PGA, Ia, and PGV) in the 'reduced set' with assigned probability P_n. One realization of randomly generated PGA, Ia, and PGV maps is

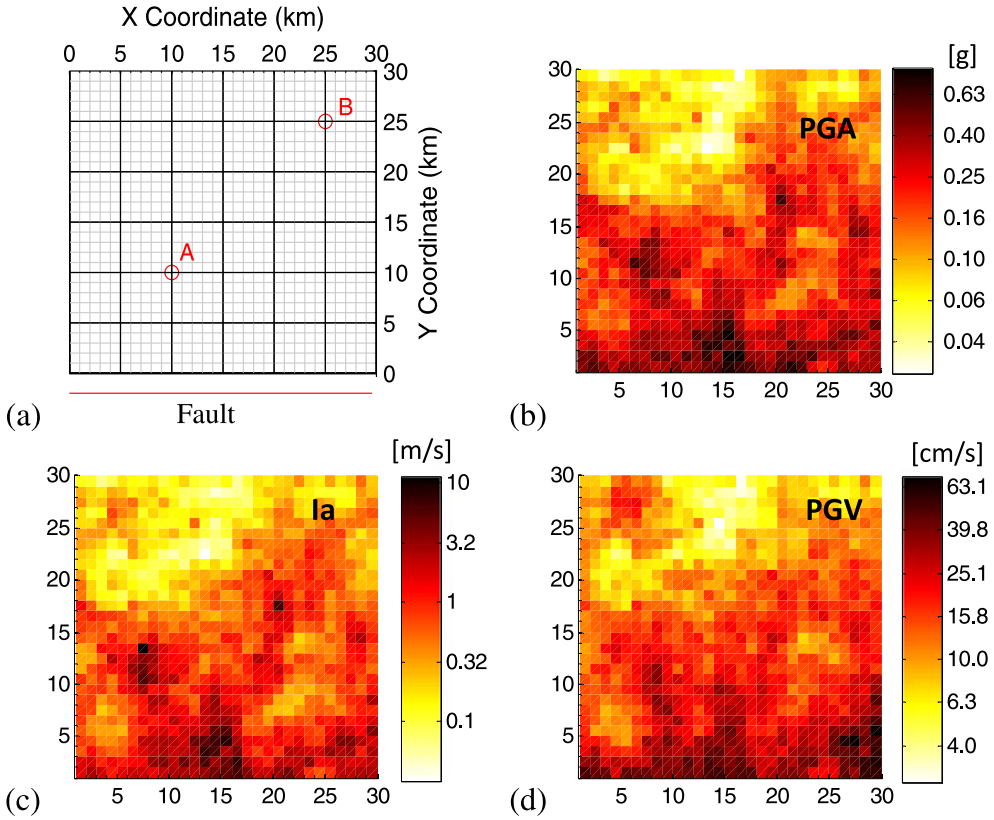


Figure 4. (a) A 30 × 30 km area divided into 1 × 1 km grids showing the location of the fault trace. Site A and Site B are two selected sites for comparison. Subplots (b), (c), and (d) show one realization of spatially correlated vector IM fields for a moment magnitude 7 earthquake scenario for (b) PGA (in unit of g), (c) Arias intensity (Ia) (in unit of m/s), and (d) peak ground velocity (PGV) (in unit of cm/s), respectively, all presented in $\log_{10}$ scale.

DOI: 10.1002/eqe

illustrated in Figure 4 for a moment magnitude 7 earthquake scenario. By applying SS method, a total of 3000 displacement maps are generated in the reduced set each with assigned probability P_D.

After the earthquake scenarios are sampled using IS, the conventional MCS method is used to generate 100 sets of inter and intra-event residuals for each IM and each earthquake scenario for comparison (i.e., a total of 9000 IM maps, called the 'large set'). For each group of intensity maps, the MCS is also carried out to obtain the 60 sets of displacement maps, resulting in a total number of 540,000 displacement maps. It is also noted that if the conventional MCS is conducted from the beginning of the procedure (IS is not used), the required computational cost should increase by more than three orders of magnitude.

5.2. *Hazard consistency of the reduced sets*

In this section, the hazard consistency is checked by comparing the intensity and displacement hazard curves obtained from the reduced set and the large set of IM maps and displacement maps. The simulated intensity hazard curve can be obtained by calculating the probability of exceedance for each IM and each site as

$$\lambda_{y^*} = \sum_{n=1}^{N} P_n \cdot I(y_{in} \geqslant y^*) \tag{40}$$

where P_n is the probability for the n-th IM map, y_{in} represents IM value at site i on the n-th IM map, and $I(y_{in} \geqslant y^*)$ is a binary function (equals 1 if $y_{in} \geqslant y^*$, 0 otherwise). On the other hand, the analytical (true) intensity hazard curve for IM Y at each site can be computed using PSHA approach:

$$\widetilde{\lambda}_{y^*} = \iint P(Y \geqslant y^* | m, r) f(m) f(r) dm dr \tag{41}$$

where $P(Y \geqslant y^* | m,r)$ is computed using GMPEs by assuming lognormal distribution of IM.

Equation (37) is used to compute the probability of exceedance of Newmark displacement. For the scalar model (Equation (1a) and (1b)), the analytical displacement hazard curve can be computed as

$$\widetilde{\lambda}_{D(x)} = \int P(D \geqslant x | \mathrm{IM} = y) f_{\mathrm{IM}}(y) dy \tag{42}$$

where $f_{\mathrm{IM}}(y)$ means the probability of occurrence of IM level y and $P(D \geqslant x | \mathrm{IM} = y)$ refers to the probability that the displacement level x is exceeded for the given IM value y. For the models using two-IMs (Equations (2–3)), the joint probability density function for the vector **IM** must be considered [37]. The analytical displacement hazard curve is calculated as

$$\widetilde{\lambda}_{D(x)} = \int P(D \geqslant x | \mathrm{IM}_1 = y, \mathrm{IM}_2 = z) f_{\mathrm{IM}_1,\mathrm{IM}_2}(y, z) dz dy \tag{43}$$

where $f_{\mathrm{IM}_1,\mathrm{IM}_2}(y, z)$ is the joint occurrence probability for IM_1 equals y and IM_2 equals z [1]. The analytical function for the three-IM case is more complicated and will not be presented in this paper.

Intensity hazard curves obtained from the reduce set and the large set are compared with the analytical hazard curves for PGA, Ia, and PGV in Figure 5 for two representative sites A and B shown in Figure 4(a). Quite consistent results can be observed for both sites and all IMs. The displacement hazard curves for the two sites are also compared in Figure 6 using the scalar (PGA) model and (PGA, Ia) model. It can be seen that the displacement hazard curves obtained by using data reduction technique are in reasonable agreement with these obtained using conventional MCS method, although the former only requires about one-180th number of realizations of the latter. Also, all these hazard curves obtained from sampling maps are consistent with the analytical hazard curves (Equations (42, 43)). By this example, it is demonstrated that the proposed computational framework can result in stable, fast, and hazard-consistent results and can be used to estimate the seismic risk over a large region.

DOI: 10.1002/eqe

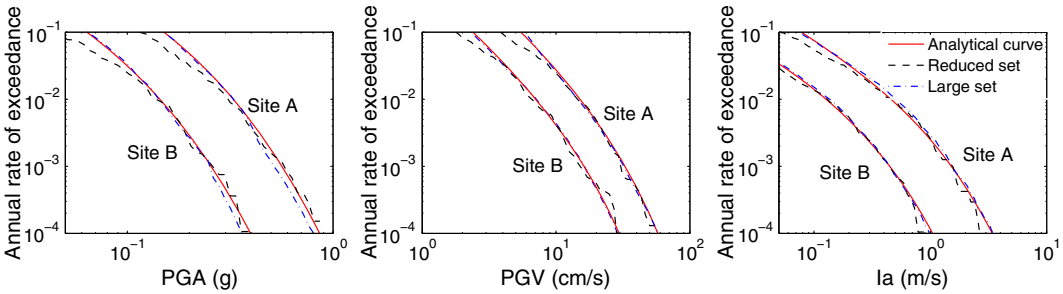


Figure 5. Comparison of seismic hazard curves obtained using the reduced intensity measure maps (400 maps) and the large intensity measure maps (9000 maps) method for PGA, peak ground velocity, and Arias intensity at site A and site B. The analytical curves obtained by Equations (42, 43) are also shown.

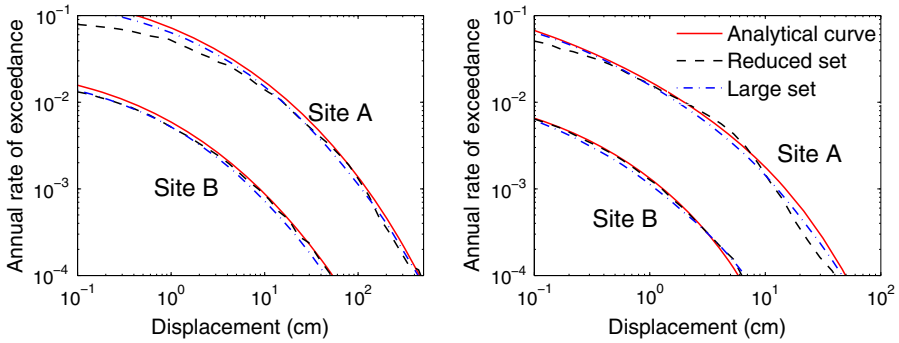
5.3. *Importance of spatial correlation*

In this section, the importance of spatial correlation on the regional-scale hazard analysis is highlighted by considering several special cases. Given a specified value of D (denoted as D^*) and its exceedance area ratio AR^*(defined as the ratio of the areas where displacements exceed the specified D^* value against the total area of the region), the annual rate of exceedance (termed as 'aggregated displacement hazard curve') can be computed as

$$\lambda = \sum_{j=1}^{N_D} P_{Dj} I_j(D > D^* \,\&\, AR > AR^*) \tag{44}$$

where P_{Dj} is the occurrence probability for displacement map j and $I_j(D > D^* \,\&\, AR > AR^*)$ is a binary function for displacement map j (equals 1 if the argument is true and 0 otherwise).

On the other hand, the analytical displacement hazard curve for each individual site can be computed independently using Equations (42–43), as is discussed before. Figure 7 is such an example showing contours of the displacements over this region for a return period of 2475 years obtained by the analytical methods, termed as the equal hazard displacement maps. Figure 8 shows the aggregated displacement hazard curve using the scalar (PGA) model and the (PGA, Ia) model, by assuming AR^* is 25%. The spatial correlation of IMs is assumed to follow zero correlation, model correlation via Equation (20), and perfect correlation. Accordingly, the range of spatial correlation of displacement residuals, $\rho_D(h)$, is assumed to be zero, 10 km, and infinite for these cases. The results demonstrate that, ignoring spatial correlation would yield an underestimated displacement hazard curve, especially for the rare cases. The case of perfect correlation, on the other hand, would overestimate the displacement hazard. It may also be tempting to convolve the



(a) PGA model (b) (PGA, Ia) model

Figure 6. Comparison of displacement hazard curves at two sites using the reduced set and the large set. (a) PGA model and (b) (PGA, Ia) model. The analytical curves obtained by Equations (42, 43) are also shown.

DOI: 10.1002/eqe

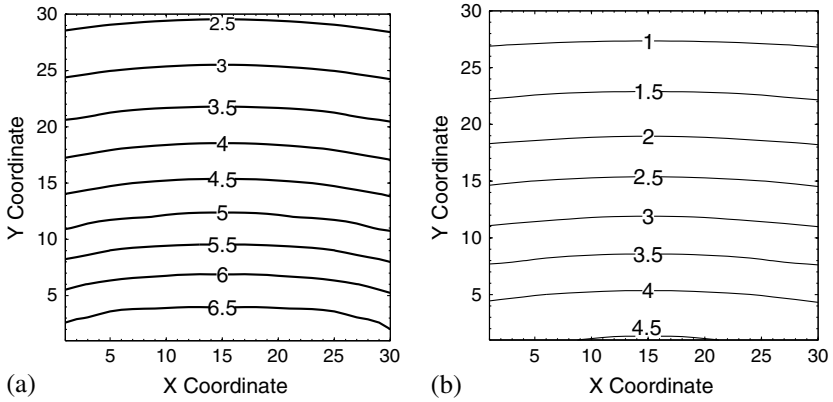


Figure 7. The equal hazard displacement maps (in cm) for a return period of 2475 years obtained by analytical methods using (a) PGA model and (b) (PGA, Ia) model.

analytical displacement hazard curve at individual sites directly to obtain the aggregated displacement hazard curve for a given annual rate of exceedance value and AR^*. More specifically, the 'convolved curve' simply chooses the $1–AR^*$ percentile displacement directly from the equal hazard displacement maps (such as Figure 7) for each return period. However, the process implicitly assumes that the sliding displacements for a certain return period occur simultaneously at all locations, that is, they are perfectly correlated. It can be further verified in Figure 8 that the 'convolved curve' is very close to the 'perfect correlation' curve, and both of them overestimate hazard estimate for spatially distributed slopes. Although only the scalar (PGA) model and the vector (PGA, Ia) model are used in Figure 8, similar conclusions can be drawn if other vector models are used.

5.4. *Influence of displacement prediction models*

In this section, the displacement hazard curves obtained using four different displacement prediction models are compared. Figure 9 shows the displacement hazard curves for individual sites (A and B), as well as for a given area ratio AR^* of 5% and 25%, respectively. The spatial correlation of vector IMs are computed using Equation (20). For all these cases, the displacement hazard curves obtained from the scalar PGA model are significantly higher than these obtained from vector models. The large discrepancy is not unexpected since a scalar displacement prediction model usually cannot satisfy the sufficiency requirement, for example, the model exhibits systematic bias over earthquake magnitudes [1] On the other hand, rather consistent results are obtained using three vector IM models, because the sufficiency requirement can be more easily satisfied using a vector model. The results clearly demonstrated the advantage of using vector IMs in displacement hazard analysis.

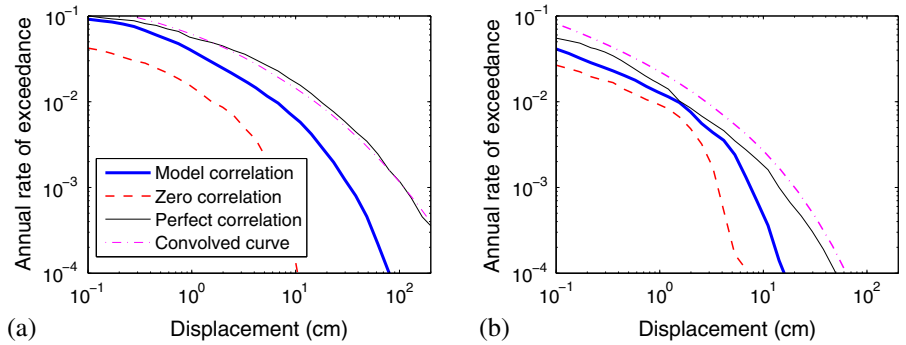


Figure 8. Displacement hazard curves for exceedance area ratio AR^* as 25% using (a) PGA model, and (b) (PGA, Ia) model.

DOI: 10.1002/eqe

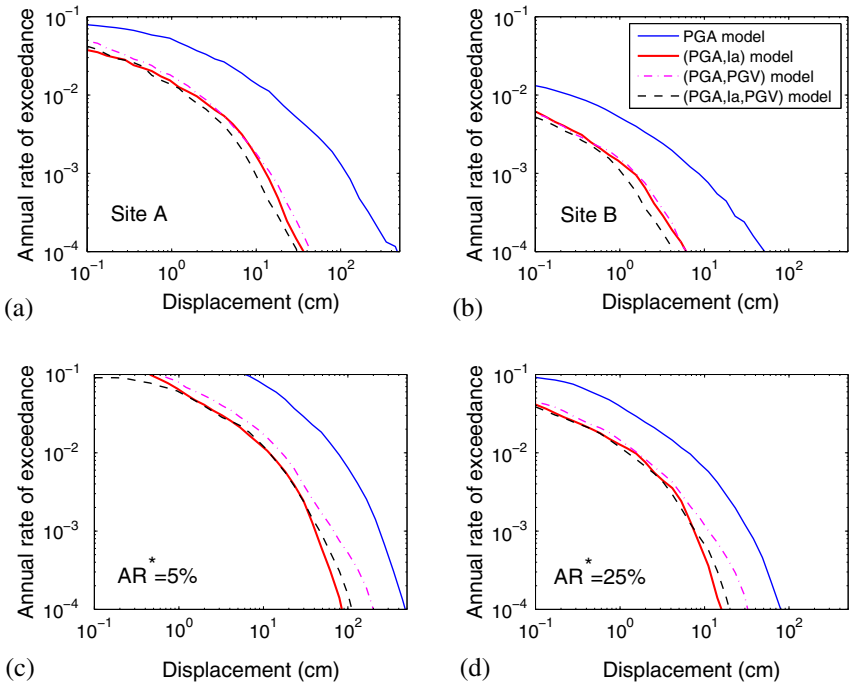


Figure 9. Displacement hazard curves using by different displacement prediction models for (a) site A, (b) site B, (c) exceedance area ratio AR^* as 5%, and (d) AR^* as 25%.

6. CONCLUSIONS

This paper provided a computationally efficient framework to conduct a fully probabilistic hazard analysis for spatially distributed slopes. The cross-correlations between the vector **IM** [PGA, Ia, PGV] are developed on the basis of geostatistical modeling of strong motion data from 11 recent earthquakes. The coregionalization matrices provide a positive definite covariance matrix that enables generation of random fields of the vector IM preserving their spatial correlations.

The developed covariance model for the vector **IM** [PGA, Ia, PGV] is then used to quantify the spatial variability in a hypothetical region. The intensity maps and sliding displacement maps are generated by Monte Carlo method, and aleatory variability is incorporated in each step. To reduce the computational cost, several state-of-the-art data reduction techniques are also applied. The difference of hazard curves between 'reduced set' and 'large set' implies that results obtained from data reduction techniques can provide accurate results. The importance of spatial correlation is also emphasized using the example. Neglecting the spatial variability or using the convolved analytical solution would result in either underestimated or overestimated displacement hazard curves.

Finally, the displacement hazard curves computed by different predictive models (PGA model, (PGA, Ia) model, (PGA, PGV) model and (PGA, Ia, PGV) model) are compared. Except for the scaler PGA model, all vector models yield consistent displacement hazard curves for individual sites as well as for the whole region. The vector models demonstrate significant advantages over the scalar model, underlining the importance of using spatially correlated vector IM in seismic hazard analysis of spatially distributed slopes.

ACKNOWLEDGEMENTS

The authors acknowledge financial support from Hong Kong Research Grants Council Grant 620311, and DAG08/09.EG13, DAG11EG03G to conduct this study. Strong motion data are obtained from resources in the public domain, including CESMD strong motion database (http://strongmotioncenter.org/, last accessed Nov. 2012), COSMOS strong motion database (http://www.cosmos-eq.org/, last accessed Nov. 2012), K-NET strong motion database (http://www.k-net.bosai.go.jp/, last accessed Nov. 2012) and PEER-NGA database (http://peer.berkeley.edu/products/strong_ground_motion_db.html, last accessed Nov. 2012).

DOI: 10.1002/eqe

REFERENCES

1. Rathje EM, Saygili G. Probabilistic seismic hazard analysis for the sliding displacement of slopes: scalar and vector approaches. *Journal of Geotechnical and Geoenvironmental Engineering* 2008; **134**(6):804–814.
2. Rathje EM, Saygili G. Estimating fully probabilistic seismic sliding displacements of slopes from a pseudoprobabilistic approach. *Journal of Geotechnical and Geoenvironmental Engineering* 2011; **137**(3):208–217.
3. Crowley H, Bommer JJ. Modelling seismic hazard in earthquake loss models with spatially distributed exposure. *Bulletin of Earthquake Engineering* 2006; **4**:249–273.
4. Bommer JJ, Crowley H. The influence of ground-motion variability in earthquake loss modelling. *Bulletin of Earthquake Engineering* 2006; **4**:231–248.
5. Arias A Measure of earthquake intensity. In *Seismic design for nuclear power plants*, Hansen RJ (ed.). Massachusetts Institute of Technology Press: Cambrige, MA, 1970; 438–483.
6. Boore DM, Gibbs JF, Joyner WB, Tinsley JC, Ponti DJ. Estimated ground motion from the 1994 Northridge, California, earthquakes at the site of the Interstate 10 and La Cienega Boulevard bridge collapse, West Los Angeles, California. *Bulletin of the Seismological Society of America* 2003; **93**(6):2737–2751.
7. 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.
8. Jayaram N, Baker JW. Correlation model for spatially distributed ground-motion intensities. *Earthquake Engineering and Structural Dynamics* 2009; **38**:1687–1708.
9. Foulser-Piggott R, Stafford PJ. A predictive model for Arias intensity at multiple sites and consideration of spatial correlations. *Earthquake Engineering and Structural Dynamics* 2012; **41**(3):431–451.
10. 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.
11. Loth C, Baker JW. A spatial cross-correlation model of ground motion spectral accelerations at multiple periods. *Earthquake Engineering & Structural Dynamics* 2013; **42**(3):397–417.
12. Sokolov V, Wenzel F. Influence of ground-motion correlation on probabilistic assessments of seismic hazard and loss: sensitivity analysis. *Bulletin of Earthquake Engineering* 2011; **9**(5):1339–1360.
13. Bal IE, Bommer JJ, Stafford PJ, Crowley H, Pinho R. The influence of geographical resolution of urban exposure data in an earthquake loss model for Istanbul. *Earthquake Spectra* 2010; **26**(3):619–634.
14. Jayaram N, Baker JW. Efficient sampling and data reduction techniques for probabilistic seismic risk assessment. *Earthquake Engineering & Structural Dynamics* 2010; **39**:1109–1131.
15. Han Y, Davidson RA. Probabilistic seismic hazard analysis for spatially distributed infrastructure. *Earthquake Engineering & Structural Dynamics* 2012; **41**(15):2141–2158.
16. Newmark NM. Effects of earthquakes on dams and embankments. *Geotechnique* 1965; **15**(2):139–160.
17. Jibson RW. Regression models for estimating coseismic landslide displacement. *Engineering Geology* 2007; **91**:209–218.
18. Ambraseys NN, Menu JM. Earthquake-induced ground displacements. *Earthquake Engineering and Structural Dynamics* 1988; **16**:985–1006.
19. Saygili G, Rathje EM. Empirical predictive models for earthquake-induced sliding displacements of slopes. *Journal of Geotechnical and Geoenvironmental Engineering* 2008; **134**(6):790–803.
20. Hsieh SY, Lee CT. Empirical estimation of Newmark displacement from the Arias intensity and critical acceleration. *Engineering Geology* 2011; **122**:34–42.
21. Kaklamanos J, Baise LG. Model validations and comparisons of the next generation attenuation of ground motions (NGA-West) project. *Bulletin of the Seismological Society of America* 2011; **101**(1):160–175.
22. Abrahamson NA, Youngs RR. A stable algorithm for regressionanalyses using the random effects model. *Bulletin of the Seismological Society of America* 1992; **82**:505–510.
23. Joyner WB, Boore DM. Methods for regression analysis of strong-motion data. *Bulletin of Seismological Society of America* 1993; **83**:469–487.
24. Campbell KW, Bozorgnia Y. NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5% damped linear elastic response spectra for periods ranging from 0.1 to 10 s. *Earthquake Spectra* 2008; **24**(1):139–171.
25. Campbell KW, Bozorgnia Y. A comparison of ground motion prediction equations for Arias intensity and cumulative absolute velocity developed using a consistent database and functional form. *Earthquake Spectra* 2012; **28**(3):931–941.
26. Jayaram N, Baker JW. Statistical tests of the joint distribution of spectral acceleration values. *Bulletin of Seismological Society of America* 2008; **98**(5):2231–2243.
27. Goovaerts P *Geostatistics for Natural Resources Evaluation*. Oxford University Press: Oxford, New York, 1997.
28. Journel AG, Huijbregts CJ. *Mining Geostatistics*. Academic Press: London, 1978; 600.
29. Goulard M, Voltz M. Linear coregionalization model: tools for estimation and choice of crossvariogram matrix. *Mathematical Geology* 1992; **24**(3):269–286.
30. 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), in press. DOI: 10.1785/0120130061
31. Kiremidjian A, Stergiou E, Lee R. Issues in seismic risk assessment of transportation networks. In *Earthquake Geotechnical Engineering*, Pitilakis KD (ed.). Springer, Dordrecht, The Netherlands: 2007; 461–480.

DOI: 10.1002/eqe

32. Cochran W. *Sampling Techniques* (3rd edn). Wiley: New York, 1977.
33. Neyman J. On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society* 1934; **97**:558–608.
34. Chang S, Shinozuka M, Moore J. Probabilistic earthquake scenarios: extending risk analysis methodologies to spatially distributed systems. *Earthquake Spectra* 2000; **16**(3):557–572.
35. Vaziri P, Davidson R, Apivatanagul P, Nozick L. Identification of optimization-based probabilistic earthquake scenarios for regional loss estimation. *Journal of Earthquake Engineering* 2012; **16**(2):296–315.
36. Wells DL, Coppersmith KJ. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of Seismological Society of America* 1994; **84**(4):974–1002.
37. Bazzurro P, Cornell CA. Vector-valued probabilistic seismic hazard analysis (VPSHA). *The 7th US national conference on Earthquake Engineering*, Boston, Massachusetts, July 21–25, 2002.

DOI: 10.1002/eqe