Bulletin of the Seismological Society of America, Vol. 103, No. 2A, pp. 1117–1129, April 2013, doi: 10.1785/0120120185

Intra-Event Spatial Correlations for Cumulative Absolute Velocity, Arias Intensity, and Spectral Accelerations Based on Regional Site Conditions

by Wenqi Du and Gang Wang

Abstract Spatial correlations of ground-motion intensity measures (IMs) are essential for seismic analysis of spatially distributed systems. In this paper, geostatistical analysis is conducted to calculate the spatial correlations for cumulative absolute velocity (CAV), Arias intensity (Ia), and spectral accelerations (SA) using a total number of more than 1500 earthquake records from nine recent earthquakes occurred in Taiwan, California, and Japan. The results indicate that the spatial correlations for these IMs are closely related to the regional site conditions, and they can be predicted based on the spatial correlations of shear-wave velocity in the top 30 m (V_{S30}). In general, an IM recorded from a relatively homogeneous regional site condition tends to have a larger spatial correlation range than that from a heterogeneous site condition. Due to their intrinsic similarity to represent the integration of acceleration time histories, CAV and Ia have similar spatial correlation coefficients. Besides, the range of spatial correlation of SA generally increases as the spectral period increases. Simple predictive equations are proposed in this study to quantify the spatial correlations of CAV, Ia, and SA based on regional site conditions. Methods for data correction are also proposed to eliminate artificial correlations due to biased distance scaling and V_{S30} estimation in the database. Finally, Monte Carlo method is used to generate spatially distributed IMs. The results demonstrate that the annual frequency of exceedance curves for spatially distributed IMs differ significantly if different ranges of spatial correlations are used.

Introduction

Considering spatial distribution of ground-motion intensity measures (IMs) is important in seismic-hazard analysis of spatially distributed infrastructure systems such as long-span bridges, lifelines, railways, or geohazards (Lee and Kiremidjian, 2007). Traditional ground-motion prediction equations (GMPEs) usually provide the statistical characteristics of IMs at a particular location for a casual earthquake event. However, the statistical characteristics of IMs at spatially separated locations are often overlooked. In recent years, spatial correlations of some important IMs, such as the peak ground acceleration (PGA) and spectral acceleration (SA), have been developed by several researchers (Boore *et al.*, 2003; Wang and Takada, 2005; Goda and Hong, 2008; Jayaram and Baker, 2009; Goda and Atkinson, 2010; Sokolov *et al.*, 2010, Esposito and Iervolino, 2011). The influence of considering spatial correlation on developing GMPEs for spectral accelerations is also studied (Jayaram and Baker, 2010). To the best of the authors' knowledge, there is no spatial correlation study for cumulative absolute velocity (CAV) available in the literature. The spatial correlation of Arias intensity (Ia) was recently studied based on ground-motion data from the Northridge and Chi-Chi earthquakes (Piggott and Stafford, 2012). However, the study concluded that it is not executable to find a generic spatial correlation model for Ia because the derived spatial correlation coefficients differ significantly from event to event. This study is aimed at developing predictive models for the spatial correlation of these cumulative IMs and their relationship to regional site conditions.

CAV has been found to be a good index that is well correlated to structural damages (Electrical Power Research Institute [EPRI], 1988). CAV is defined as the time integration of absolute acceleration as follows:

$$\mathrm{CAV} = \int_0^{t_{\mathrm{tot}}} |a(t)|dt, \tag{1}$$

where $|a(t)|$ is the absolute value of the acceleration time history, and t_{tot} is the total duration of the ground motion time history. Similar to CAV, Ia is the integration of the square

of ground-motion acceleration time history over the total duration, given by the following equation (Arias, 1970):

$$\mathrm{Ia} = \frac{\pi}{2g}\int_0^{t_{\mathrm{tot}}} a(t)^2 dt, \tag{2}$$

where g is the acceleration of gravity. By definition, CAV and Ia share intrinsic similarity in that they both incorporate the cumulative effect of an acceleration time history. Hence they can capture multiple characteristics of ground motion, including the amplitude, the frequency content, and duration of the ground-motion time histories implicitly. Previous researches indicate that these cumulative IMs, such as CAV and its variants (e.g., CAV_5 in Kramer and Mitchell, 2006) as well as Ia, have improved efficiency in predicting soil liquefaction and earthquake-induced landslides than the other IMs related to the peak response of structures (e.g., SA; Kayen and Mitchell, 1997; Liyanapathirana and Poulos, 2004; Kramer and Mitchell, 2006; Jibson, 2007).

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 1994 Northridge and 1999 Chi-Chi earthquakes (e.g., Boore *et al.*, 2003; Wang and Takada, 2005; Jayaram and Baker, 2009). Pooled data of multiple earthquakes in some specific regions are also used (e.g., Goda and Hong, 2008; Goda and Atkinson, 2010; Sokolov *et al.*, 2010, Esposito and Iervolino, 2011). In this study, more than 1500 recorded strong-motion data from nine recent earthquakes occurred in Taiwan, Japan, and California are systematically compiled to evaluate the spatial correlations of various IMs. Geological information for each recording station is also assembled. The abundance of new data enables more robust estimation of spatial correlations, as well as a better understanding of their relationships to different regional site conditions. In the following sections, methods to estimate empirical spatial correlations are briefly described. Spatial correlations of CAV, Ia, and SA are calculated for each earthquake event. Simple empirical equations are provided to relate these spatial correlations to regional site conditions, which is represented by the spatial correlation of shear-wave velocity of soils in upper 30 m (V_{S30}). Correction methods are also proposed in this study to eliminate artificial correlations due to biased distance scaling and V_{S30} estimation in the database. Finally, an illustrative example is provided to highlight the importance of spatial correlations in estimating the annual frequency of exceedance curves for spatially distributed IMs.

Spatial Correlation Model

GMPEs typically assume that IMs follow lognormal distribution. The observed logarithmic IM, denoted as $\ln Y_{ij}$, of a ground-motion record can be written as follows:

$$\ln Y_{ij} = \overline{\ln Y_{ij}(M,R,\theta)} + \eta_i + \varepsilon_{ij}, \tag{3}$$

where Y_{ij} is the ground-motion IM of the jth record of the ith earthquake event, and $\overline{\ln Y_{ij}(M,R,\theta)}$ is the predicted median value of $\ln Y_{ij}$ based on the magnitude (M), rupture distance (R), and other parameters (θ) of this event. In this study, several recently developed GMPEs are used to estimate the predicted median values for CAV, Ia, and SA, respectively (e.g., Campbell and Bozorgnia, 2008, 2010, 2012), although the results are not particularly sensitive to the particular GMPEs used in the analysis. η_i and ε_{ij} represent the inter- and intra-event residuals, respectively, which are both assumed to be normally distributed independent random variables with zero means and standard deviations of τ_i and σ_{ij} (Brillinger and Preisler, 1984, 1985; Abrahamson and Youngs, 1992; Joyner and Boore, 1993). The standard deviation of the total residual term can be calculated by $\sigma_T = \sqrt{\sigma_{ij}^2 + \tau_i^2}$. Then the observed logarithmic IM follows lognormal distribution, denoted as $\ln Y_{ij} = N(\overline{\ln Y_{ij}}; \sigma_T)$. Accordingly, the normalized intra-event residuals can be computed as

$$\varepsilon'_{ij} = \frac{\varepsilon_{ij}}{\sigma_{ij}} \approx \frac{\ln Y_{ij} - \overline{\ln Y_{ij}(M,R,\theta)}}{\sigma_{ij}}, \tag{4}$$

where ε'_{ij} is the normalized residuals, and σ_{ij} is the intra-event standard deviation for site j which can be either estimated from samples or obtained from GMPEs. Note that equation (4) approximates the intra-event residuals using total residuals. Because the interevent residual is constant for each site during one earthquake event, this approximation does not affect the results of spatial correlation presented below.

For a given earthquake event, the interevent residual η_i is identical for all sites, but the value of ε_{ij} varies from site to site. Therefore, the total spatial correlation can be expressed as (Park *et al.*, 2007)

$$\rho_T(h) = \frac{\tau_i^2 + \rho_\varepsilon(h) \times \sigma_{ij}^2}{\tau_i^2 + \sigma_{ij}^2}, \tag{5}$$

where $\rho_\varepsilon(h) = \rho_{\varepsilon_{i,j1},\varepsilon_{i,j2}}(h)$ represents the empirical spatial correlation obtained from normalized intra-event residuals ε'_{ij}, and it is a function of the separation distance h between different sites. The values of and are usually provided by GMPEs, and $\rho_\varepsilon(h)$ can be estimated using empirical semivariogram that will be discussed in details in the following session.

Estimation of Empirical Semivariogram

Semivariogram is a useful statistical tool widely used to estimate the empirical spatial correlation of IMs. The semivariogram $\gamma(\boldsymbol{h})$ measures the average dissimilarity between data separated by a vector $\boldsymbol{h}$ (Goovaerts, 1997). Under the assumptions that the spatial correlation is isotropic and

second-order stationary, a scalar variable $h = \|\boldsymbol{h}\|$ can be used in the empirical semivariogram formulation:

$$\gamma(h) = \frac{1}{2}\mathrm{Var}[z_{u_i+h} - z_{u_i}] = \mathrm{Var}(z)[1 - \rho_\varepsilon(h)], \quad (6)$$

where z_{u_i} is a random distributed variable at position u_i. z_{u_i+h} is a random variable at a position separated by distance h from position u_i (Cressie, 1993). In this paper, z_{u_i} refers to the normalized intra-event residual ε'_{ij}. As an estimate of the theoretical semivariogram equation (6), an empirical semivariogram can be calculated from a sample dataset using estimators. For an earthquake event, all empirical data pairs are grouped if the separation distance of the data pair falls into a bin bounded by $[h - \Delta h/2, h + \Delta h/2]$, where Δh is the size of the bin. A classical estimator based on the method of moments can be defined as

$$\tilde{\gamma}(h) = \frac{1}{2|N(h)|}\sum_{i=1}^{N(h)}[z_{u_i+h} - z_{u_i}]^2, \quad (7)$$

where $\tilde{\gamma}$ represents empirical semivariogram, and $N(h)$ is the number of data pairs within this distance bin, z_{u_i+h} and z_{u_i} represents the ith data pair in this bin. Alternatively, a robust estimator (Cressie and Hawkins, 1980) is less sensitive to data outliers, and it is defined as

$$\tilde{\gamma}(h) = \frac{\left(\frac{1}{N(h)}\sum_{i=1}^{N(h)}|z_{u_i+h} - z_{u_i}|^{0.5}\right)^4}{0.914 + 0.988/N(h)}. \quad (8)$$

An example of the empirical semivariogram for the normalized intra-event CAV residuals for the Chi-Chi earthquake is demonstrated in Figure 1. Both classical and robust estimators yield similar empirical semivariograms. To get consistent results, the robust estimator is chosen throughout the study. As a suitable size of distance bin is also essential for the empirical estimation, some rules are applied. It has been proposed that size of the bin shall be no larger than a half of the maximum separation distance between data pairs, and each bin shall contain at least 30 pairs of data (Journel, 1978).

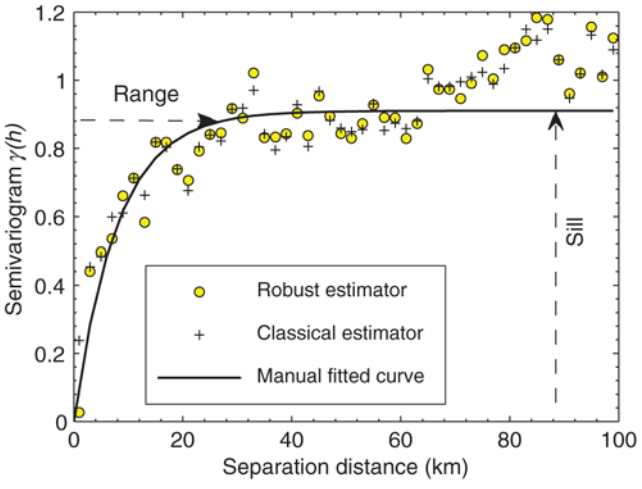

Figure 1. Example of empirical semivariogram of normalized intra-event CAV residuals for the Chi-Chi earthquake using classical and robust estimators. The color version of this figure is available only in the electronic edition.

Parametric Function

A parametric function is useful to represent the empirical semivariogram. Three basic second-order stationary and isotropic models can be considered, namely, Gaussian, spherical, and exponential models (Goovaerts, 1997). Because the exponential model has distinct advantage of simplicity, it is adopted in this study. The exponential model approximates the empirical semivariogram using the following functional form:

$$\tilde{\gamma}(h) = a[1 - \exp(-3h/b)], \quad (9)$$

where a is the sill of the semivariogram and also the population variance of empirical data, b is the range of the semivariogram, defined as the separation distance h at which $\tilde{\gamma}(h)$ equals 95% of the sill. In another word, the range b is the separation distance where 95% of the correlation vanishes. For illustration, the sill and the range are marked in Figure 1. It is worth pointing out that intra-event standard deviation $[\mathrm{Var}(z)]$ only affects the estimate of sill a, and it will not affect the value of range b if the function form of equation (9) is used. By normalization process, ε'_{ij} follows a normal standard distribution. Therefore, the sill should be equal to 1. Accordingly, the relationship between the spatial correlation and semivariogram can be simplified via equations (6) and (9) as

$$\rho_\varepsilon(h) = \exp(-3h/b). \quad (10)$$

Therefore, range b is the only unknown parameter to quantify the spatial correlation. Several approaches have been proposed in previous studies to fit the empirical data using the exponential model, such as the weighted-least-square method and the manual fitting method (Jayaram and Baker, 2009). Although the manual fitting method is relatively subjective, it has greater flexibility to better fit the empirical data. Therefore, it is employed in this study.

Spatial Correlations for CAV, Ia, and SA

Strong-Motion Database

A total of 1588 ground-motion recordings from nine earthquakes are compiled to develop the spatial correlation models for CAV, Ia, and SA. These earthquakes occurred in California (1994 Northridge earthquake, 2004 Parkfield earthquake, 2005 Anza earthquake, 2007 Alum Rock earthquake, and 2008 Chino Hills earthquake), in Japan (2000 Tottori earthquake, 2004 Niigata earthquake, and 2007 Chuetsu earthquake), and in the Taiwan region (1999 Chi-Chi

Table 1
Earthquake Events Used in This Study

Earthquake Name	Date (yyyy/mm/dd)	Moment Magnitude	Hypocenter Latitude (°)	Hypocenter Longitude (°)	Fault Mechanism	Number of Recordings
Chi-Chi	1999/09/20	7.62	23.860	120.800	Reverse-oblique	381
Northridge	1994/01/17	6.69	34.206	−118.554	Reverse	149
Parkfield	2004/09/28	6	35.817	−120.365	Strike-slip	89
Anza	2005/06/12	5.2	35.533	−116.578	Reverse-oblique	111
Alum Rock	2007/10/30	5.4	37.432	−121.776	Strike-slip	161
Chino Hills	2008/07/29	5.4	33.955	−117.765	Reverse-oblique	337
Tottori	2000/10/06	6.61	35.275	133.350	Strike-slip	112
Niigata	2004/10/23	6.63	37.307	138.839	Reverse	134
Chuetsu	2007/07/16	6.8	37.538	138.617	Reverse	114

Only recorded data within rupture distance of 120 km is included for Japan earthquakes.

earthquake). The recorded time histories for these events are obtained from CESMD, COSMOS for U.S. earthquakes, and K-NET for Japan earthquakes. The seismic information and site conditions are obtained from the Pacific Earthquake Engineering Research–Next Generation Attenuation (PEER–NGA) database and the table S1 database provided by Kaklamanos and Baise (2011). The V_{S30} data for recording stations for the Chi-Chi earthquake is updated according to the Taiwan Strong-Motion Instrumentation Program (TSMIP; Kuo *et al.*, 2012). The detailed information of the earthquakes is summarized in Table 1. The moment magnitude and rupture distance distribution of the data in the database is illustrated in Figure 2.

Following the procedures illustrated in the previous section, the normalized intra-event residuals of recorded ground motions from each event are computed and used to derive the empirical semivariograms for CAV, Ia, and SA, respectively. As will be discussed in details in later sections, the spatial correlations of these IMs are observed to be dependent on regional site conditions. The shear-wave velocity in the top 30 m (V_{S30}) at the record station is chosen as an index of the local site condition. Accordingly, the spatial correlation of

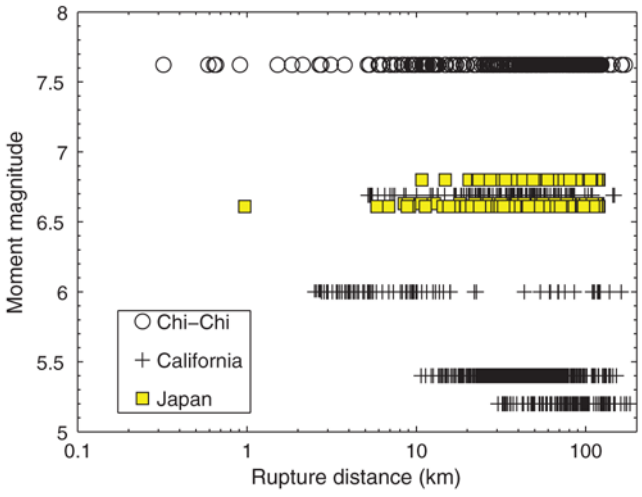

Figure 2. Magnitude and rupture distance distribution of records in the database. The color version of this figure is available only in the electronic edition.

V_{S30} among the recording stations for each event is quantified and used as an index to represent the characteristics of site conditions for that region.

Spatial Correlations of CAV and Ia

Before proceeding, the distribution of the residuals of IMs is examined against the rupture distance as well as the V_{S30} data to check whether the data exhibits significant biases. Figures 3 and 4 plot the distribution of the residuals against rupture distances and V_{S30} values for nine earthquake events, together with their linear trend lines, respectively. It is noted that GMPEs used in this study (e.g., Campbell and Bozorgnia, 2008, 2010, 2012) are developed from the PEER-NGA database. Therefore, it is not unexpected to observe significant biases in distance scaling for events out of the NGA database (e.g., all events except for the Chi-Chi and Northridge earthquakes). These residual terms, if used without correction, would result in artificial spatial correlation due to systematical predictive biases in the distance scaling (Piggott and Stafford, 2012) or in the local site condition represented by V_{S30} (Sokolov *et al.*, 2010). On the other hand, the residual terms of events in the PEER-NGA database, such as the Chi-Chi and Northridge earthquakes, exhibit no obvious bias. It is also noted that presently no predictive equation for CAV and Ia is available to incorporate these events out of the NGA database. Therefore, in order to reduce artificial correlations, a distance term and a V_{S30} term are added to correct the residuals for each event as follows:

$$\varepsilon_{\text{corr}} = \ln Y_{ij} - \overline{\ln Y_{ij}(M,R,\theta)} - [\varphi_1 + \varphi_2 \ln(R_{ij}) + \varphi_3 \ln(V_{s30})], \qquad (11)$$

where R_{ij} is the rupture distance of the jth recording and ith event, and φ_1, φ_2, and φ_3 are the coefficients obtained by linear regression. The corrected residuals show no significant bias against rupture distances and V_{S30} values for all events.

The second step is to calculate the normalized intra-event residuals. Because only two earthquakes (the Northridge and Chi-Chi earthquakes) are selected as part of the PEER-NGA database to develop GMPEs, it is not surprising that other

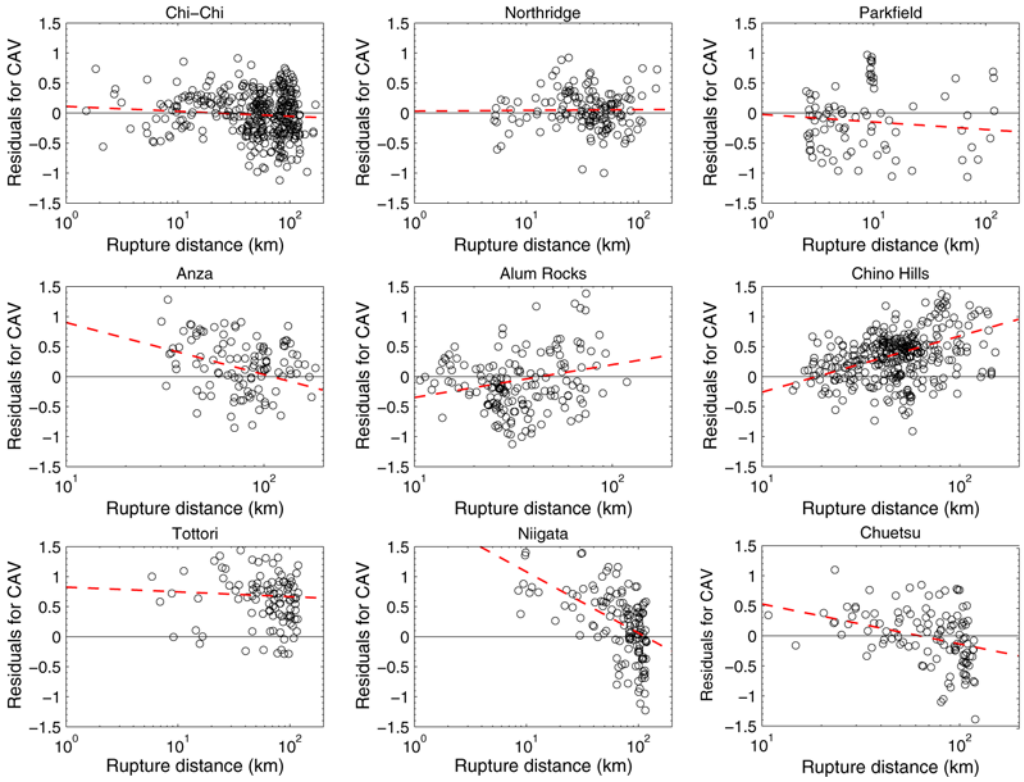

Figure 3. Distributions of residuals versus rupture distance (km) for nine earthquakes. The color version of this figure is available only in the electronic edition.

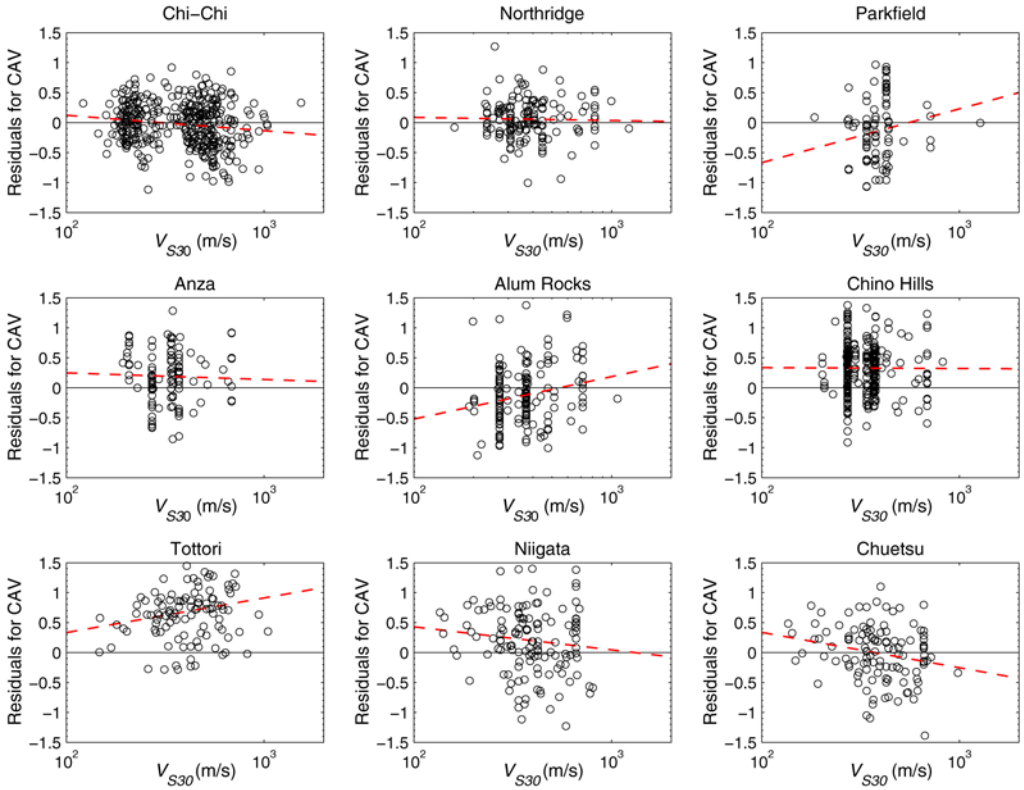

Figure 4. Distributions of residuals versus V_{S30} (m/s) for nine earthquakes. The color version of this figure is available only in the electronic edition.

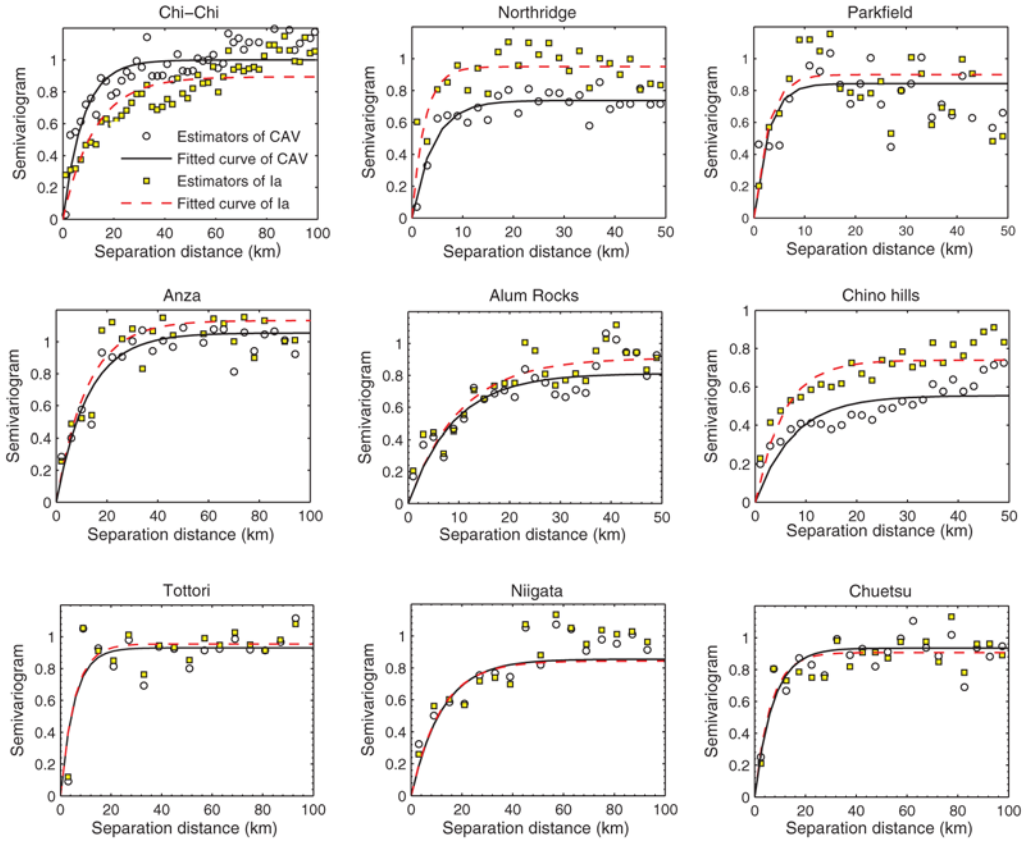

Figure 5. Experimental semivariograms for intra-event CAV and Ia residuals. The color version of this figure is available only in the electronic edition.

events' sample variations are somewhat different from these provided by GMPEs. In this study, the event-specific sample variances are used instead of these provided by GMPEs. Based on the total number of sites and their separation distances, the size of distance bins varies from 2 to 6 km for different events to ensure at least 30 pairs of data in each bin.

The empirical semivariograms and the manually fitted parametric functions for CAV and Ia are shown in Figure 5. All fitting curves perform reasonably well at the short distance range, which is the most concerned and important region. The computed values of range b are compiled in Table 2. The considerable difference in range b implies that the spatial effects of IMs may be influenced by other factors, such as the characteristics of regional site conditions.

Spatial Correlations of V_{S30}

The effects of the regional geology conditions on the spatial IM correlation have been observed (e.g., Jayaram and Baker, 2009) due to the fact that the travel path and frequency contents of earthquake waves are strongly dependent on the regional geological conditions. For this purpose, the V_{S30} values are normalized using the mean and standard deviation of V_{S30} values at all sites of the event. The normalized V_{S30} values have a zero mean and a unit variance to enable direct comparison of the semivariograms. The correlation range of the normalized V_{S30} values can be used to represent the homogeneity of the regional geological conditions, as a larger correlation range of the V_{S30} implies a more uniform geological condition.

However, most V_{S30} values in the strong-motion database are not directly measured but inferred from other geological information. In this situation, a constant V_{S30} value is often assigned to sites that are identified as being in the same site category. The inference results in reduced variability of the V_{S30} distribution compared with the variability of

Table 2
Values of Correlation Range b (km) for CAV, Ia, PGA, and V_{S30}

Earthquake	CAV	Ia	PGA	Original V_{S30}	Redistributed V_{S30}
Chi-Chi	23.9	37.5	42.5	27	26
Northridge	12.6	7.6	7.9	0	0
Parkfield	8.6	8.2	6.4	3.2	3.5
Anza	37.7	35.4	40.2	31.2	20.26
Alum Rock	27.0	30.1	23.5	17.1	14.2
Chino Hills	22.6	16.8	17.5	33	14.5
Tottori	16.2	16.5	20.4	19.1	18.75
Niigata	36.2	34.8	54	26	21.8
Chuetsu	21.0	19.0	25	27	20.8

Table 3
Numbers of Measured and Inferred V_{S30} in Database

Earthquake Name	Number of Recordings	Number of Measured V_{S30}	Number of Inferred V_{S30}	Percentage of Measured Data
Chi-Chi	381	256	125	67.2%
Northridge	149	50	99	33.6%
Parkfield	89	12	77	13.5%
Anza	111	7	104	6.3%
Alum Rock	161	17	144	10.6%
Chino Hills	337	16	321	4.8%
Tottori	112	35	77	31.3%
Niigata	134	19	115	14.2%
Chuetsu	114	18	96	15.8%

measured data for some regions. These inferred V_{S30} values tend to imply a more uniform geological condition than the actual case, and consequently, an artificially increased spatial correlation range of V_{S30} (Baker and Miller, 2011). Table 3 lists the number of inferred and measured V_{S30} values for all events. Even for well-recorded events such as the Northridge earthquake, the number of measured V_{S30} values is less than 40% of the total data. In addition, only approximately 5% of measured V_{S30} values are available for the Anza and Chino Hills earthquakes. Figure 6a,c displays the distribution of original V_{S30} values against the rupture distance for the Chino Hills and Anza earthquakes, where a large number of identical V_{S30} values are inferred. As is explained previously, the situation could greatly increase the range of spatial correlation for V_{S30}.

A simple correction method is implemented to reduce the fake spatial correlations of V_{S30}. First, all V_{S30} data are randomly redistributed around its value using Monte Carlo method by assuming V_{S30} follows a lognormal distribution with a specific standard deviation. The standard deviation of $\ln V_{S30}$ is assumed to be 0.1 for measured data and 0.3 for inferred data, as was provided by Chiou *et al.* (2008). It is worth pointing out that the random redistribution does not consider spatial correlation of V_{S30} residuals (i.e., the difference between the redistributed value and the original value). They are independent even for neighboring sites. In reality, the residuals may also be spatially correlated. However, the uncertainty associated with the V_{S30} is mainly due to measurement errors, human errors, errors associated with the inference of V_{S30} from geological conditions and estimated site categories, and so on. Presently, there is no systematic approach to quantify the spatial correlation of this uncertainty. Our preliminary study also reveals that the spatial correlation of V_{S30} residuals is usually much smaller than that of the normalized V_{S30} values, and they would not significantly affect the latter. Figure 6b,d displays one realization of the randomly redistributed V_{S30} values for each event. Second, the Monte Carlo simulations are repeated for two thousand times. For each realization, the value of correlation range b of the normalized V_{S30} is obtained. Finally, the mean value of

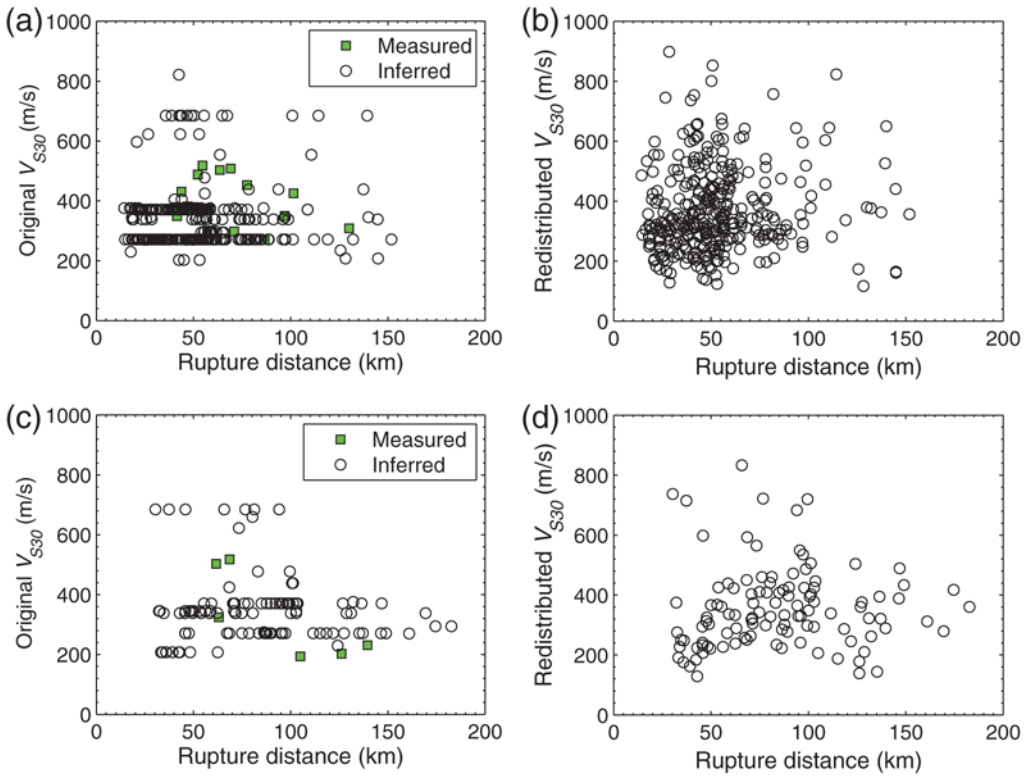

Figure 6. Original and redistributed V_{S30} values versus rupture distances for (a,b) the Chino Hills earthquake, and (c,d) the Anza earthquake. The color version of this figure is available only in the electronic edition.

these computed b values are chosen as the corrected range of (normalized) V_{S30}. After this correction, the correlation ranges of V_{S30} for some events are greatly reduced as expected. For instance, the V_{S30} range for the Chino Hills event is reduced from 33 to 14.7 km after correction. The corrected V_{S30} ranges for nine aforementioned earthquake events are also listed in Table 2.

Predictive Models Based on Regional Site Condition

In this subsection, relationships between the ranges of IMs (CAV, Ia, and PGA) with that of V_{S30} values are examined. The GMPE (Campbell and Bozorgnia, 2008) is used to compute the ranges of PGA as well as SA (T). As is shown in Figure 7, the ranges of IMs are in general positively correlated with the range of V_{S30}. That is, the correlation range of an IM increases if the range of V_{S30} values increases. For practical purpose, three simple fitting functions are provided as follows:

$$\text{CAV}: b_{\text{CAV}} = 10.9 + 0.8 b_{vs} \quad (\sigma = 7.7\ \text{km}, R^2 = 0.46) \tag{12}$$

$$\text{Ia}: b_{\text{Ia}} = 5.8 + 1.1 b_{vs} \quad (\sigma = 7.4\ \text{km}, R^2 = 0.65) \tag{13}$$

$$\text{PGA}: b_{\text{PGA}} = 7.45 \exp(0.07 b_{vs}) \quad (\sigma = 9.2\ \text{km}, R^2 = 0.72), \tag{14}$$

where b_{vs}, b_{CAV}, b_{Ia}, b_{PGA} represent the correlation range (in km) related to the normalized V_{S30} and the normalized IM (CAV, Ia, PGA) residuals, respectively. As is shown in Figure 7, the predictive model shows reasonably good agreement with the empirical data. The standard deviations σ for the predicted models are also provided in equations (12)–(14) to quantify the uncertainty associated with the estimation. R^2 indicates the proportion of data variability that can be explained by the proposed model. It is observed that the ranges of spatial correlations for CAV, Ia, and PGA are similar if the range of V_{S30} is smaller than 15 km (the site condition is relatively heterogeneous), whereas the difference becomes more pronounced when the range of V_{S30} increases (the site condition is more homogeneous). For a relatively homogeneous geological condition (the range of V_{S30} larger than 15 km), PGA appears to have a stronger spatial correlation (i.e., a larger correlation range) than CAV and Ia. In meanwhile, Ia have a slightly larger correlation range than that of CAV, probably due to its stronger correlation with PGA. For example, if the range of the normalized V_{S30} is 20 km, the predicted range for CAV, Ia, and PGA are 26.9, 27.8, and 30.2 km, respectively.

The spatial correlations of spectral accelerations at a spectral period T [termed as SA (T)] are also studied. Figure 8 shows the distribution of the ranges of SA (T) versus these of V_{S30} for spectral periods equal 0.5, 1, 2, and 5 s, respectively. Again, a linear regression can be used to reasonably approximate the relationship between two ranges, expressed as

$$b_{\text{Sa}(T)} = c_1 + c_2 \cdot b_{vs}, \tag{15}$$

where c_1 and c_2 are regression coefficients listed in Table 4. At the spectral period of 10 s, a constant value

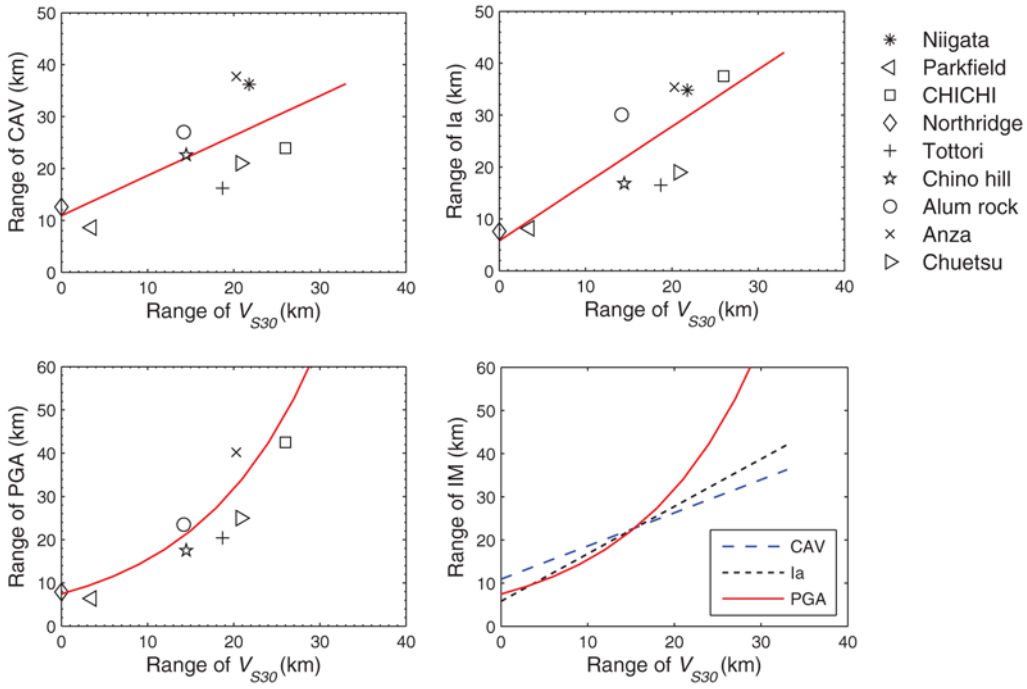

Figure 7. Ranges of CAV, Ia, and PGA versus ranges of normalized V_{S30} values for nine events. The color version of this figure is available only in the electronic edition.

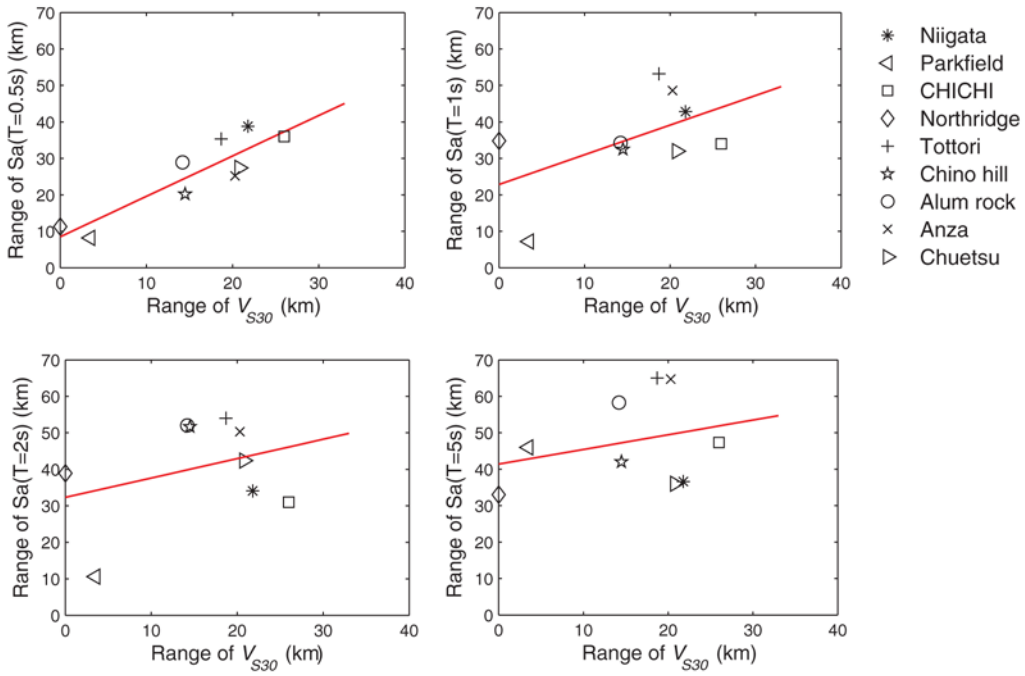

Figure 8. Ranges of spectral accelerations at $T = 0.5$, 1, 2, and 5 s versus the range of normalized V_{S30} values for nine events. The color version of this figure is available only in the electronic edition.

$b_{\mathrm{Sa}(T=10\ \mathrm{s})} = 60$ km can be assigned to all cases. Linear interpolation between these $b_{\mathrm{Sa}(T)}$ values can be employed for periods other than 0, 0.5, 1, 2, 5, or 10 s. The results indicate that, as the spectral period T increases, the influence of the regional site condition (represented by the range of V_{S30} values) on the spatial ranges of spectral accelerations becomes weaker. The observation is consistent with the results by other scholars (e.g., Jayaram and Baker, 2009). This is not unexpected because the high-frequency component of the earthquake waves is sensitive to the heterogeneities of the regional site conditions during wave propagation. However, the long-period component of the earthquake waves is less affected by the regional site condition, especially for the spectral accelerations at periods longer than 2 s.

Comparisons and Discussions

In previous sections, particular GMPEs are chosen to compute the correlation range of the IMs, so the results might be influenced by the choice of GMPEs. In this study, other GMPEs available in literature are also used to compute the empirical semivariograms. For Ia, the GMPE by Travasarou and Bray (2003; termed as TB03) and Piggott and Stafford (2012; termed as PS12) are chosen and compared with Campbell and Bozorgnia (2012; termed as CB12). For spectral accelerations, three other NGA GMPEs by Boore and Atkinson (2008; termed as BA08), Chiou and Youngs (2008; termed as CY08), and Abrahamson and Silva (2008; termed as AS08) are used and compared with Campbell and Bozorgnia (2008; termed as CB08). Figure 9 shows the obtained empirical semivariograms computed by these GMPEs for the Chi-Chi earthquake for Ia and PGA, respectively. Although not presented here, the results from an alternative GMPE for CAV (Du and Wang, 2012) are also compared with that from Campbell and Bozorgnia (2010) model. In general, very similar results can be derived by using different GMPEs, indicating that obtained spatial correlations are not dependent on specific GMPEs.

Compared with a previous study (Piggott and Stafford, 2012) on the spatial correlation for Ia using only the Chi-Chi and Northridge earthquakes, the present study shows comparable results. For instance, the reported correlation range

Table 4
Model Coefficients of Predictive Equation for the Correlation Range of SA (T)

Model Coefficients	SA ($T = 0.2$ s)	SA ($T = 0.5$ s)	SA ($T = 1$ s)	SA ($T = 2$ s)	SA ($T = 5$ s)
c_1	4.4	8.5	22.8	32.3	41.4
c_2	1.1	1.1	0.8	0.5	0.4
σ	8	5.3	11.8	14.1	12.6
R^2	0.63	0.79	0.29	0.11	0.08

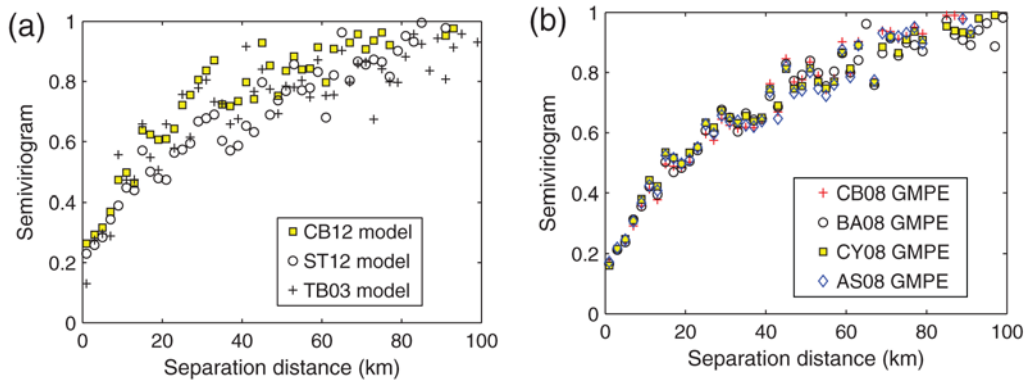

Figure 9. Empirical semivariograms computed by various GMPEs for the Chi-Chi earthquake for Ia and PGA, respectively. The color version of this figure is available only in the electronic edition.

for the Chi-Chi earthquake is about 31 km in Piggott and Stafford (2012) compared with 38 km obtained in this study. The difference can be attributed to the epistemic uncertainty by using different GMPEs and fitting technique, and it is insignificant for application. There is no published result for spatial correlation of CAV available in the literature for comparison.

Furthermore, the spatial correlations of spectral accelerations are compared with a previous study. Figure 10a plots the range of empirical semivariograms versus periods for nine events, and Figure 10b plots the predicted values of correlation ranges computed by equation (13) versus periods, assuming b_{vs} is 1, 10, and 20 km, respectively. In comparison, the correlation ranges of SA (T) are specified as a function of the spectral periods T in Jayaram and Baker (2009):

$$
\begin{aligned}
b_{\mathrm{Sa}(T)} &= 8.5 + 17.2T, \quad \text{if } T < 1 \text{ s} \\
&\quad \text{and homogeneous site conditions,} \\
b_{\mathrm{Sa}(T)} &= 40.7 - 15T, \quad \text{if } T < 1 \text{ s} \\
&\quad \text{and heterogeneous site conditions.} \\
b_{\mathrm{Sa}(T)} &= 22 + 3.7T, \quad \text{if } T > 1 \text{ s} \\
&\quad \text{and applicable for all site conditions.}
\end{aligned} \tag{16}
$$

Their results (termed as JB09) are compared with the present study in Figure 10b. Similar trend can be observed for two models, except that the present study gives a slightly larger range compared with JB09's model at periods greater than 2 s. The discrepancy may be due to different databases used in each study: this study included an addition of three recent earthquakes that occurred in Japan, and they were not considered in the previous study (Jayaram and Baker, 2009).

Finally, an illustrative example is provided to highlight the importance of the spatial correlation of IMs in seismic-hazard analysis. Considering a hypothetical region 40 km × 40 km in size located near a point source at the origin, the area is divided into cells of 1 km × 1 km in size. Assuming that the style of faulting is a reverse fault, and the regional average site conditions are deep soils (the averaged V_{S30} = 240 m/s). For illustration purposes, a scenario earthquake with moment magnitude of 7 and corresponding annual rate of exceedance $\lambda_M = 1/500$ alone is considered, for which the predicted median and the standard deviation of Ia values can be obtained by GMPEs (i.e., Campbell and Bozorgnia, 2012) at each site location. Based on equation (13), the correlation range of the intra-event residuals for Ia can be calculated given a certain range of V_{S30}. Copula functions (Nelson, 2006) are used to generate the intra-event residuals,

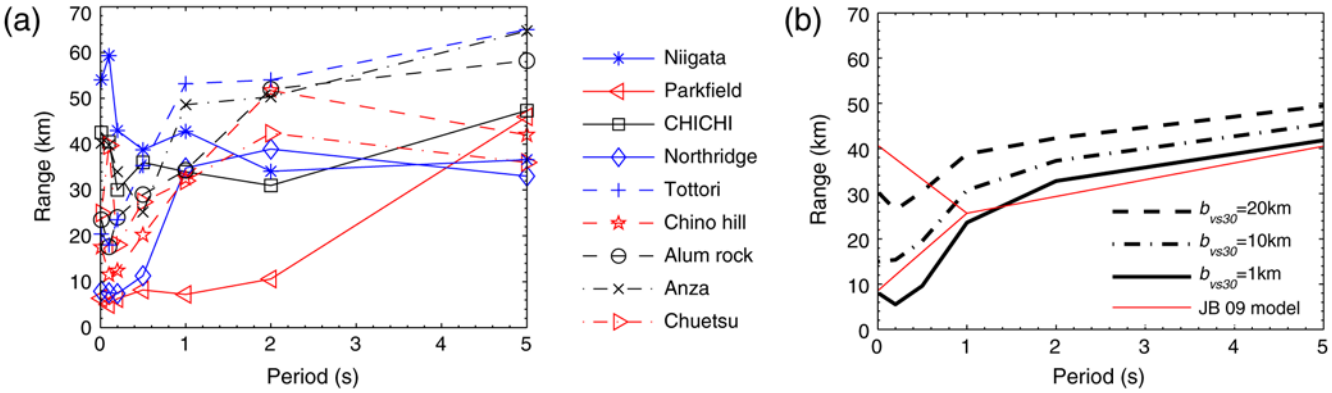

Figure 10. (a) Ranges of spectral accelerations versus periods for nine events; (b) comparison of predicted ranges by assuming varying b_{vs} values with results from JB09 (Jayaram and Baker, 2009). The color version of this figure is available only in the electronic edition.

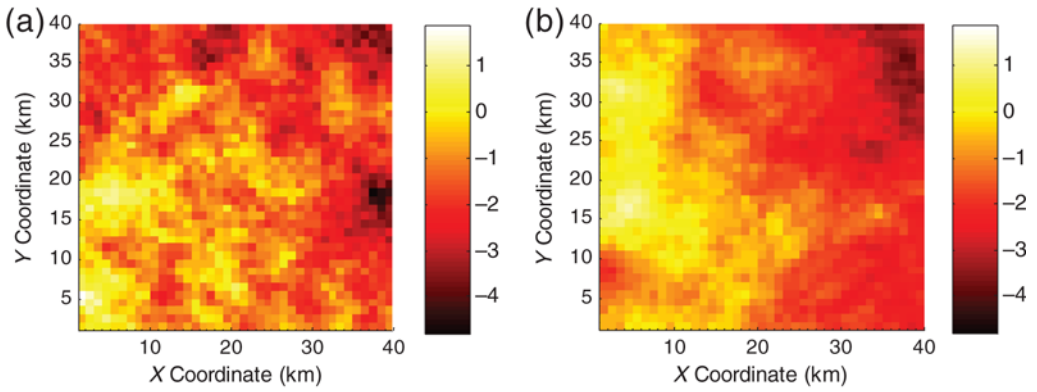

Figure 11. Examples of realized spatially correlated field of Ia (in natural log scale, unit of m/s). (a) correlation range 10 km, and (b) correlation range 40 km. The color version of this figure is available only in the electronic edition.

which are normally distributed with a zero mean and a standard deviation specified by GMPEs at each site, whereas at the same time, they are spatially correlated according to the given spatial correlation. Detailed procedures can be found in the references (e.g., Johnson, 1987; Sokolov and Wenzel, 2011a). Two Monte Carlo realizations of Ia values (in natural log scale) are shown in Figure 11 by assuming the correlation ranges of 10 and 40 km, respectively. As is expected, a large correlation range corresponds to a more uniform spatial distribution of Ia, as shown in Figure 11b.

The Monte Carlo method is applied to generate 10,000 realizations of spatially correlated Ia values over the region. Given a specified value of Ia (denoted as Ia*) and its exceedance area ratioAR^* (defined as ratio of the areas for which the Ia values exceed the specified Ia* value against the total area of the region), the annual rate of exceedance can be computed from the Monte Carlo realizations as

$$\lambda = \lambda_M \cdot P(\text{Ia} > \text{Ia}^* \,\&\, AR > AR^*). \qquad (17)$$

Several special cases are considered for comparison by assuming that the intra-event residuals of Ia are independently distributed without spatial correlation (correlation range is zero), or spatially correlated with the correlation range of 10 and 40 km, respectively; or perfectly correlated (correlation range is infinite such that the residuals are identical over the region). Figure 12 shows the annual exceedance curve for Ia with the exceedance area AR^* of 5% and 25%, respectively. The results imply that, considering spatial effect does not always increase the designed IM value, which actually depends on the specific seismic-hazard level as well as the exceeded ratio being considered. For example, given an annual rate of exceedance as 10^{-4}, the predicted Ia values for 25% area of exceedance is 1.06 m/s for an uncorrelated distribution, and 1.45 m/s for a correlated distribution with a range of 40 km. It is obvious that the hazard level will be underestimated if no spatial correlation is considered in this case. However, if a relative low seismic-hazard level (an annual rate of exceedance value of 10^{-3}) is considered, the predicted Ia values for 25% area of exceedance for uncorrelated distribution and correlated distribution with a range of 40 km will be 0.65 m/s and 0.6 m/s, respectively. Hence, the hazard is slightly overestimated for the case of uncorrelated spatial distribution. Nevertheless, the spatial correlation does increase the probability of rare occurrence, especially when the mean annual rate of exceedance is smaller than 10^{-4} in this case study. The results corroborate some recent studies, for example, Esposito and Iervolino (2011). These annual curves can be further used for the damage and loss estimation

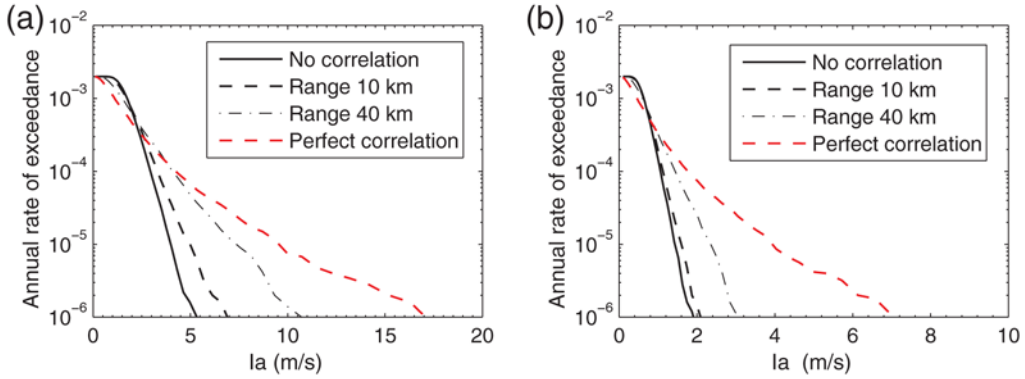

Figure 12. Annual exceedance hazard curves for Ia considering different spatial correlations for (a) exceedance area ratioAR^* as 5%, and (b) AR^* as 25%. The color version of this figure is available only in the electronic edition.

of spatially distributed infrastructure, such as Sokolov and Wenzel (2011b).

Conclusions

In this study, the spatial correlations of CAV, Ia, and SA are obtained by semi-empirical estimation using a comprehensive database containing more than 1500 recorded ground motions from nine recent earthquakes. The correlation range of normalized V_{S30} values among the sites has been found to be a good indicator to represent the homogeneity of the regional site conditions, and it is closely related to the correlation ranges of IMs. In general, an IM recorded from a more homogenous regional site condition tend to have a larger spatial correlation range than that from a heterogeneous site condition. In addition, the range of spatial correlation of SA generally increases as the spectral period increases. Methods for data correction are proposed to reduce the artificial correlation arising from biased distance scaling of the GMPEs and the estimation of V_{S30} values.

Simple predictive equations are proposed in this study to quantify the spatial correlations of CAV, Ia, and SA based on regional site conditions. The model can be easily used in seismic-hazard analysis or lose estimate by generating spatially distributed IMs following the specified range of correlation. An illustrative example is provided to highlight the importance of the spatial correlation.

Data and Resources

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

Acknowledgments

The authors acknowledge financial support from Hong Kong Research Grants Council Grant Number 620311. Brian Chiou provided useful information for this study. Constructive comments from the Associate Editor and an anonymous reviewer were very helpful for improving the manuscript.

REFERENCES

Abrahamson, N. A., and W. J. Silva (2008). Summary of the Abrahamson & Silva NGA ground-motion relations, *Earthq. Spectra* **24,** no. 1, 67–97.

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

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

Baker, J. W., and M. K. Miller (2011). Effects of earthquake source geometry and site conditions on spatial correlation of earthquake ground motion hazard, in *4th IASPEI/IAEE International Symposium*, University of California, Santa Barbara, California, 23–26 August 2011.

Boore, D. M., and G. M. Atkinson (2008). Ground-motion prediction equations for the average horizontal component of PGA, PGV and 5%-damped PSA at spectral periods between 0.01 s and 10 s, *Earthq. Spectra* **24,** no. 1, 99–138.

Boore, D. M., J. F. Gibbs, W. B. Joyner, J. C. Tinsley, and D. J. Ponti (2003). 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, *Bull. Seismol. Soc. Am.* **93,** no. 6, 2737–2751.

Brillinger, D. R., and H. K. Preisler (1984). An exploratory analysis of the Joyner-Boore attenuation data, *Bull. Seismol. Soc. Am.* **74,** no. 4, 1441–1450.

Brillinger, D. R., and H. K. Preisler (1985). Further analysis of the Joyner-Boore attenuation data, *Bull. Seismol. Soc. Am.* **75,** no. 2, 611–614.

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

Campbell, K. W., and Y. Bozorgnia (2010). A ground motion prediction equation for the horizontal component of cumulative absolute velocity (CAV) based on the PEER-NGA strong motion database, *Earthq. Spectra* **26,** no. 3, 635–650.

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

Chiou, B., and R. R. Youngs (2008). An NGA model for the average horizontal component of peak ground motion and response spectra, *Earthq. Spectra* **24,** no. 1, 173–215.

Chiou, B. S. J., R. Darragh, N. Gregor, and W. Silva (2008). NGA project strong-motion database, *Earthq. Spectra* **24,** 23–44.

Cressie, N. (1993). *Statistics for Spatial Data*, Revised edition, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc., New York, 928 pp.

Cressie, N., and D. M. Hawkins (1980). Robust estimation of variogram, *Math. Geol.* **12,** no. 2, 115–125.

Du, W., and G. Wang (2012). A simple ground-motion prediction model for cumulative absolute velocity and model validation, *Earthq. Eng. Struct Dynam.*, doi: 10.1002/eqe.2266.

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

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

Goda, K., and G. M. Atkinson (2010). Intraevent spatial correlation of ground-motion parameters using SK-net data, *Bull. Seismol. Soc. Am.* **100,** no. 6, 3055–3067.

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

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

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

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

Jibson, R. W. (2007). Regression models for estimating coseismic landslide displacement, *Eng. Geol.* **91,** 209–218.

Johnson, M. E. (1987). *Multivariate Statistical Simulation*, Wiley Series in Probability and Mathematical Statistics, Wiley and Sons, Inc., New York, 240 pp.

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

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

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

Kayen, R. E., and J. K. Mitchell (1997). Assessment of liquefaction potential during earthquakes by Arias intensity, *J. Geotech. Geoenvir. Eng.* **123,** 1162–1174.

Kramer, S. L., and R. A. Mitchell (2006). Ground motion intensity measures for liquefaction hazard evaluation, *Earthq. Spectra* **22,** no. 2, 413–438.

Kuo, C. H., K. L. Wen, H. H. Hsieh, C. M. Lin, T. M. Chang, and K. W. Kuo (2012). Site classification and Vs30 estimation of free-field TSMIP stations using the logging data of EGDT, *Eng. Geol.* **129–130,** 68–75.

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

Liyanapathirana, D. S., and H. G. Poulos (2004). Assessment of soil liquefaction incorporating earthquake characteristics, *Int. J. Soil Dynam. Earthq. Eng.* **24,** 867–875.

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

Park, J., P. Bazzurro, and J. W. Baker (2007). Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation, in *Applications of Statistics and Probability in Civil Engineering*, J. Kanda, T. Takada, and H. Furuta (Editors), Taylor & Francis Group, London, 1–8.

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

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

Sokolov, V., and F. Wenzel (2011b). Influence of ground-motion correlation on probabilistic assessments of seismic hazard and loss: Sensitivity analysis, *Bull. Earthq. Eng.* **9,** 1339–1360.

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

Travasarou, T., and J. D. Bray (2003). Empirical attenuation relationship for Arias intensity, *Earthq. Eng. Struct. Dynam.* **32,** 1133–1155.

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

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

Manuscript received 30 May 2012