Bulletin of the Seismological Society of America, Vol. 101, No. 5, pp. 2532–2541, October 2011, doi: 10.1785/0120110117

Short Note

PGA and PGV Spatial Correlation Models Based on European Multievent Datasets

by Simona Esposito and Iunio Iervolino

Abstract Spatial modeling of ground motion intensity measures (IMs) is required for risk assessment of spatially distributed engineering systems. For example, when a lifeline system is of concern, classical site-specific hazard tools, which treat IMs at different locations independently, may not be adequate to accurately assess the seismic risk. In fact, in this case, modeling of ground motion as a random field is required; it basically consists of assigning a correlation structure to the IM of interest. This work focuses on semiempirical estimation of the correlation coefficient, as a function of intersite separation distance, between residuals with respect to ground motion prediction equations (GMPEs) of horizontal peak ground acceleration (PGA) and peak ground velocity (PGV). In particular, subsets of the European Strong-Motion Database (ESD) and the Italian Accelerometric Archive (ITACA) were employed to evaluate the intraevent residual correlation based on multiple earthquakes, considering different GMPEs fitted to the same records. The analyses were carried out through geostatistical tools, which enabled results to be found that are generally consistent between the two datasets. Correlation for PGV appears to attenuate more gradually with respect to PGA. In order to better understand the dependency of the results on the adopted estimation approach and dataset, some aspects related to the working hypotheses are critically discussed. Finally, estimated correlation models are used to develop illustrative applications of regional probabilistic seismic-hazard analysis.

Introduction

Seismic-risk analysis of distributed systems and infrastructures requires a different approach with respect to the one commonly used for site-specific structures. In fact, systemic seismic performance may be conditional upon the behavior of many different components, each of which may respond differently to the input ground motion in the region where the system is deployed. In the seismic-risk assessment of such systems, one of the key issues, at least on the demand side, is to account for the existence of a spatial statistical correlation between ground motion intensity measures (IMs).

Traditionally, ground motion is modeled, for engineering purposes, via ground motion prediction equations (GMPEs), which provide probabilistic distribution of the chosen IM, conditional on earthquake magnitude, source-to-site distance, and other parameters such as local geological conditions. GMPEs are obtained by regression of recorded data from historical events. The model's residual is usually expressed as the sum of two components: an interevent term, which is constant for each earthquake (common for all sites) and represents average source effects not explicitly appearing in the model covariates, and an intraevent term representing site-to-site variability of the IM (Strasser *et al.*, 2009). Boore *et al.* (2003) demonstrated that intraevent residuals, for example those referring to peak ground acceleration (PGA), are spatially correlated[1]. Therefore, IMs at different sites are correlated because of both inter- and intraevent residuals, and it is important to account for these dependencies in seismic-risk assessment when a region is of concern (Crowley and Bommer, 2006; Park *et al.*, 2007; Goda & Hong, 2008b, Crowley *et al.*, 2008).

Several correlation models available in the literature depend uniquely on intersite separation distance. Most of the studies are based on dense observations of single events (e.g., Boore *et al.*, 2003; Wang and Takada, 2005; Jayaram and Baker, 2009) from different major earthquakes outside

[1]This kind of spatial correlation of ground motion consists of similarity between IMs (e.g., peak values of time history) observed at different sites within the same event. It is also worth mentioning here the coherency of ground-motion signals, which represents the similarity of ground motion in the frequency domain and describes the degree of positive or negative correlation between amplitudes and phase angles of two time histories at each of their component frequencies (e.g., Zerva and Zervas, 2002).

Europe, such as Northridge (1994) or Chi-Chi (1999). A few works have, instead, combined data from multiple events to obtain a unique estimate of correlation (e.g., Goda and Hong; 2008a, Goda and Atkinson, 2009; Goda and Atkinson, 2010; Sokolov *et al.*, 2010).

Different authors, for a given IM, provide different distance limits for correlation to disappear (i.e., distance beyond which IMs may be considered uncorrelated), and this is supposed to depend on the dataset considered, the GMPE chosen to compute residuals, and the working assumptions of the estimation. For example, Goda and Atkinson (2009) investigated the effects of earthquake types (i.e., shallow and deep events) on correlation using datasets from K-NET and KiK-net Japanese strong-motion networks without finding any significant dependency. On the other hand, Sokolov *et al.* (2010), starting from the strong-motion database of Taiwan Strong Motion Instrumentation Program (TSMIP) network, estimated correlation for various areas, site classes, and geological structures, asserting that a single generalized spatial model may not be adequate for all of Taiwan territory.

In some cases (e.g., Wang and Takada, 2005; Jayaram and Baker, 2009), existing GMPEs are used, while in others (e.g., Goda and Hong, 2008a; Goda and Atkinson, 2009; Sokolov *et al.*, 2010), *ad hoc* fit on the chosen dataset is adopted. Generally, regression analysis used to develop prediction equations does not incorporate the correlation structure of residuals as a hypothesis. Hong *et al.* (2009) and Jayaram and Baker (2010) evaluated the influence of considering the correlation in fitting a GMPE, finding a minor influence on regression coefficients and a more significant effect on the variance components.

Goda and Atkinson (2010) investigated the influence of the estimation approach, emphasizing its importance when residuals are strongly correlated.

In Figure 1, several models for PGA and peak ground velocity (PGV) as mentioned in the preceding paragraph, are shown; the correlation coefficient is expressed by equation (1), where a, b, and c are the model parameters (to follow), and h is the intersite separation distance (in kilometers):

$$\rho(h) = \max\{(1-c) + c \cdot e^{-a \cdot \mathbf{h}^b}, 0\}. \tag{1}$$

In this paper, the evaluation of the spatial correlation of PGA and PGV intraevent residuals is carried out using the European Strong-Motion Database (ESD) and the Italian Accelerometric Archive (ITACA). Because each earthquake in the chosen datasets is characterized by a relatively small number of records, which may be insufficient to evaluate correlation, data from multiple events are pooled to fit a unique model.

The GMPEs, with respect to which residuals are computed, are those of Akkar and Bommer (2010) for ESD and Bindi *et al.* (2010) for ITACA. Subsets of the same records used to estimate the considered GMPEs are used to estimate intraevent spatial correlation models.

The analysis of correlation was performed through geostatistical tools, and in order to better understand the dependency of the results on the approach adopted, some aspects related to the working hypotheses are discussed. Finally, developed correlation models are employed within the framework of regional seismic-hazard assessment to compare with the case when spatial correlation is not considered.

Semiempirical Modeling of Spatial Correlation

GMPEs model the logs of ground-motion intensities and related heterogeneity at a site p due to earthquake j as in equation (2):

$$\log Y_{pj} = \overline{\log Y_{pj}}(M, R, \boldsymbol{\theta}) + \eta_j + \varepsilon_{pj}. \tag{2}$$

Y_{pj} denotes the IM of interest; $\overline{\log Y_{pj}}(M, R, \boldsymbol{\theta})$ is the mean of the logs conditional on parameters such as magnitude (M), source-to-site distance (R), and others ($\boldsymbol{\theta}$); η_j denotes the interevent residual, which is a constant term for all sites in a given earthquake and represents a systematic deviation

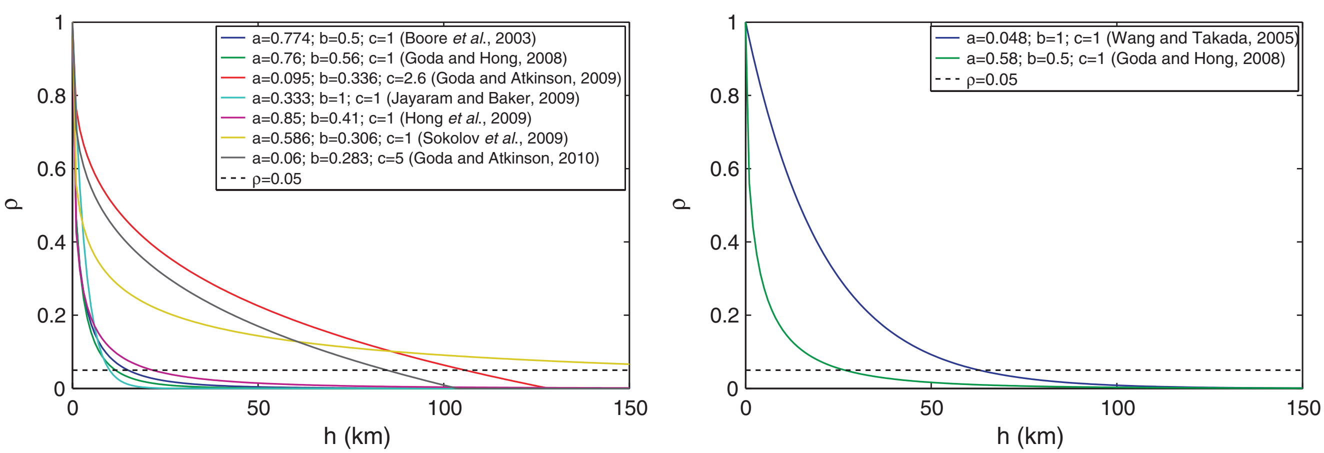


Figure 1. Some correlation models available in literature for PGA (left) and PGV (right); a, b, and c are the model parameters in equation (1). The black dotted line intersects the curves at the distance at which the correlation is conventionally considered as almost lost, which is the correlation coefficient equal to 0.05.

from the mean of the specific seismic event; and ε_{pj} is the intraevent variability of ground motion. Residuals ε_{pj} and η_j are usually assumed to be independent random variables, normally distributed with zero mean and standard deviation σ_{intra} and σ_{inter}, respectively. Then, $\log Y_{pj}$ is modeled as a normal random variable with mean $\overline{\log Y_{pj}}(M, R, \boldsymbol{\theta})$ and standard deviation σ_T, where $\sigma_T^2 = \sigma_{\text{intra}}^2 + \sigma_{\text{inter}}^2$. Appropriately plugging this distribution into the probabilistic seismic-hazard analysis leads to the distribution of the IM at the site of interest (McGuire, 2004).

If the hazard assessment at two or more sites is of concern, the joint probability density function (PDF) for the IMs at all locations is required. A simple way to model for the joint PDF of the IM, conditional on the GMPE covariates, is the multivariate normal (e.g., Jayaram and Baker, 2008). It is assumed that the logs of IM form a Gaussian random field (GRF), defined as a set of random variables $\log[IM(\mathbf{u})]$, one for each site $\mathbf{u}$ in the study area $S \in R^2$, $\{\log[IM(\mathbf{u})], \forall \mathbf{u} \in S\}$. For any set of n sites $\mathbf{u}_p$, $p = 1, \ldots, n$, corresponds to a vector of n random variables that is characterized by the covariance matrix, Σ, as in equation (3), where the first term produces perfectly correlated interevent residuals (Malhotra, 2008), while the second term (symmetrical) produces partially correlated intraevent residuals:

$$\Sigma = \sigma_{\text{inter}}^2 \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} + \sigma_{\text{intra}}^2 \cdot \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & 1 \end{bmatrix}. \qquad (3)$$

In equation (3), the correlation is heterogeneous as it depends on the pairs of sites considered, and the intraevent variance is homoscedastic as it is constant for all sites (this is assumed in most of GMPEs, although some studies have found dependence of intraevent variability on distance, magnitude, and nonlinear site effects; Strasser *et al.*, 2009). If the spatial correlation for intraevent residuals is a function of the relative location of sites, it becomes as in equation (4), where p and q are two locations at the ends of $\mathbf{h_{pq}}$ (the separation vector between the two sites):

$$\rho_{pq} = \rho(\mathbf{h_{pq}}). \qquad (4)$$

To briefly review the features of a GRF, let $\mathbf{u_p} \in R^2$ be the generic site in a two-dimensional Euclidian space and suppose that the intraevent residual in a specific earthquake, $\varepsilon_j(\mathbf{u}_p) = \varepsilon_{pj}$, is a GRF in a domain $S \in R^2$ (the region of interest). The GRF is fully described by the mean $E[\varepsilon_j(\mathbf{u})]$ for each site and the covariance C_j between two generic locations $\mathbf{u_p}$ and $\mathbf{u_q}$ defined in equation (5):

$$C_j(\mathbf{u}_p, \mathbf{u}_q) = E[\varepsilon_j(\mathbf{u}_p) \cdot \varepsilon_j(\mathbf{u}_q)] - E[\varepsilon_j(\mathbf{u}_p)] \cdot E[\varepsilon_j(\mathbf{u}_q)]. \qquad (5)$$

Under the hypothesis of second-order stationarity of the GRF (Goovartes, 1997), the mean is constant and the covariance is location-independent as defined in equation (6):

$$C_j(\mathbf{h}) = E[\varepsilon_j(\mathbf{u}) \cdot \varepsilon_j(\mathbf{u} + \mathbf{h})] - E[\varepsilon_j(\mathbf{u})] \cdot E[\varepsilon_j(\mathbf{u} + \mathbf{h})]. \qquad (6)$$

In this case, the two points' statistics depend only on the separation vector $\mathbf{h}$, and the reference to a particular location $\mathbf{u}$ can be dropped. Here, it is assumed that intraevent residuals may be modeled as a stationary GRF, and all data available from different earthquakes and regions, therefore deemed homogeneous, are used to fit a unique model.

If the GRF is isotropic, correlation depends only on the separation distance $h = \|\mathbf{h}\|$. An anisotropic random field implies, instead, the possibility of having a spatial variability that depends on the direction considered. Past research has shown that the hypothesis of isotropic random field, which is also retained herein, is reasonable (Wang and Takada, 2005; Jayaram and Baker, 2009).

Geostatistical Analysis of Intraevent Residuals

A common tool to quantify spatial variability of random fields is the semivariogram, $\gamma_j(h)$. It measures the average dissimilarity between georeferenced data, and it is used to model the covariance structure of GRF through suitable functions. Under the hypothesis of second-order stationarity and isotropy it is defined as in equation (7):

$$\gamma_j(h) = \frac{1}{2} \mathrm{Var}[\varepsilon_j(\mathbf{u} + h) - \varepsilon_j(\mathbf{u})] = \mathrm{Var}[\varepsilon_j(\mathbf{u})] - C_j(h). \qquad (7)$$

Therefore, for an isotropic and homogenous random field, considering that for $h \to 0$, $C_j(h) = \mathrm{Var}(\varepsilon_j)$, the semivariogram results as in equation (8),

$$\gamma_j(h) = C_j(0) - C_j(h) = \mathrm{Var}(\varepsilon_j) \cdot [1 - \rho_j(h)], \qquad (8)$$

where $\rho_j(h)$ denotes the spatial correlation coefficient between $\varepsilon_j(\mathbf{u} + h)$ and $\varepsilon_j(\mathbf{u})$; see Cressie (1993) for more details. In fact, estimation of correlation usually develops in three steps: (1) computing the empirical semivariogram (assuming a common semivariogram for different events, that is, invariant through earthquakes, allows us to neglect the subscript j in the following equations), (2) choosing a functional form, and (3) estimating the correlation parameters by fitting the empirical data with the functional model.

Empirical semivariograms are computed as a function of site-to-site separation distance, with different possible estimators. The classical estimator is the method-of-moments (Matheron, 1962) which is defined for an isotropic random field in equation (9),

$$\hat{\gamma}(h) = \frac{1}{2 \cdot |N(h)|} \cdot \sum_{N(h)} [\varepsilon(\mathbf{u} + h) - \varepsilon(\mathbf{u})]^2, \tag{9}$$

where $N(h)$ is the set of pairs of sites separated by the same distance h, and $|N(h)|$ is the cardinal of $N(h)$. To compute the semivariogram, it may be useful, when dealing with earthquake records, to define tolerance bins around each possible h value. The selection of distance bins has important effects: if its size is too large, correlation at short distances may be masked; conversely, if it is too small, empty bins, or bins with samples small in size, may impair the estimate. A rule of thumb is to choose the maximum bin size as a half of the maximum distance between sites in the dataset and to set the number of bins so that there are at least 30 pairs per bin (Journel and Huijbregts, 1978).

The method-of-moments estimator is unbiased; however, it can be badly affected by atypical observations (Cressie, 1993). Therefore, Cressie and Hawkins (1980) proposed a more robust estimator (less sensitive to outliers), as in equation (10):

$$\hat{\gamma}(h) = \frac{1}{2} \left\{ \left[\frac{1}{|N(h)|} \sum_{N(h)} |\varepsilon(\mathbf{u} + h) - \varepsilon(\mathbf{u})|^{0.5} \right]^4 \Big/ \left(0.457 + \frac{0.494}{|N(h)|} \right) \right\}. \tag{10}$$

The fitting analytical model, under stationary and isotropic hypotheses, may be of different kinds, for example, exponential, Gaussian, or spherical (Goovartes, 1997). In particular, the exponential model, which is the most common one, is described in equation (11), where c_0 is defined as the nugget and represents the limit value of the semivariogram when h is zero because of variations at distances smaller than the sampling interval and measurement errors, which cause a discontinuity at the origin (Matheron, 1962); c_e is the sill, or the population variance of the random field (Barnes, 1991); and b is the range defined as the intersite distance at which $\gamma(h)$ equals the sill. For the exponential model, the sill is asymptotic, and it is possible to define a practical range as the separation distance at which $\gamma(h)$ equals 95% of the sill:

$$\gamma(h) = c_0 + c_e \cdot (1 - e^{-3 \cdot h/b}). \tag{11}$$

The goodness of fit of a model can be determined via several criteria that have been proposed in the geostatistical literature. Studies dealing with earthquake data sometimes use visual or trial-and-error approaches in order to appropriately model the semivariogram structure at short site-to-site distances (Jayaram and Baker, 2009). In this work, experimental semivariograms are fitted visually, although using the least-squares estimation as a starting point (described in the PGA and PGV Correlations from ESD and ITACA section).

Datasets

The estimation of the correlation starts from the characterization of residuals of empirical data with respect to a GMPE. To this aim, subsets of the ESD and the ITACA datasets were considered (see the Data and Resources section). The ESD dataset is comprised of 480 records from 87 events recorded between 1973 and 2003 and characterized by moment magnitudes from 5 to 7.6 and the closest horizontal distance to the vertical projection of the rupture (i.e., Joyner–Boore distance, R_{jb}) from 0 to 100 km. The number of considered recordings for the ITACA subset is 1112 from 162 events over the 4–6.9 magnitude range (moment or local), and R_{jb} up to 196 km. Characteristics of the datasets, with respect to explanatory variables of the considered prediction equations (magnitude, distance, and local site conditions) are shown in Figure 2. ESD is a smaller database of stronger and closer-to-the-source records from European events, while ITACA is a denser dataset of Italian earthquakes within a lower magnitude range and a broader distance range. A limited number of records (150 from 19 events) are in common between the two sets of data. In Figure 3, the distributions of data pairs as a function of separation distance bins (4-km width for ESD and 1-km width for ITACA) are also shown.

Estimating Correlation on Multievent Data

To compute the empirical semivariogram, normalized intraevent residuals are obtained for a single earthquake j and a generic site p as $\varepsilon^*_{pj} = \varepsilon_{pj}/\sigma_p$ where σ_p is the standard deviation of the intraevent residual at the site p (in the study, the intraevent standard deviation is common for all sites consistent with GMPEs used to compute residuals). In this case, equation (8) becomes equation (12), where the superscript represents an empirical estimate:

$$\hat{\gamma}(h) = 1 - \hat{\rho}(h). \tag{12}$$

Equation (12) shows that standardization enables us to not estimate the sill, as it should be equal to one. This applies if standardization is carried out using the true population's variance. With earthquake data, the sample variance or the standard deviation provided by the GMPE can be used to obtain standardized residuals[2]. Another option is to use the sample variance as an estimate of the true variance (e.g., Jayaram and Baker, 2010). Goda and Atkinson (2010) used the intraevent standard deviation inferred from the large-separation-distance plateau of the semivariogram, assuming that at those distances, residuals are not correlated. In this work, the variance provided by the GMPEs was preferred. In fact, evaluation of possible alternatives for standardization leads to results that seem to be not significantly affected by one choice with respect to another (as discussed in the Influence of Standardization section).

Because geostatistical estimation needs a relatively large number of data to model the semivariogram (i.e., many records have more than 30 pairs in each h bin), which are not available for individual events in the chosen datasets, all data available

[2]It is usual to use the sample variance as an estimator of the sill for the experimental semivariogram, but this may be improper in some circumstances; see Barnes (1991) for a discussion.

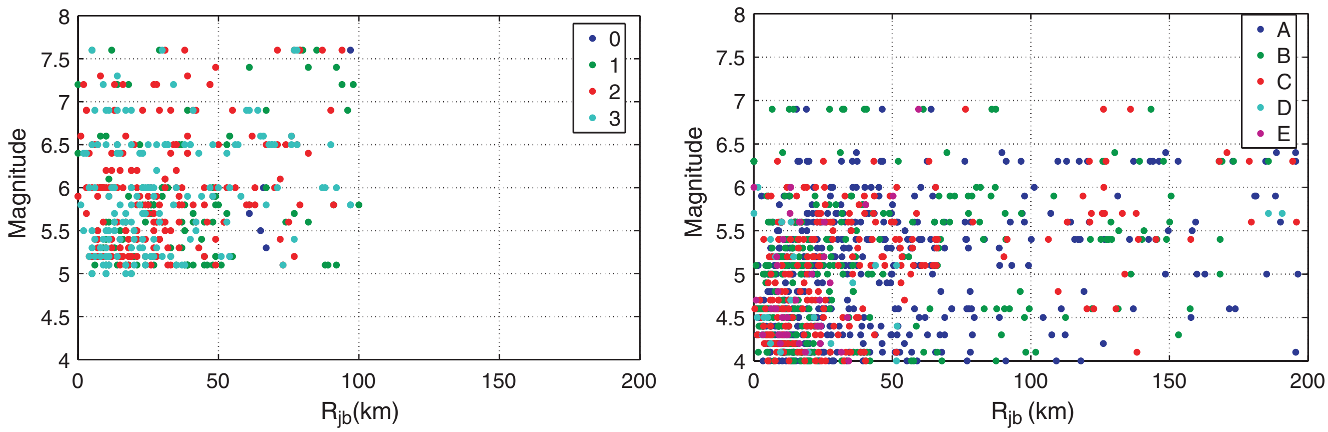


Figure 2. Left: The ESD subsets with respect to M, R_{jb}, and local site conditions: rock (3), stiff soil (2), soft soil (1), and very soft soil (0). Right: The ITACA strong-motion subsets with respect to M, R_{jb}, and local site conditions according to Eurocode 8 (2004).

from multiple events (and regions) are used herein to fit a unique correlation model. The same isotropic semivariogram with the same parameters for all earthquakes is assumed.

The experimental semivariogram becomes that of equation (13), where n_j is the number of records for the j-th event, and $|N(h)|$ is the number of pairs in the specific h bin. Equation (14) shows how individual events (k in total) are kept separated in computing the empirical semivariogram. In fact, the differences of residuals of equation (13) are computed only between pairs of residuals (standardized with the common standard deviation from the GMPE) from the same earthquake; then, differences from different earthquakes are pooled. This process is visually sketched in Figure 4, from which it is possible to note that the empirical semivariogram point at h_1 is not the average of experimental semivariograms from different earthquakes, as $|N(h)|$ is the number of pairs in the specific h bin from all earthquakes:

$$\hat{\gamma}(h) = \frac{1}{2 \cdot |N(h)|} \sum_{N(h)} [\varepsilon^*_{pj} - \varepsilon^*_{qj}]^2, \tag{13}$$

$$N(h) = \{(j, \varepsilon^*_{pj}, \varepsilon^*_{qj}) : \|\varepsilon^*_{pj} - \varepsilon^*_{qj}\| = h; \; p, q = 1, \ldots n_j; j = 1, \ldots k\}. \tag{14}$$

PGA and PGV Correlations from ESD and ITACA

For the European subset, each bin has a 4-km width, as this also allows there to be at least 30 pairs per bin and no empty bins until half of the maximum distance between pairs in the dataset. Both estimators (classical and robust) were

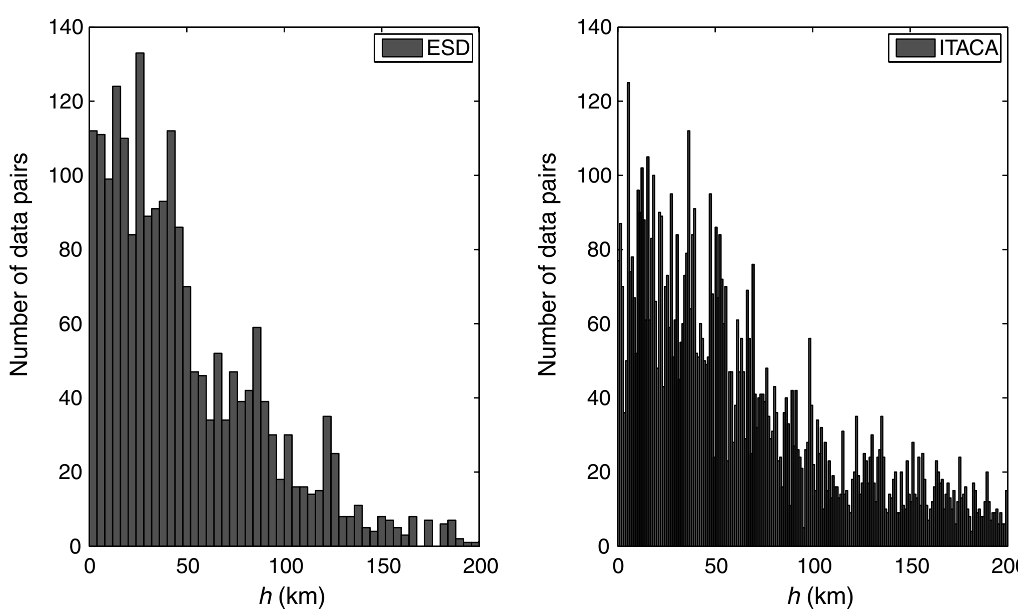


Figure 3. Histograms of the number of data pairs as a function of site-to-site separation distance.

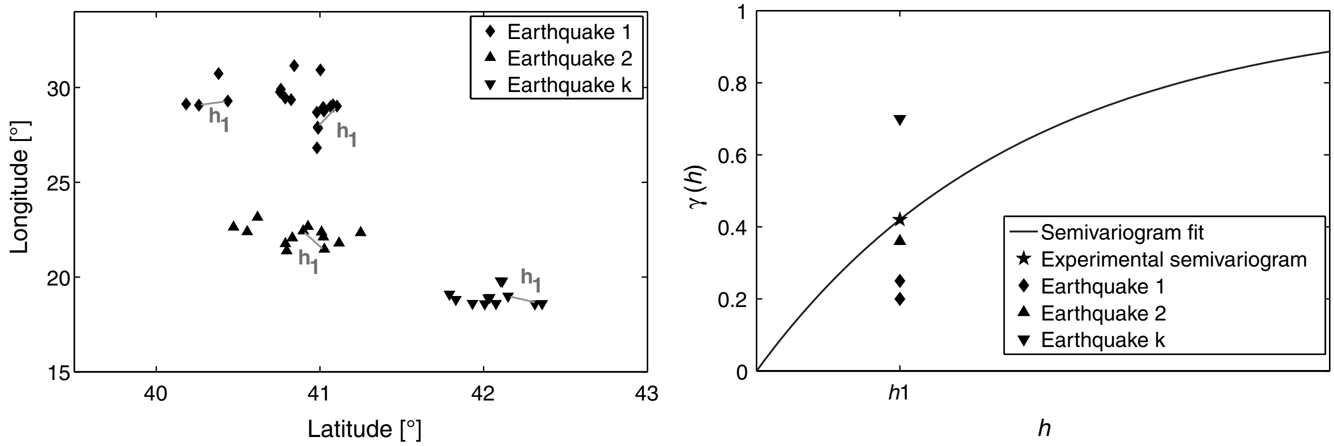


Figure 4. Pooling standardized intraevent residuals of multiple events (left) to compute experimental semivariogram (right).

used; no significant difference was found in the shape of the fitted semivariogram.

Of the three basic models (Gaussian, spherical, and exponential), the exponential model provided the best fit at the small separation distances (where the correlation is expected to be strong). The least-squares method (LSM) was used as a starting point to fit the semivariogram. Because the LSM minimizes the fitting error over the whole distance interval data, and in order to give more importance to the small separation distances, the LSM has been applied until a limit separation distance (of the same order of magnitude of the range where correlation is expected to disappear). LSM results are then used as a reference to manually fit a model in the empirical semivariogram. This approach was used to estimate the correlation of both the PGA and PGV. Because the chosen GMPE refers to the geometric mean of the horizontal components, the correlation was estimated for this IM. Assuming that there is no nugget effect (this study does not investigate variations on a smaller scale with respect to that of the tolerance), the only parameter to estimate is the range b whose results equal 13.5 km for PGA and 21.5 km for PGV, as shown in Figure 5.

It should be noted that in Esposito *et al.* (2010), the proposed methodology was used to estimate the correlation of the PGA (horizontal and vertical components) intraevent residuals starting from a less-recent GMPE, the Ambraseys *et al.* (2005a,b), and the dataset was also not exactly the same. However, the resulting range was quite similar, 12 km and 18 km for the horizontal and vertical components, respectively.

For the Italian dataset, the spatial correlation ranges of residuals obtained from the Bindi *et al.* (2010) GMPE were 11.5 km and 14.5 km for PGA and PGV, respectively (Fig. 6). In this case, because of the denser dataset, it was possible to consider a 1-km bin width; however, it seems that estimates are not significantly dependent on such size. In fact, the exponential model for PGA obtained with a bin width of 4 km is characterized by a range of 13.5 km.

Discussion

IM and Dataset Effects

In the cases of both ESD and ITACA, correlation ranges are higher for PGV than for PGA. In fact, the acceleration

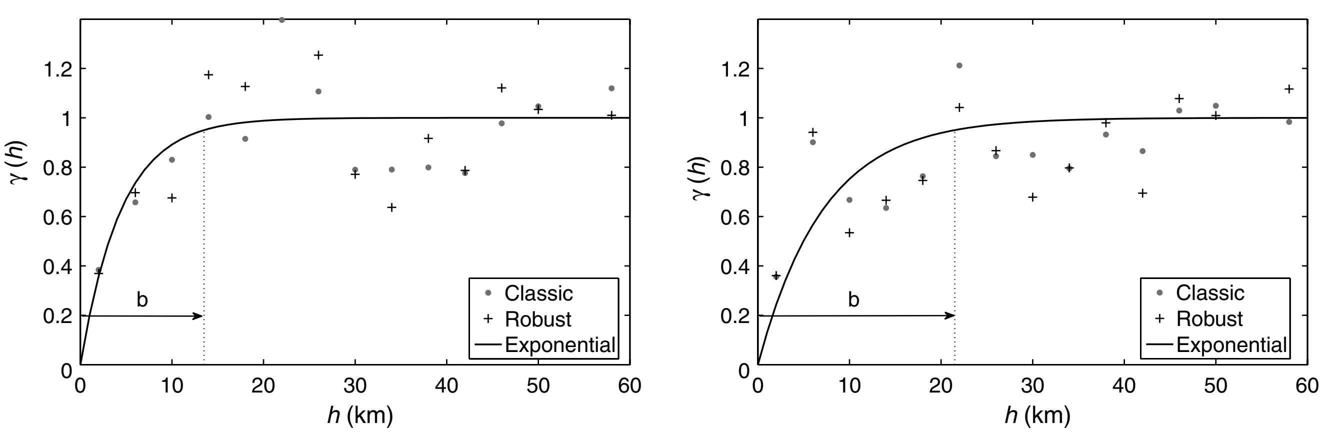


Figure 5. The ESD empirical semivariogram and fitted exponential model for PGA (left) and PGV (right) considering a bin width of 4 km.

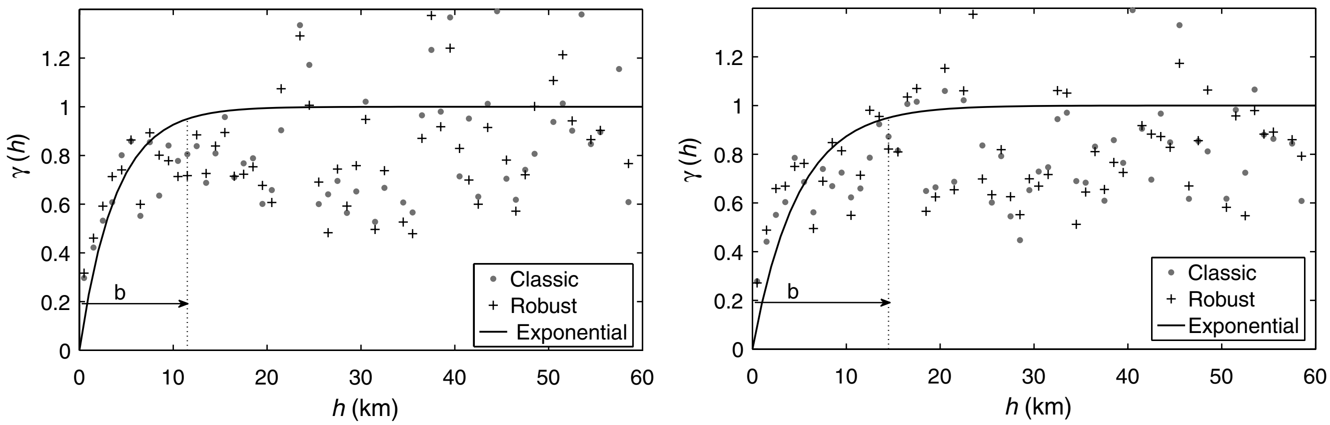


Figure 6. ITACA empirical semivariogram and fitted exponential model for PGA (left) and PGV (right) considering a bin width of 1 km.

time history shows a significant proportion of relatively high frequency, while velocity records shows substantially less high-frequency motion and are likely to yield higher correlations (e.g., Kramer, 1996). This seems to be consistent with past studies of ground-motion coherency (Zerva and Zervas, 2002). In fact, the coherency describes the degree of correlation between amplitudes and phase angles of two time histories at each of their component frequencies. Considering that coherency decreases with increasing distance between measuring points and with increasing frequency, it may be reasonable to expect more coherent ground motion, as velocity that corresponds to low frequency exhibits more correlated peak amplitudes.

In principle, residuals model what is not explained by the covariates of the GMPE; therefore, because the datasets are different, differences in the results between ESD and ITACA may be legitimate[3]. However, for a given IM, practical ranges are definitely comparable, and the differences are probably not significant, although the latter is difficult to assess because the estimation methodology does not provide the statistics of the range, which would allow us to quantitatively assess differences by means, for example, of a hypothesis test.

Influence of Standardization

As mentioned, there are different possibilities to obtain standardized residuals. As suggested in Goda and Atkinson (2010), positive correlations among intraevent residuals may lead to underestimated intraevent sample variance in GMPEs. Hence, intraevent standard deviations inferred from the large-separation-distance plateau of the semivariograms were used to estimate practical ranges of correlation in the two subsets. In particular, intraevent residuals without any standardization were used to estimate the sill (population variance) under the assumptions that at large-separation distances, those residuals are not correlated. The resulting estimates are practically the same (less than 10% differences) with respect to those of the GMPEs. This is also because there are relatively few data at short separation distances in the datasets. As a result, it was possible to infer that, at least in the considered case studies, the GMPEs variance can be used for the standardization.

Regional Hazard

Developed correlation models can be used, for example, to obtain the exceedance probability of the IM in a region and in a time interval of interest. The hazard integral of equation (15),

$$\lambda = \nu \cdot \int_M \int_{\boldsymbol{R}} P[IM_1 > im^*, \ldots\ldots\ldots, IM_n > im^* | M, \boldsymbol{R}] \cdot f_{M,\boldsymbol{R}}(m, \boldsymbol{r}) \cdot dm \cdot d\boldsymbol{r}, \tag{15}$$

provides such an annual rate of joint exceedance in a region if the same assumptions of site-specific hazard analysis are retained (McGuire, 2004). In equation (15), $f_{M,\boldsymbol{R}}(m, \boldsymbol{r})$ is the joint distribution of magnitude and distances referred to a particular seismic source; ν is the rate of occurrence of earthquakes on it; and $P[IM_1 > im^*, \ldots\ldots\ldots, IM_n > im^* | M, \boldsymbol{R}]$, the term affected by spatial correlation, is the conditional probability that the same[4] im^* threshold is exceeded at the n sites in which the region is discretized and whose distances from the source are represented by the vector $\boldsymbol{R} = \{R_1, \ldots, R_n\}$ (the integral is conventionally written as if it were a scalar). As an example, a regional hazard was developed considering the Paganica fault as a source, on which the 2009 L'Aquila (central Italy) earthquake originated, and the Bindi *et al.*, (2010) GMPE under the assumption that all the sites have the same rock local site conditions.

[3]One may argue that the larger ranges found for ESD with respect to ITACA are an effect of different distribution of magnitude in the two datasets. However, the proposed correlation models do not incorporate dependency on magnitude also based on the findings of Jayaram and Baker (2009).

[4]This is only a possible criterion, and others are possible.

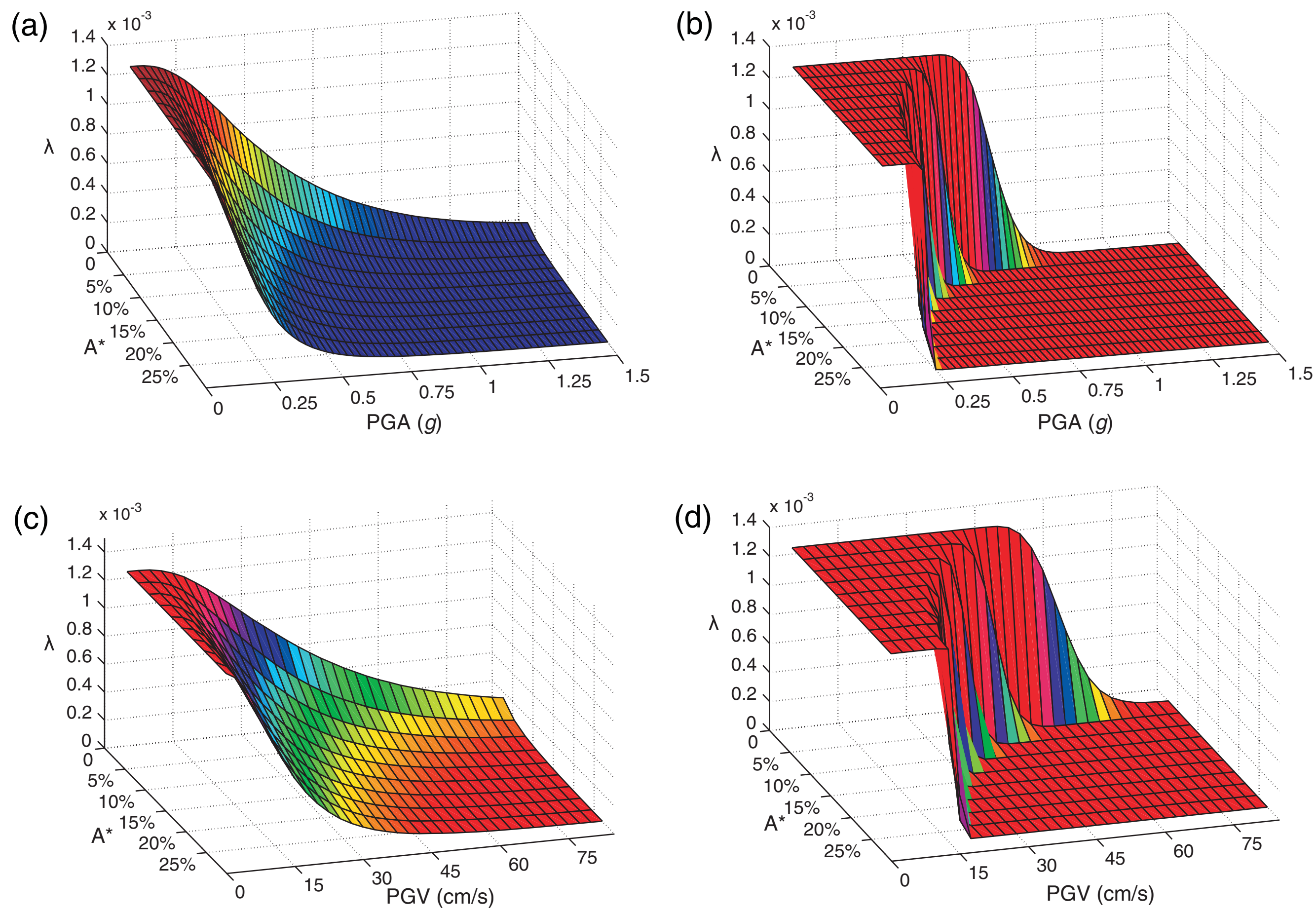


Figure 7. Regional-hazard surface considering the (a,c) correlated residuals, and the (b,d) independent residuals for PGA and PGV.

PGA and PGV hazards, considering ranges of 11.5 km and 14.5 km, respectively, were computed for a characteristic earthquake of moment magnitude 6.3 and occurrence rate on the source $\nu = 1/750$ (Pace *et al.*, 2006).

In Figure 7, surfaces are a function of IMs (as in traditional hazard curves) and exceedance areas (A^*), which are fractions, between 2.5% to 25%, of a region of 2500 km^2 around the fault. Referring, for example, to PGA, entering the plot with a pair of two A^* and PGA values, the surface returns the mean annual rate of exceedance of that PGA value over an area at least equal to A^*. For comparison, the hazard considering uncorrelated intraevent residuals was also computed.

For both PGA and PGV, correlation does not always provide higher rates with respect to the independent case. This is because, in the simulated case, the n sites constituting the A^* exceedance region are not necessarily adjoining. Given that im^* is exceeded (not exceeded) at a given site, correlation increases the probability of having neighboring sites exceeding (not exceeding) im^* as well.

If an alternate hazard criterion is considered, for example, im^* has to be exceeded at exactly n points constituting A^*, the joint hazard for correlated residuals is always higher with respect to the independent case (Fig. 8). This seems also consistent with the results of Sokolov and Wenzel (2011).

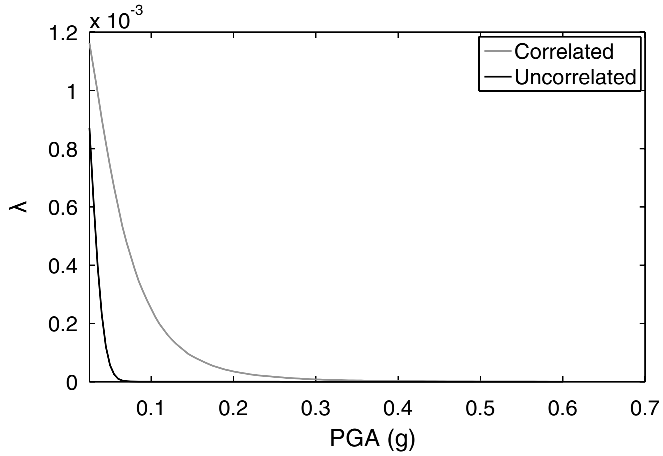
Conclusions

This study focused on the semiempirical estimation of spatial correlation of PGA and PGV using subsets of ESD and ITACA.

The hypotheses of the geostatistical analysis are stationarity and isotropy of the random fields. Consistent with the available literature on the topic, standardized intraevent residuals were used to compute experimental semivariograms that are a function of the site-to-site separation distance. Because a relatively small number of records for each earthquake was available, records from multiple events and regions were pooled to develop a unique model fitted with a large number of observations.

Exponential correlations were calibrated by finding practical ranges (the separation distance at which the correlation is technically lost) for ESD (ITACA), 13.5 km (11.5 km) for PGA and 21.5 km (14.5 km) for PGV.

The proposed results provide the basis for regional probabilistic seismic-hazard analysis, in other words, hazard



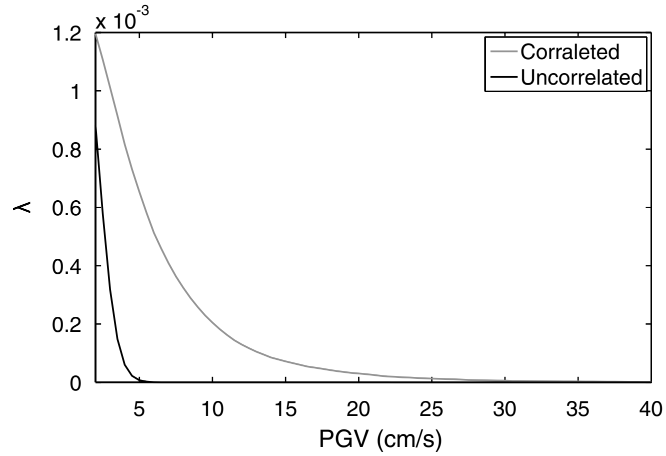


Figure 8. Regional-hazard curve considering the correlated and independent residuals for PGA (left) and for PGV (right).

analysis for spatially distributed systems. Illustrative examples show the differences in hazard assessment considering or ignoring the estimated correlations in the case of adopting different criteria.

Data and Resources

The ground motions and related information were provided by the authors of the Akkar and Bommer (2010) and Bindi *et al.* (2010) GMPEs for the ESD and ITACA datasets, respectively. In particular, this study considered subsets of data used to fit these GMPEs; in other words, this study used only free-field records from earthquakes for which more than one record was available.

Acknowledgments

This work was supported by AMRA scarl (http://www.amracenter.com) under the frame of SYNER-G (seventh framework program of the European Community for research, technological development, and demonstration activities; project contract number 244061). We want to thank Sinan Akkar (Middle East Technical University, Turkey), Dino Bindi (Deutsches GeoForschungsZentrum, Germany), and Francesca Pacor (Istituto Nazionale di Geofisica e Vulcanologia, Italy) for kindly providing us with the datasets used in this study. Moreover, we want to thank the associate editor of *BSSA*, Zhigang Peng, the two reviewers, Katsuichiro Goda (University of Bristol, United Kingdom) and Julian J. Bommer (Imperial College London, United Kingdom) for their comments, which improved the quality and readability of the paper, as well as Racquel K. Hagen (Stanford University, United States) for proofreading the manuscript. Finally, discussions with Massimiliano Giorgio (Seconda Università di Napoli, Italy) are gratefully acknowledged.

References

Akkar, S., and J. J. Bommer (2010). Empirical equations for the prediction of PGA, PGV and spectral accelerations in Europe, the Mediterranean Region and the Middle East, *Seismol. Res. Lett.* **81,** no. 2, 195–206.

Ambraseys, N. N., J. Douglas, S. K. Sarma, and P. M. Smit (2005a). Equations for the estimation of strong ground motions from shallow crustal earthquakes using data from Europe and the Middle East: Horizontal peak ground acceleration and spectral acceleration, *Bull. Earthq. Eng.* **3,** no. 1, 1–53.

Ambraseys, N. N., J. Douglas, S. K. Sarma, and P. M. Smit (2005b). Equations for the estimation of strong ground motions from shallow crustal earthquake using data from Europe and the Middle East: Vertical peak ground acceleration and spectral acceleration, *Bull. Earthq. Eng.* **3,** no. 1, 55–73.

Barnes, R. J. (1991). The variogram sill and the sample variance, *Math. Geol.* **23,** no. 4, 673–678.

Bindi, D., L. Luzi, and A. Rovelli (2010). Ground motion prediction equations (GMPEs) derived from ITACA, Deliverable No. 14, *Project S4: Italian Strong Motion Data Base*, http://esse4.mi.ingv.it.

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, earthquake 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.

Cressie, N. (1993). *Statistics for Spatial Data*, Revised edition, Wiley, New York, 900 pp.

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

Crowley, H., and J. J. Bommer (2006). Modelling seismic hazard in earthquake loss models with spatially distributed exposure, *Bull. Earthq. Eng.* **4,** no. 3, 249–275.

Crowley, H., P. J. Stafford, and J. J. Bommer (2008). Can earthquake loss models be validated using field observations, *J. Earthq. Eng.* **12,** no. 7, 1078–1104.

Esposito, S., I. Iervolino, and G. Manfredi (2010). PGA semi-empirical correlation models based on European data, *Fourteenth Europ. Conf. Earthq. Eng.*, Ohrid, Macedonia, 30 August–3 September, Paper no. 998.

Eurocode 8 (2004). Design of structures for earthquake resistance, Part 1: General rules, seismic actions and rules for buildings, EN 1998-1, *European Committee for Standardization (CEN)*, http://www.cen.eu/cenorm/homepage.htm.

Goda, K., and G. M. Atkinson (2009). Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan, *Bull. Seismol. Soc. Am.* **99,** no. 5, 3003–3020.

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

Goda, K., and H. P. Hong (2008b). Estimation of seismic loss for spatially distributed buildings, *Earthq. Spectra* **24,** no. 4, 889–910.

Goovartes, P. (1997). *Geostatistics for Natural Resources Evaluation*, Oxford University Press, New York, 496 pp.

Hong, H. P., Y. Zhang, and K. Goda (2009). Effect of spatial correlation on estimated ground motion prediction equations, *Bull. Seismol. Soc. Am.* **99,** no. 2A, 928–934.

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

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground motion intensities, *Earthq. Eng. Struct. Dynam.* **38,** no. 15, 1687–1708.

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

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

Kramer, S. L. (1996). *Geotechnical Earthquake Engineering*, Prentice Hall, Upper Saddle River, New Jersey, 653 pp.

Malhotra, P. (2008). Seismic design loads from site-specific and aggregate hazard analyses, *Bull. Seismol. Soc. Am.* **98,** no. 4, 1849–1862.

Matheron, G. (1962). *Traité de géostatistique appliquée*, Editions Technip, Paris, 333 pp. (in French).

McGuire, R. K. (2004). *Seismic Hazard and Risk Analysis*, Earthquake Engineering Research Institute, MNO-10, Oakland, California, 178 pp.

Pace, B., L. Perruzza, G. La Vecchia, and P. Boncio (2006). Layered seismogenic source model and probabilistic seismic-hazard analyses in central Italy, *Bull. Seismol. Soc. Am.* **96,** no. 1, 107–132.

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, *Tenth Int. Conf. Appl. Stat. Prob. Civ. Eng. (ICASP10)*, Tokyo, Japan, 2007.

Sokolov, V., and F. Wenzel (2011). Influence of ground motion correlation on probabilistic assessments of seismic hazard and loss: Sensitivity analysis, *Bull. Earthq. Eng.* doi 10.1007/s10518-011-9264-4.

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

Strasser, F. O., N. A. Abrahamson, and J. J. Bommer (2009). Sigma: Issues, insights, and challenges, *Seismol. Res. Lett.* **80,** no. 1, 40–56.

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

Zerva, A., and V. Zervas (2002). Spatial variation of seismic ground motion, *Appl. Mech. Rev.* **55,** no. 3, 271–297.

Dipartimento di Ingegneria Strutturale
Università degli Studi di Napoli Federico II
via Claudio 21, 80125, Naples, Italy
iunio.iervolino@unina.it

Manuscript received 22 April 2011