Bulletin of Earthquake Engineering (2019) 17:1099–1115
https://doi.org/10.1007/s10518-018-0506-6

ORIGINAL RESEARCH

Uncertainty in intraevent spatial correlation of elastic pseudo-acceleration spectral ordinates

Pablo Heresi[1] · **Eduardo Miranda**[1]

Received: 16 July 2018 / Accepted: 20 October 2018 / Published online: 24 October 2018

Abstract
The probabilistic nature of seismic ground motion intensity measures such as peak ground acceleration and spectral acceleration ordinates has been extensively studied during the last decades. However, their spatial correlation is mostly considered without any event-to-event variability, using a mean estimate from a number of seismic events. The present study quantitatively evaluates the event-to-event uncertainty of intraevent spatial correlations, using 39 well-recorded earthquakes. Results indicate a high event-to-event variability in the correlation model parameters, which if taken explicitly into account, would improve regional hazard and risk analyses. Event magnitude was found to be a statistically significant predictor variable of the model parameter, however it explains less than 20% of the total event-to-event variability. Moreover, clustering of site conditions, tectonic region, and fault mechanism are not statistically significant as predictor variables of the spatial correlation model parameter. Finally, this paper proposes a simple Monte Carlo approach for considering the high event-to-event variability of spatial correlation models, taking advantage of the Markov dependence of residuals for reducing the number of correlated variables to be simulated. This approach can be used with different intraevent spatial correlation models, as long as proper estimates of the dispersion of their parameters are considered.

Keywords Spatial correlation · Uncertainty · Ground motion intensity measure · Regional risk assessment

1 Introduction

Peak ground acceleration (*PGA*) and spectral acceleration ordinates (*Sa*) are the most widely used ground motion intensity measures (IM) for seismic hazard analyses. The probabilistic nature of these parameters is a well-researched topic and several ground motion prediction equations (GMPE), which provide estimations of their median and dispersion, have been developed (e.g., the NGA-West2 generation of GMPEs Bozorgnia et al. 2014). Moreover, estimating seismic risk on spatially distributed infrastructure (e.g., lifelines) or on many structures

✉ Pablo Heresi
pheresi@stanford.edu

[1] John A. Blume Earthquake Engineering Center, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA 94305-4020, USA

in a region requires not only the estimation of mean and dispersion of ground motion parameters at each location, but also requires the characterization of correlations of their residuals (e.g., Wesson and Perkins 2001; Lee and Kiremidjian 2007). Note that this correlation is applied to ground motion residuals, therefore it is related to changes in variability during different earthquakes and at different locations.

The spatial correlation of *Sa* values at different locations has been investigated in several studies over the last 15 years (Wesson and Perkins 2001; Boore et al. 2003; Kawakami and Mogi 2003; Wang and Takada 2005; Park et al. 2007; Goda and Hong 2008; Goda and Atkinson 2009, 2010; Hong et al. 2009; Jayaram and Baker 2009; Goda 2011; Sokolov et al. 2012; Sokolov and Wenzel 2013a; Loth and Baker 2013). The interested reader is also referred to the exhaustive literature review developed by Sokolov and Wenzel (2013b) about spatial correlations of ground motions. In general, previous studies grouped sets of a few different events for developing overall spatial correlation equations, mainly because, with the exception of the 1999 Chi–Chi and 1994 Northridge earthquakes, individual earthquakes have not produced the sufficient number of records for accurate estimations. Although this procedure generates smooth estimations of spatial correlations, it has the drawback of neglecting the variability between different events: the "event-to-event" variability. Even though some authors have commented on this variability when comparing the results of California events with the 1999 Chi–Chi earthquake and its aftershocks (Goda and Hong 2008; Jayaram and Baker 2009), the first author to highlight the large event-to-event variability was Goda (2011), who compared spatial correlations of intraevent terms (also referred to as intraevent spatial correlations) computed from 41 different earthquakes. He found that, for example, the intraevent correlation at 10 km between *Sa* values at a vibration period $T=0.2$ s has a median of 0.5, but it could vary approximately between 0.1 and 0.8. These two values would lead to significantly different results in a regional seismic risk estimation from those computed using a single value obtained from an approximate equation that groups all the events and neglects the high variability (Goda and Atkinson 2009; Sokolov and Wenzel 2011).

In this context, the main objective of the present work is to quantitatively evaluate the event-to-event uncertainty of the spatial correlation of *PGA* and *Sa* ordinates, and to propose an approach for explicitly considering it in seismic hazard and risk analyses. In particular, this investigation presents a correlation model, which is then fitted to 39 well-recorded earthquakes. A statistical study of the resulting parameters is conducted in order to propose a new methodology for considering the uncertainty of the spatial correlation model in future regional seismic hazard and risk analyses.

2 Intraevent spatial correlation model

The intraevent spatial correlation model described in this section is similar to the one presented by other authors (Goda and Hong 2008; Goda and Atkinson 2009, 2010; Hong et al. 2009; Jayaram and Baker 2009; Goda 2011). Several studies have shown that response spectral ordinates can be assumed to be lognormally distributed random variables (e.g., Abrahamson 1988; Jayaram and Baker 2008), and GMPEs can be used to estimate their median and dispersion as a function of the event magnitude, source-to-site distance, and other variables, such as the local soil conditions, fault mechanism, and tectonic region. In the case of spectral acceleration ordinates, this is expressed as follows:

$$\ln Sa_{ik}(T) = f\left(T, M_k, R_{ik}, \boldsymbol{\theta}_{\mathbf{ik}}\right) + \eta_k(T) + \varepsilon_{ik}(T), \tag{1}$$

where $Sa_{ik}(T)$ is the pseudo-acceleration spectral ordinate for a vibration period T at the i-th site in the k-th event. The function $f(T, M, R, \boldsymbol{\theta})$ is the logarithmic mean value estimated by a GMPE as a function of the event magnitude M, source-to-site distance R, and a set of other explanatory variables $\boldsymbol{\theta}$. Note that this function is deterministic, given a set of input parameters. The randomness of the intensity measure is accounted by $\eta_k(T)$ and $\varepsilon_{ik}(T)$, which are the interevent (also referred to as between-event) and intraevent (also referred to as within-event) residual terms, respectively. These terms are assumed to be independent and normally distributed random variables with zero mean and standard deviations $\sigma_\eta(T)$ and $\sigma_\varepsilon(T)$, respectively. Note that in some models these standard deviations might be expressed also as a function of the event magnitude and other variables. The interevent term, $\eta_k(T)$, represents the variability between different earthquake events, independent of the site, while the intraevent term, $\varepsilon_{ik}(T)$, represents the site-to-site variability within an earthquake event. Finally, since $\eta_k(T)$ and $\varepsilon_{ik}(T)$ are assumed to be independent, then the total standard deviation of $\ln Sa_{ik}(T)$ is $\sigma_T(T)$, given by:

$$\sigma_T(T) = \sqrt{\sigma_\eta^2(T) + \sigma_\varepsilon^2(T)}. \tag{2}$$

Estimating a correlation between spectral ordinates $Sa(T)$ at different locations would not be appropriate, even when using records from a single earthquake. This is because the underlying distribution of each realization (i.e., values of the IM) is different, as they have different values of R_{ik} and $\boldsymbol{\theta_{ik}}$. However, realizations of the residual terms $\eta_k(T) + \varepsilon_{ik}(T)$ come from the same distribution (normal distribution with zero mean and standard deviation $\sigma_T(T)$, as explained above). Thus, the correlation between residual terms, $\eta(T) + \varepsilon(T)$, at two sites i and j separated by a distance Δ_{ij}, and for two different periods T_i and T_j, respectively, can be demonstrated to be given by:

$$\rho_T\left(\Delta_{ij}, T_i, T_j\right) = \frac{\rho_\eta\left(T_i, T_j\right)\sigma_\eta\left(T_i\right)\sigma_\eta\left(T_j\right) + \rho_\varepsilon\left(\Delta_{ij}, T_i, T_j\right)\sigma_\varepsilon\left(T_i\right)\sigma_\varepsilon\left(T_j\right)}{\sigma_T\left(T_i\right)\sigma_T\left(T_j\right)}, \tag{3}$$

where $\rho_T(\Delta_{ij},T_i,T_j)$ is the correlation between $\eta_k(T_i) + \varepsilon_{ik}(T_i)$ and $\eta_k(T_j) + \varepsilon_{jk}(T_j)$, $\rho_\eta(T_i,T_j)$ is the interevent correlation between $\eta_k(T_i)$ and $\eta_k(T_j)$, and $\rho_\varepsilon(\Delta_{ij},T_i,T_j)$ is the intraevent correlation between $\varepsilon_{ik}(T_i)$ and $\varepsilon_{jk}(T_j)$. Note that, even when for very large separation distances $\rho_\varepsilon(\Delta_{ij},T_i,T_j)$ is expected to decay to zero, there will be always some total correlation $\rho_T(\Delta_{ij},T_i,T_j)$, due to the correlation between interevent residuals of the same earthquake. Goda and Hong (2008) proposed that the total spatial correlation can be approximated by:

$$\rho_T\left(\Delta_{ij}, T_i, T_j\right) \cong \frac{\rho_0\left(T_i, T_j\right)\left[\sigma_\eta\left(T_i\right)\sigma_\eta\left(T_j\right) + \rho_\varepsilon\left(\Delta_{ij}, T_{\max}, T_{\max}\right)\sigma_\varepsilon\left(T_i\right)\sigma_\varepsilon\left(T_j\right)\right]}{\sigma_T\left(T_i\right)\sigma_T\left(T_j\right)}, \tag{4}$$

where T_{max} is the largest value of the two periods, T_i and T_j, and $\rho_0(T_i,T_j)$ represents the empirical approximation of $\rho_T(\Delta_{ij}=0,T_i,T_j)$, previously studied by several authors (e.g., Inoue and Cornell 1990; Baker and Cornell 2006; Abrahamson and Silva 2007; Baker and Jayaram 2008; Abrahamson et al. 2013). The approximation from Eqs. (3) to (4) comes from the assumption of a Markov dependence of residuals at different periods, and its correctness was demonstrated by Loth and Baker (2013). Thus, the focus of this paper is the study of the intraevent spatial correlation $\rho_\varepsilon(\Delta_{ij},T_{max},T_{max})$, which will be denoted $\rho_\varepsilon(\Delta,T)$ hereafter.

In order to estimate $\rho_\varepsilon(\Delta,T)$, a regression analysis has to be carried out in a first step in order to determine $f(T, M, R, \boldsymbol{\theta})$, $\sigma_\eta(T)$ and $\sigma_\varepsilon(T)$. In this paper the GMPE developed by Boore et al. (2014) is used for computing the mean and dispersions of the regression. The regression residuals are then used to evaluate the intraevent spatial correlation. For a given event and period, the interevent residual term, $\eta_k(T)$, is a constant for all the sites, therefore the residuals $\ln Sa_{ik}(T) - f(T, M_k, R_{ik}, \boldsymbol{\theta_{ik}}) = \eta_k(T) + \varepsilon_{ik}(T)$ only give information about the intraevent spatial correlation. One approach for estimating $\rho_\varepsilon(\Delta,T)$ is to directly compute the covariance and the correlation coefficient between residuals at sites separated at a certain distance Δ. Another approach, first recommended by Goda and Hong (2008), which is consistent with the geostatistical practice, and also used by Jayaram and Baker (2009) and Loth and Baker (2013), is to assume stationarity and isotropy, then using the sample semivariogram (Goovaerts 1997), $[\sigma_d(\Delta,T)]^2/2$. The term $\sigma_d(\Delta,T)$ is the standard deviation of $\varepsilon_d(\Delta,T) = \varepsilon_{ik}(T - \varepsilon_{jk}(T)$, where the i-th and j-th sites have a separation distance Δ (note that, as it is unlikely to find several data points with an exact separation Δ, the data points are organized in bins of separation distance). In other words, all the pairs of residuals at sites with a separation distance within the range of distances of a given bin are subtracted to generate a new variable $\varepsilon_d(\Delta,T)$ and then $\sigma_d(\Delta,T)$ is computed as its standard deviation. Finally, the intraevent spatial correlation is evaluated as:

$$\rho_\varepsilon(\Delta, T) = 1 - \frac{1}{2}\left[\frac{\sigma_d(\Delta, T)}{\hat{\sigma}_\varepsilon(T)}\right]^2, \tag{5}$$

where $\hat{\sigma}_\varepsilon(T)$ is the intraevent standard deviation of the event k. It is paramount to note that the intraevent standard deviation from the original GMPE, $\sigma_\varepsilon(T)$, must not be used as $\hat{\sigma}_\varepsilon(T)$ in Eq. (5), because this is a constant value for different events (all the events considered in the development of the GMPE), while in reality $\hat{\sigma}_\varepsilon(T)$ may vary significantly from one event to another. Therefore, using a constant $\hat{\sigma}_\varepsilon(T) = \sigma_\varepsilon(T)$ would introduce a bias into the estimation of $\rho_\varepsilon(\Delta,T)$. Thus, it is generally recommended to assume that the semivariogram reaches a plateau at long separation distances, where the correlation is theoretically zero. Therefore $\hat{\sigma}_\varepsilon(T)^2$ can be assumed to be equal to $0.5\ \sigma_d(\Delta,T)^2$ for a very large separation distance. In this work, the semivariogram approach is taken, separation distances that differ by no more than 3 km are grouped into the same bin, and $\hat{\sigma}_\varepsilon(T)^2$ is computed as the plateau value of $0.5\ \sigma_d(\Delta,T)^2$ at distances between 85 and 180 km. Using other width ranges of separation distances in each bin and large distances for $\hat{\sigma}_\varepsilon(T)$ produce almost the same results than those shown in the following sections.

At this point it is important to note that the methodology previously described is in theory inconsistent if the GMPE that estimates $f(T, M, R, \boldsymbol{\theta})$, $\sigma_\eta(T)$, and $\sigma_\varepsilon(T)$ of Eq. (1) does not explicitly include the spatial correlation of residuals, which is the common practice (e.g., the NGA-West2 generation of GMPEs assume no spatial correlation). However, Hong et al. (2009) included the spatial correlation in their regression analysis for developing a GMPE with a set of California records, concluding that the effect is negligible when comparing it with a GMPE without spatial correlation. Therefore, even when the procedure previously described is in theory inconsistent, in practice it can be used with standard GMPEs without significant errors.

Once the empirical $\rho_\varepsilon(\Delta,T)$ was obtained using Eq. (5), the following functional form was fitted to the data:

$$\hat{\rho}_\varepsilon(\Delta, T) = \exp\left[-\left(\frac{\Delta}{\beta(T)}\right)^{\alpha(T)}\right], \tag{6}$$

where α and β are the model parameters obtained using a nonlinear regression, which are function of the vibration period T. The parameter α controls the decaying rate of correlation with increasing distance Δ, and β is the distance at which the correlation is $exp(-1) = 0.368$. Note that for a fixed value of β, a higher value of α produces higher correlation values for distances $\Delta < \beta$, and lower correlation values for $\Delta > \beta$. A least-squares regression was used to fit the model and obtain the parameters α and β as a function of vibration period. However, as correlation coefficients have non-constant standard errors, the following transformation was used, known as Fisher z transformation:

$$z = \frac{1}{2} \log \left(\frac{1+\rho}{1-\rho} \right), \tag{7}$$

where ρ is the estimated correlation from Eq. (5) and z is the transformed value, now with a constant standard error. Then, in order to obtain the model parameters, α and β, the least-squares regression was conducted with the z values.

Although this paper is focused on $\rho_\varepsilon(\Delta,T)$, please note that once this correlation is estimated, the total correlation between residual values $\rho_T(\Delta_{ij},T_i,T_j)$, can be easily computed using Eq. (4). Furthermore, as the residual terms are the only source of uncertainty in Eq. (1), the correlation between ln *Sa(T)* values is equal to the total correlation between residual terms.

3 Ground motion database

The correlation model expressed in Eq. (6) was fitted to the empirical correlations of different earthquakes individually. The reliability of these empirical correlations increases with the number of stations that recorded the earthquake (i.e., as the sample size increases). Therefore, only well-recorded events, with more than 200 recorded ground motions (i.e., 100 stations with two horizontal perpendicular directions), were considered. This is also consistent with the selection criteria used by Goda (2011). A total of 39 earthquakes were selected, with magnitudes ranging between 4.0 and 7.9 in order to study the influence of the event magnitude in the spatial correlation. Here, only ground motions recorded in stations with NEHRP site class C or D (average shear wave velocity of the top 30 m, V_{S30}, between 180 and 760 m/s) were considered, as these are the most common site classes encountered in most urban areas. Table 1 shows the information about the earthquakes considered. All the ground motions used in this study were obtained from the NGA-West2 database (Ancheta et al. 2014).

It should be noted that the reliability of estimates of correlation coefficients computed from a given sample size increases as the absolute value of the correlation coefficient increases. Thus, correlation coefficients from this study are particularly well estimated for separation distances smaller than 30 km, which are the ones that have the largest influence in regional seismic hazard and risk analyses of urban areas.

4 Variability of intraevent spatial correlation

The procedure described in Sect. 2 was applied to every earthquake event, for *PGA* and for pseudo-acceleration spectral ordinates at periods of vibration between 0.1 and 6.0 s. Figure 1 illustrates, as an example, the results of the 2007 Chuetsu-oki, Japan

Table 1 Summary of considered earthquakes

Event name	Year	Magnitude	Fault mechanism	Latitude [°]	Longitude [°]	Number of stations[a]
Chuetsu-oki	2007	6.80	Reverse	37.538	138.617	543
Niigata, Japan	2004	6.63	Reverse	37.307	138.839	464
Chi–Chi, Taiwan	1999	7.62	Reverse	23.850	120.820	371
Tottori, Japan	2000	6.61	Strike-Slip	35.275	133.350	366
El Mayor-Cucapah	2010	7.20	Strike-Slip	32.300	− 115.267	333
Iwate	2008	6.90	Reverse	39.027	140.878	332
Chi–Chi, Taiwan-05	1999	6.20	Reverse	23.810	121.080	300
Chi–Chi, Taiwan-02	1999	5.90	Reverse	23.940	121.010	273
Chi–Chi, Taiwan-06	1999	6.30	Reverse	23.870	121.010	273
Wenchuan, China	2008	7.90	Reverse	30.986	103.364	262
10370141	2009	4.45	Strike-Slip	34.108	− 117.306	249
Chi–Chi, Taiwan-03	1999	6.20	Reverse	23.810	120.850	231
Chi–Chi, Taiwan-04	1999	6.20	Strike-Slip	23.600	120.820	231
14312160	2007	4.66	Reverse	34.298	− 118.626	210
10275733	2007	4.73	Strike-Slip	33.733	− 117.492	201
Anza-02	2001	4.92	Normal	33.508	− 116.514	198
14383980	2008	5.39	Reverse	33.947	− 117.767	187
40204628	2007	5.45	Strike-Slip	37.432	− 121.777	173
10410337	2009	4.70	Strike-Slip	33.928	− 118.354	169
40199209	2007	4.20	Strike-Slip	37.806	− 122.185	168
14138080	2005	4.59	Normal	34.998	− 119.193	164
14186612	2005	4.69	Reverse	35.017	− 119.025	153
71336726	2010	4.05	Strike-Slip	37.481	− 121.799	153
Northridge-01	1994	6.69	Reverse	34.206	− 118.554	147
9983429	2004	4.34	Strike-Slip	35.017	− 119.149	142
21522424	2006	4.30	Strike-Slip	37.100	− 121.491	140
Parkfield-02, CA	2004	6.00	Strike-Slip	35.817	− 120.365	137
14151344	2005	5.20	Strike-Slip	33.533	− 116.570	135
51207740	2008	4.10	Strike-Slip	37.862	− 122.000	135
9753485	2002	4.18	Reverse	34.362	− 118.664	132
Hector Mine	1999	7.13	Strike-Slip	34.598	− 116.265	124
21530368	2006	4.50	Strike-Slip	38.367	− 122.588	115
Whittier Narrows-01	1987	5.99	Reverse	34.049	− 118.081	113
21465580	2005	4.77	Strike-Slip	39.314	− 120.064	111
21437727	2005	4.18	Strike-Slip	37.393	− 121.486	110
14155260	2005	4.88	Reverse	34.066	− 117.001	107
Yorba Linda	2002	4.27	Strike-Slip	33.917	− 117.776	120
Big Bear City	2003	4.92	Strike-Slip	34.310	− 116.848	100
Darfield, New Zealand	2010	7.00	Strike-Slip	− 43.615	172.049	100

[a]Only stations with NEHRP site class C or D

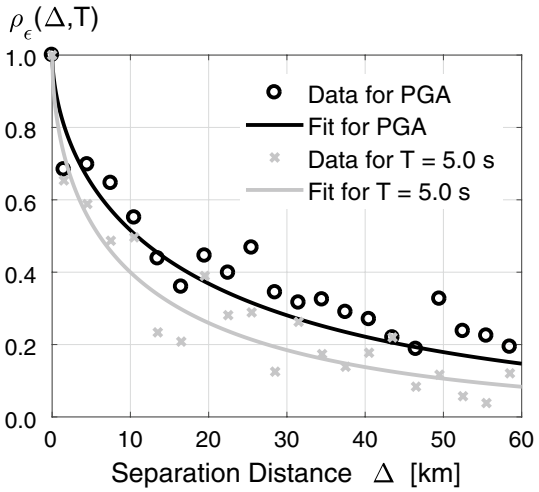

Fig. 1 Empirical and fitted spatial intraevent correlation for *PGA* and *Sa* (5.0 s) in the 2007 Chuetsu-oki, Japan earthquake. For *PGA* (α, β) = (0.59, 20.05 km), while for *Sa* (5.0 s) (α, β) = (0.56, 11.69 km)

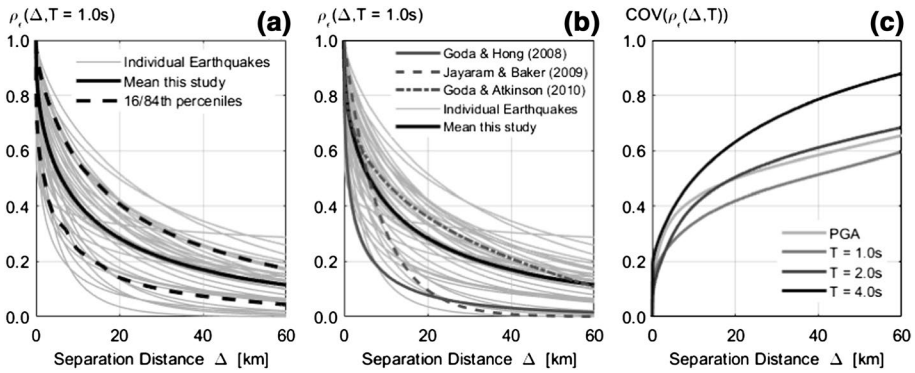

Fig. 2 **a** Fitted spatial intraevent correlation for *Sa*(1.0 s) for 39 individual earthquake events, their mean, and their 16/84th percentiles. **b** Comparison of the fitted spatial intraevent correlation for 39 individual earthquake events and their mean, with previous models. **c** Coefficient of variation (COV) of the intraevent correlation coefficient as a function of separation distance, for four different periods of vibration

earthquake for *PGA* and *Sa*(5.0 s). The fitted parameters α and β for this event are (α, β) = (0.59, 20.05 km) for *PGA*, and (α, β) = (0.56, 11.69 km) for *Sa* (5.0 s). Figure 2a shows the results of fitting the model of Eq. (6) to the 39 earthquake events of Table 1, for *Sa* (1.0 s). The mean value of the correlation coefficient and the 16th and 84th percentiles, as function of separation distance, are also shown in the same figure. Figure 2b compares the spatial correlation results for *Sa(1.0 s)* from the 39 events and their mean with previous models developed by Goda and Hong (2008), Jayaram and Baker (2009), and Goda and Atkinson (2010). Figure 2c presents the coefficient of variation (COV) of the correlation coefficients for *Sa* at four different periods of vibration, as a function of separation distance. As can be observed, there is a significant variability in the intraevent correlation coefficient at a given separation distance for every period. In order to evaluate this variability with an overall measure, the following sections are focused on the probability distribution of the model parameters α and β.

4.1 Central tendency and variability of correlation model parameters

The resulting values of α and β as a function of period for each individual earthquake are shown in Fig. 3a, b, in light gray lines. The central tendencies and the counted 16th and 84th percentile values are also presented in the same figures. As noted before, the parameters fitted to earthquakes with more records are smoother and more reliable than those fitted to events with less ground motion records, thus the central tendency is computed as the weighted geometric mean, where each earthquake result is weighted by the square of the number of stations (see Table 1 for the number of stations of each event). Dispersions of the parameters α and β as a function of period were computed as the weighted standard deviation of the natural logarithm of α and β values, and are shown in Fig. 3c. As can be seen, the parameter α is fairly constant across periods, with a value approximately equal to 0.55, and it has a significantly lower dispersion than the parameter β. Thus, the total variability of $\hat{\rho}_{\varepsilon}(\Delta, T)$ is dominated by the dispersion of β. Therefore, in order to simplify the model of Eq. (6), a new functional form can be used:

$$\hat{\rho}_{\varepsilon}(\Delta, T) = \exp\left[-\left(\frac{\Delta}{\beta(T)}\right)^{0.55}\right]. \tag{8}$$

This also simplifies the comparison between different curves, since for a fixed α, a higher value of β is directly translated into higher spatial correlations. New β values were then computed with regression analyses using Eq. (8), consistent with the fixed value of α=0.55. Figure 4a illustrates the resulting β values for each individual event (in gray lines) with its corresponding weighted geometric mean, plotted as a function of period of vibration. The weighted geometric mean of β is also shown in Table 2. In order to compare these results to previous correlation models, β values are compared with the distances at which the correlation equals $exp(-1) = 0.368$ according to the models proposed by Goda and Hong (2008) and by Jayaram and Baker (2009). The former model was developed with 39 California earthquakes,

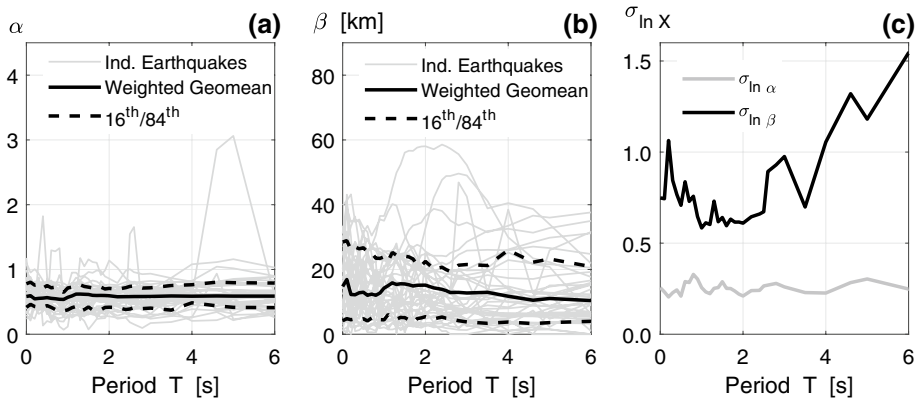

Fig. 3 **a** Variation of parameter α as a function of the period of vibration for individual earthquakes, and its central tendency weighted by the square of the number of stations of each event. **b** Variation of parameter β as a function of the period of vibration for individual earthquakes, and its central tendency weighted by the square of the number of stations of each event. **c** Dispersion (weighted logarithmic standard deviation) of parameters α and β as a function of the period of vibration

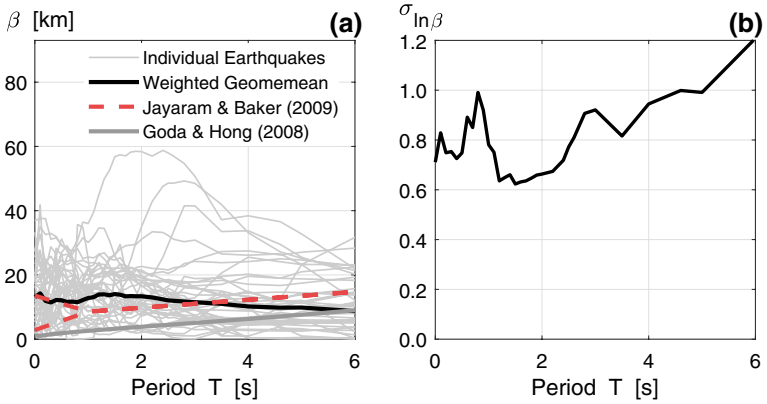

Fig. 4 **a** Variation of β (for $\alpha = 0.55$) as a function of the period of vibration for individual earthquakes, and its central tendency weighted by the square of the number of stations of each event. The results are compared with the models proposed by Goda and Hong (2008) and by Jayaram and Baker (2009). **b** Dispersion (weighted logarithmic standard deviation) of β (for $\alpha = 0.55$) as a function of the period of vibration

Table 2 Weighted geometric mean and weighted logarithmic standard deviation of β

Period T [s]	Geomean β [km]	$\sigma_{\ln\beta}$	Period T [s]	Geomean β [km]	$\sigma_{\ln\beta}$	Period T [s]	Geomean β [km]	$\sigma_{\ln\beta}$
0.0	13.10	0.71	1.2	13.91	0.64	2.6	11.96	0.81
0.1	14.25	0.83	1.3	13.95	0.65	2.8	11.72	0.91
0.2	11.90	0.75	1.4	13.68	0.66	3.0	11.51	0.92
0.3	11.44	0.75	1.5	14.04	0.62	3.5	11.01	0.82
0.4	12.20	0.73	1.6	13.84	0.63	4.0	10.18	0.94
0.5	12.11	0.75	1.7	13.32	0.64	4.6	9.81	1.00
0.6	11.58	0.89	1.8	13.42	0.65	5.0	9.85	0.99
0.7	11.69	0.85	1.9	13.35	0.66	6.0	8.73	1.21
0.8	11.52	0.99	2.0	13.36	0.66	7.0	8.15	1.24
0.9	12.20	0.92	2.2	12.98	0.67	8.0	7.57	1.36
1.0	12.81	0.78	2.4	12.39	0.72	9.0	7.72	1.49
1.1	13.09	0.75	2.5	12.08	0.77	10.0	9.04	1.16

while the latter only considered seven events and has two branches for periods shorter than 1.0 s, depending on the clustering of site conditions. Figure 4b and Table 2 show the dispersion of the β values computed for individual events, around the central tendency, for a fixed $\alpha = 0.55$. Note the high dispersion of the parameter β, between 0.6 and 1.2. As a reference, these dispersion values, which are currently neglected in regional seismic hazard analyses, are higher than those of spectral acceleration ordinates in GMPEs, dispersion values that are routinely incorporated in probabilistic seismic hazard analyses.

4.2 Probability distribution of β

At every period of vibration, the empirical cumulative probability distribution of β was computed. For this, the β values were sorted in ascending order and for each observation i, a probability (i.e., plotting position) equal to $(i - 3/8)/(n + 1/4)$ was assigned, where n is the sample size (i.e., the number of earthquakes in this case). This plotting position, proposed by Blom (1958), has been demonstrated to be a suitable approximation of the unbiased plotting position (Cunnane 1978). An example for $T = 2.0$ s is presented in Fig. 5, where a positive-skewed distribution (higher upper tail) of the data points can be observed. Thus, a lognormal distribution is evaluated to determine if it can characterize the probability distribution of β, using the Kolmogorov–Smirnov (K-S) goodness-of-fit test (Massey 1951). The fitted lognormal distribution and its K-S 10% significance confidence boundaries are also presented in Fig. 5. This test was repeated for every period of vibration. Figure 6 shows the maximum absolute difference between the empirical cumulative distribution of β, F_β, and the fitted lognormal distribution of β, F^*_β, as a function of period, along with the K-S 10% significance limit for this sample size, $D_{crit,10\%}$. As can be seen, β can be adequately assumed to have a lognormal distribution for all the periods.

4.3 Influence of earthquake magnitude and clustering of site conditions

In order to evaluate the influence of the event moment magnitude, M_w, β values were plotted against the magnitude of their corresponding events for each period of vibration. An example of this evaluation for $T = 3.0$ s is shown in Fig. 7. The Pearson's empirical correlation coefficient between β and M_w at this vibration period is 0.35, and a slight influence can be observed, where higher magnitudes are correlated with higher β values. This correlation was found to be higher for periods greater than 1.0 s than for shorter periods, as can be seen in Fig. 8a, which presents the Pearson's empirical correlation coefficient between β and event moment magnitude as a function of the period of vibration. However, despite this relatively important level of correlation for periods of vibration larger than 1.0 s, the variability of β is only slightly decreased (less than 20%) when the event magnitude is

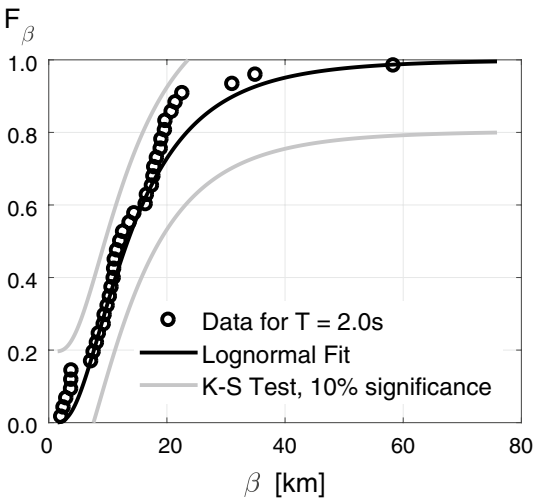

Fig. 5 Empirical cumulative distribution of β for $T = 2.0$ s, along with a fitted lognormal distribution and its K-S 10% significance confidence boundaries

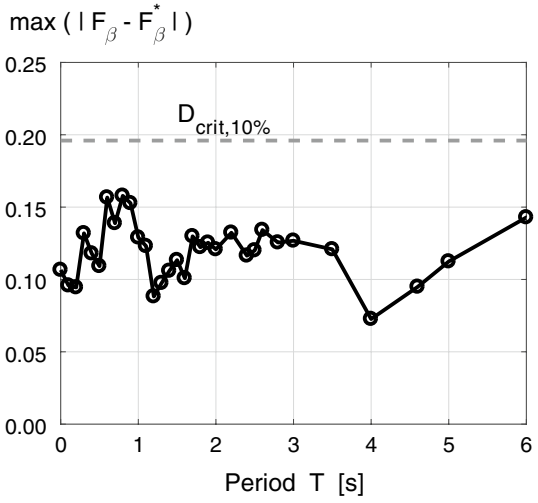

Fig. 6 Maximum absolute difference between the empirical cumulative distribution of β, F_β, and the fitted lognormal distribution of β, F^*_β, as a function of period, along with the K-S 10% significance limit for this sample size, $D_{crit,10\%}$

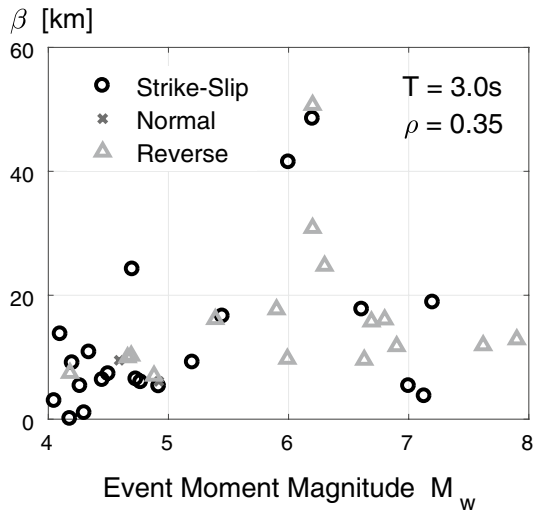

Fig. 7 Influence of event moment magnitude on β for $T = 3.0$ s. The corresponding Pearson's empirical correlation coefficient is 0.35

explicitly taken into account for estimating β. This is shown in Fig. 8b, which compares the dispersion of β before and after taking into account the earthquake magnitude for estimating it, considering a linear trend between β and M_w. This means that only a small fraction of the high variability of β is explained by changes in M_w. Considering a nonlinear model between β and M_w did not improved these results, as no clear trend, either linear or nonlinear, is observed between β and M_w.

On the other hand, Jayaram and Baker (2009) showed a trend between β values and the clustering of soils with similar geological conditions. To evaluate this, they used the spatial correlation of V_{S30} values as a proxy for clustering of site conditions. Considering seven earthquakes, the authors concluded that regions with higher spatial correlations of V_{S30} present higher spatial correlations between spectral ordinates at short periods of vibration. This is the reason behind the two branches for periods shorter than 1.0 s shown in Fig. 4a,

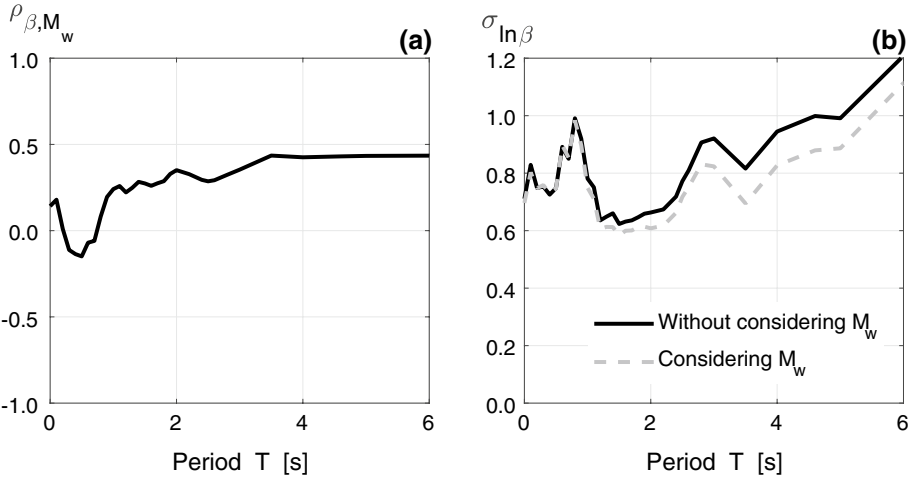

Fig. 8 **a** Pearson's empirical correlation coefficient between β and event moment magnitude as a function of the vibration period. **b** Comparison of variability of β before and after considering event magnitude, M_w, through a linear regression analysis for β as a function of M_w

where the top branch (higher β values) is for regions with clustering of V_{S30} and the bottom branch (lower β values) is for regions where the soil conditions vary widely. Sokolov et al. (2012) and Sokolov and Wenzel (2013a) drew similar conclusions about the influence of clustering of site conditions in Taiwan and Japan, using the same procedure than the one used by Jayaram and Baker (2009). In this study, we follow the same approach for evaluating the influence of clustering of site conditions. Similar to the spatial correlation of intensity measures, the empirical semivariogram was computed from V_{S30} values of each station, at every earthquake event. From the semivariograms, empirical spatial correlation coefficients were calculated using Eqs. (5) and (8) was fitted to the resulting data. The β values obtained are termed β_{VS30}, and represent a proxy for the clustering of site conditions: higher β_{VS30} values mean higher spatial correlations of V_{S30}, which are related to the clustering of site conditions. Then, β_{VS30} was used as a possible explanatory variable that could partially explain the high variability of β, by conducting a linear regression analysis for β as a function of β_{VS30}. Note, however, that this procedure only takes into account the influence of V_{S30}, and it does not consider other possible geological variables, such as the depth of sediments. Similarly to the procedure with the magnitude, Fig. 9 shows an example of a scatter plot between β and β_{VS30} for $T=0.2$ s, where the corresponding Pearson's empirical correlation coefficient is only -0.03, illustrating a negligible influence of β_{VS30} on β. Pearson's empirical correlation coefficients computed for every period of vibration are presented in Fig. 10a, again showing a very low correlation for all periods. Moreover, Fig. 10b shows that the reduction in the dispersion of β when explicitly considering a regression analysis between β and β_{VS30} is negligible, demonstrating that β_{VS30} (and thus the clustering of V_{S30}) has no influence on the spatial correlation of these intensity measures.

Finally, a linear regression analysis was conducted for every period of vibration, with the event moment magnitude, β_{VS30}, and two earthquake characteristics: tectonic region and fault mechanism. From these predictor variables, only the event moment magnitude was found to be statistically significant at a 5% significance level, and just for β values at periods greater than 1.0 s. Moreover, the resulting reduction of the variability of β is similar to the one presented in

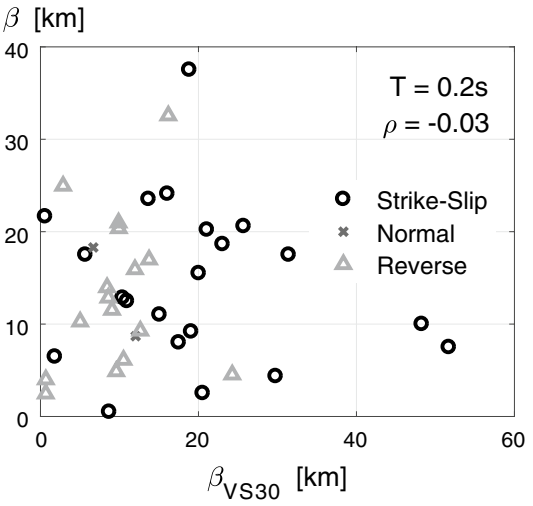

Fig. 9 Influence of β_{VS30} (proxy for clustering of site conditions) on β for $T=0.2$ s. The corresponding Pearson's empirical correlation coefficient is -0.03

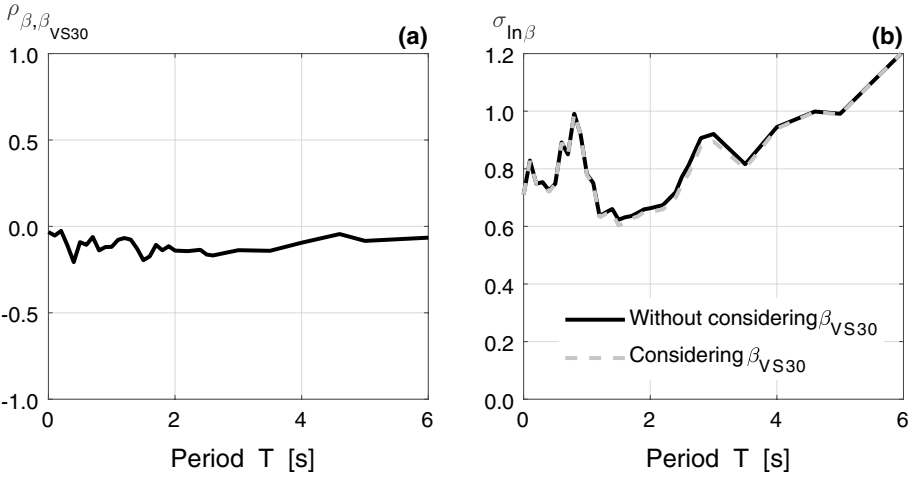

Fig. 10 **a** Pearson's empirical correlation coefficient between β and β_{VS30} (proxy for clustering of site conditions) as a function of the vibration period. **b** Comparison of variability of β before and after considering clustering of soil conditions through a linear regression analysis for β as a function of β_{VS30}

Fig. 8b, for regressions using only the event moment magnitude as explanatory variable. This is also consistent with Figs. 7 and 9, where no clear difference between fault mechanisms is observed.

5 Monte Carlo approach for considering the variability of spatial correlation

The previous section demonstrated that the intraevent spatial correlation of intensity measures during a given earthquake is characterized by a high inherent variability, and therefore, rather than just considering one correlation model (derived with either one event or a set of events), regional risk assessments can be improved by explicitly considering this dispersion. In this context, Eq. (8) can be used with β as a lognormally distributed random variable. It is proposed that the median and the dispersion of β are computed with the following simplified equations, which are also shown in Fig. 11:

$$\hat{\beta}(T) = \begin{cases} 4.231 \cdot T^2 - 5.180 \cdot T + 13.392 & T < 1.37\,\text{s} \\ 0.140 \cdot T^2 - 2.249 \cdot T + 17.050 & T \geq 1.37\,\text{s} \end{cases} \tag{9}$$

$$\sigma_{Ln\beta}(T) = 4.63x10^{-3} \cdot T^2 + 0.028 \cdot T + 0.713 \tag{10}$$

A simple and direct approach for incorporating the variability of the spatial correlation is to perform Monte Carlo simulations, by considering the spatial correlation model parameters as random variables. In most of the cases, different intensity measures (at different vibration periods) must be simulated at many sites. Therefore, taking advantage of the Markov approximation for reducing the number of correlated variables that must be simulated, the following simulation sequence for a given scenario earthquake can be adopted (note that, at each Monte Carlo simulation, this scenario earthquake can either stay fixed for estimating ground motion intensity measures for that particular event, or may vary for an event-based probabilistic seismic hazard analysis):

1. Define the set of locations (subscript $j = 1,2,\ldots J$) and periods (subscript $i = 1,2,\ldots I$) at which ground motion intensity measures will be simulated.
2. Obtain the maximum period of vibration to be simulated, $T_{max} = max(T_i)$.

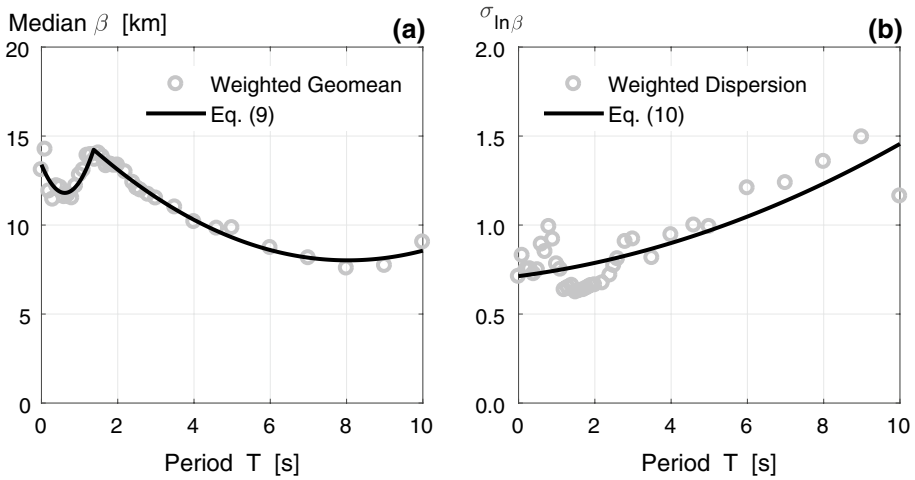

Fig. 11 **a** Computed weighted geometric mean and proposed fitted model. **b** Computed weighted logarithmic standard deviation and proposed fitted model

3. Choose a GMPE to be used for simulating the ground motion intensity measures.
4. Use Eqs. (9) and (10) with $T = T_{max}$ to compute the median and dispersion of $\beta(T_{max})$.
5. At the k-th simulation, obtain a realization of $\beta(T_{max})$ considering a lognormal distribution with median and dispersion computed at step 4.
6. Assemble the total spatial correlation model with Eqs. (4) and (8).
7. Obtain a realization of the residuals for T_{max}, $\eta_k(T_{max}) + \varepsilon_{jk}(T_{max})$, at every location $j = 1,2,\ldots J$ from a multivariate normal distribution with zero mean, total standard deviations computed from Eq. (2), and spatially correlated with the model obtained in step 6.
8. At each location j, obtain a realization of residuals for the rest of the periods, $\eta_k(T_i) + \varepsilon_{jk}(T_i)$, conditioned on the value of the residual $\eta_k(T_{max}) + \varepsilon_{jk}(T_{max})$ computed in step 7 (here is where the Markov approximation is used). The conditional distributions of the residuals for the rest of the periods, T_i, can be computed using a correlation model $\rho_0(T_i, T_{max})$, such as those proposed by Inoue and Cornell (1990), Baker and Cornell (2006), Abrahamson and Silva (2007), Baker and Jayaram (2008), or Abrahamson et al. (2013).
9. Compute $f(T_i, M_k, R_{jk}, \boldsymbol{\theta}_{jk})$ for every location j and each period i, and finally obtain the ground motion intensity measures using Eq. (1).
10. Repeat steps 5 through 9 for the total number of simulations.

This procedure can also be used with efficient sampling schemes, such as importance sampling (Rubinstein 1981), as the standard Monte Carlo simulation method is not computationally efficient for estimating low-probability high-consequence risks (Au and Beck 2003). Moreover, note that this simulation sequence can also be applied with other intraevent spatial correlation models different than the one presented in Eq. (8), incorporating a variability into the model parameters, as did with β in this study. However, values of dispersion must be estimated with a similar approach than the one presented in this paper.

6 Conclusions

This study quantitatively evaluated the event-to-event variability of the intraevent spatial correlation of common ground motion intensity measures (*PGA* and spectral acceleration ordinates). For this, 39 world-wide seismic events, each having more than 100 recording stations, were considered, for a total of 15,940 ground motion records. An exponential model as a function of separation distance and using a single parameter, β, was fitted to every event independently. A probabilistic assessment of the model parameter was conducted, showing that it follows a lognormal distribution, and that the logarithmic standard deviation around its central tendency can be as high as 1.2. Different linear regression analyses were performed, and although the event moment magnitude was found to be statistically significant as a predictor variable at long periods, it explains less than 20% of the total variability of β. Moreover, the spatial correlation of V_{S30} values (as a proxy for clustering of site geological conditions, while the depth of sediments was not considered), the tectonic region, and the fault mechanism of the earthquakes were found not statistically significant at a 5% significance level.

Finally, this paper has presented a simple and direct Monte Carlo simulation approach for explicitly considering the event-to-event variability of the spatial correlation model when performing regional seismic hazard and risk analyses. The proposed sequence of

simulation takes advantage of the Markov dependence of residuals for reducing the total number of correlated variables to be simulated, therefore greatly decreasing the computational effort involved. Explicit consideration of the event-to-event variability of the spatial correlation model will provide improved results when conducting regional hazard and risk assessments.

Acknowledgements The authors would like to acknowledge CONICYT—Becas Chile, the Nancy Grant Chamberlain Fellowship, the Charles H. Leavell Fellowship, the Shah Graduate Student Fellowship, and the John A. Blume Fellowship for their financial support to the first author for conducting his doctoral studies under the supervision of the second author. Records used in this investigation were obtained from the PEER NGA-West2 ground motion database. The authors are grateful to the various government agencies responsible for the installation and maintenance of seismic instrumentation and for making their data publicly available, and to PEER for collecting, processing, and distributing these records. The authors would also like to thank the two anonymous reviewers, whose comments helped improve the quality of this paper.

References

Abrahamson NA (1988) Statistical properties of peak ground accelerations recorded by the SMART 1 array. Bull Seismol Soc Am 78:26–41

Abrahamson NA, Silva WJ (2007) Abrahamson & Silva NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. PEER Report Draft v2, Pacific Earthquake Engineering Research Center, Berkeley, CA

Abrahamson NA, Silva WJ, Kamai R (2013) Update of the AS08 ground-motion prediction equations based on the NGA-West2 data set. PEER Report 2013-04, Pacific Earthquake Engineering Research Center, Berkeley, CA

Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BS-J, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T, Donahue JL (2014) NGA-West2 database. Earthq Spectra 30:989–1005

Au SK, Beck JL (2003) Subset simulation and its application to seismic risk based on dynamic analysis. J Eng Mech 129:901–917

Baker JW, Cornell CA (2006) Correlation of response spectral values for multicomponent ground motions. Bull Seismol Soc Am 96:215–227

Baker JW, Jayaram N (2008) Correlation of spectral acceleration values from NGA ground motion models. Earthq Spectra 24:299–317

Blom G (1958) Statistical estimates and transformed beta-variables. Wiley, New York

Boore DM, Gibbs JF, Joyner WB, Tinsley JC, Ponti DJ (2003) Estimated ground motion from the 1994 Northridge, California, earthquake at the site of interstate 10 and La Cienega Boulevard bridge collapse, West Los Angeles, California. Bull Seismol Soc Am 93:2737–2751

Boore DM, Stewart JP, Seyhan E, Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. Earthq Spectra 30:1057–1085

Bozorgnia Y, Abrahamson NA, Al Atik L, Ancheta TD, Atkinson GM, Baker JW, Baltay A, Boore DM, Campbell KW, Chiou BSJ, Darragh R, Day S, Donahue J, Graves RW, Gregor N, Hanks T, Idriss IM, Kamai R, Kishida T, Kottke A, Mahin SA, Rezaeian S, Rowshandel B, Seyhan E, Shahi S, Shantz T, Silva W, Spudich P, Stewart JP, Watson-Lamprey J, Wooddell K, Youngs R (2014) NGA-West2 research project. Earthq Spectra 30:973–987

Cunnane C (1978) Unbiased plotting positions—A review. J Hydrol 37:205–222

Goda K (2011) Interevent variability of spatial correlation of peak ground motions and response spectra. Bull Seismol Soc Am 101:2522–2531

Goda K, Atkinson GM (2009) Probabilistic characterization of spatially correlated response spectra for earthquakes in Japan. Bull Seismol Soc Am 99:3003–3020

Goda K, Atkinson GM (2010) Intraevent spatial correlation of ground-motion parameters using SK-net data. Bull Seismol Soc Am 100:3055–3067

Goda K, Hong HP (2008) Spatial correlation of peak ground motions and response spectra. Bull Seismol Soc Am 98:354–365

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, New York, p 483

Hong HP, Zhang Y, Goda K (2009) Effect of spatial correlation on estimated ground-motion prediction equations. Bull Seismol Soc Am 99:928–934

Inoue T, Cornell CA (1990) Seismic hazard analysis of multi-degree-of-freedom structures. Reliability of marine structures. Report RMS-8, Stanford, CA

Jayaram N, Baker JW (2008) Statistical tests of the joint distribution of spectral acceleration values. Bull Seismol Soc Am 98:2231–2243

Jayaram N, Baker JW (2009) Correlation model for spatially distributed ground-motion intensities. Earthq Eng Struct Dyn 38:1687–1708

Kawakami H, Mogi H (2003) Analyzing spatial intraevent variability of peak ground accelerations as a function of separation distance. Bull Seismol Soc Am 93:1079–1090

Lee R, Kiremidjian AS (2007) Uncertainty and correlation for loss assessment of spatially distributed systems. Earthq Spectra 23:753–770

Loth C, Baker JW (2013) A spatial cross-correlation model of spectral accelerations at multiple periods. Earthq Eng Struct Dyn 42:397–417

Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. J Am Stat Assoc 46:68–78

Park J, Bazzurro P, Baker JW (2007) Modeling spatial correlation of ground motion intensity measures for regional seismic hazard and portfolio loss estimation. In: 10th international conference applications of statistics and probability in civil engineering. July 31-August 3, 2007, Tokyo, Japan

Rubinstein RY (1981) Simulation and the Monte-Carlo method. Wiley, New York

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

Sokolov V, Wenzel F (2013a) Further analysis of the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA and PGV data from the K-NET and the KiK-net (Japan) networks. Bull Earthq Eng 11:1909–1926

Sokolov V, Wenzel F (2013b) Spatial correlation of ground motions in estimating seismic hazards to civil infrastructure. In: Tesfamariam S, Goda K (eds) Handbook of seismic risk analysis and management of civil infrastructure systems. Woodhead Publishing Limited, Cambridge, pp 57–78

Sokolov V, Wenzel F, Wen KL, Jean WY (2012) On the influence of site conditions and earthquake magnitude on ground-motion within-earthquake correlation: analysis of PGA data from TSMIP (Taiwan) network. Bull Earthq Eng 10:1401–1429

Wang M, Takada T (2005) Macrospatial correlation model of seismic ground motions. Earthq Spectra 21:1137–1156

Wesson RL, Perkins DM (2001) Spatial correlation of probabilistic earthquake ground motion and loss. Bull Seismol Soc Am 91:1498–1515